DOES A THIRD BOUND HELP? PARAMETRIC AND NONPARAMETRIC WELFARE MEASURES FROM A CV INTERVAL DATA STUDY.

Riccardo Scarpa[1] and Ian Bateman[2]

[1] Corresponding author: Riccardo Scarpa, DEAR, Universita` della Tuscia, Viterbo, 01100, Italy and AE&FM, University of Newcastle, UK.

[2] Ian Bateman, Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, Norwich, UK.

Abstract

Positive response density estimation from CV interval data affords efficiency gains which must be weighed against the risk of introducing potential bias during questions iteration. This study examines the effect of a eliciting a third response on a set of often-used welfare measures derived in a conventional parametric setting. It then compares these with distribution-free nonparametric estimates. A third bound increases censoring probability, introduces welfare estimates sensitivity to inclusion of a theoretically relevant covariate such as wealth which also affects efficiency gains. This might well introduce complications that outweigh the expected efficiency gain. This empirical finding supports and complements previous results obtained via simulation.

Does a third bound help? Parametric and nonparametric welfare measures from a CV interval data study.

Introduction

In the practice of contingent valuation (CV) discrete choices are generated by asking survey respondents take-it-or-leave-it questions concerning their willingness to pay (WTP) a specified sum for a given change in the provision of a good. This elicitation format originates in the seminal work of Bishop and Heberlein (1979) but has been substantially extended by Hanemann (1984) and Hanemann *et al.*, (1991) such that a initial choice response may be supplemented through asking one or more follow-up discrete choice questions. In a WTP experiment these work by increasing or decreasing the bid amount presented to respondents in the initial question such that a respondent who responds positively to the initial 'first bound' question is presented a higher bid in the follow-up 'second bound'. Similarly a respondent who rejects the first bound amount will be presented with a lower bid in the second bound question. Under a given set of behavioral assumptions, the additional information on a respondent's true WTP provided by such an approach yields more precise estimation of structural parameters and associated welfare estimates (Hanemann *et al.* 1991).

How much of an improvement is delivered by follow-ups under the conventional discrete choice modeling is an issue of critical importance when making a decision between using single or multiple bound elicitation formats in a CV survey. While the inclusion of follow-up questions brings benefits in terms of additional information, they may also increase the number of missing or incomplete responses. Such sample censoring curtails the inferential power of welfare and parameter estimates. Furthermore, the standard assumptions underpinning the validity of discrete choice CV responses (Hoehn and Randall, 1987) may fail as we move from the first to successive bounds. For example, such a move may induce changes in the respondents' perception of the nature of the contingent market, such that they feel that the probability or cost of provision may alter (Carson *et al.*, 1994). Follow-ups may also introduce various distortions in the truthfulness of subsequent responses. For example, the respondent who has initially expressed a positive response may feel that a bargain has been struck at that response amount, and so be indignant at the presentation of a subsequent higher bid amount even if it was one which, had it been presented initially, they would have agreed to pay. Conversely the respondent who refuses an initial bid amount may feel guilt for not contributing to a good social cause and agree to pay a subsequent lower amount even though they would have refused such an amount if presented at the first bound.

While a number of possible strategic behavior and relative biases can be conceived, empirical evidence for such effects is mixed and inconclusive (Cameron and Quiggin, 1994; Herriges and Shogren, 1996; McLeod and Bergland, 1997). Monte Carlo simulations run under the assumption that the sequential responses are indeed generated by a normal bivariate process, have shown that the bias of the interval-data welfare estimates tends to be low, especially in models estimated from large samples (Alberini, 1995). Given this then, under an assumption of either truthful response or that, despite untruthful responses the expected value of the unobservable component of the latent variable is still zero and has a given distribution, one could use an *n*-bounded response to estimate a series of *n* differently bounded models. When adequately designed, this bounding can be expected to improve the efficiency of the parametric estimates

(Hanemann, *et al*., 1991). Given the current cost of conducting a state-of-the-art CV survey any method that delivers a given precision of welfare estimate by relying on a reduced sample size is worth investigating.

Given the above assumptions, four issues are of general interest in the derivation of welfare measures from multiple bounded CV data. The first concerns the assessment of the gains in efficiency of parameter and welfare estimation delivered by increasing the number of bounds from two to three. A second issue concerns the effects of deriving welfare measures from computationally convenient parametric probability estimates as opposed to more robust nonparametric approaches when using multiple bounded data-set. Third, one must assess the trade-off between the loss of inferential power derived from missing and incomplete responses generated by successive bounds and the increase in efficiency due to the "bracketing" of WTP into tighter money intervals. Fourth, does the inclusion of theorethically important covariates affect the welfare estimates magnitude obtained at different bounds?

This paper examines these issues by comparing benefit estimates from a variety of common parametric specifications and a more robust nonparametric one from a CV study on a wetland preservation program (discussed subsequently). These benefit estimates are then compared with previously published results obtained from applying multilevel modelling techniques to the triple-bounded responses (Langford *et al.*, 1996).

Benefit estimates are derived through two approaches. In the first of these, two commonly employed, utility-consistent parametric specifications are applied: the linear and log-logistic models. The maximum likelihood (ML) parameter estimates are obtained and point and interval estimates of implied welfare measures are computed using standard techniques. Likelihood ratio tests are conducted to discriminate between the relative fit of the two specifications to the data. Theoretical consistency is investigated by considering whether a measure of the respondents' wealth has a positive effect on the estimated probability of a yes response. Finally we investigate truncation point bias under two common selection rules.

Since parametric estimates rely on quite strict and untestable *a-priori* assumptions, their implied welfare measures are inconsistent under misspecification. For this reason our second approach extends the interval data analysis by employing a ML nonparametric estimator for the survival curve of a positive response to bid levels. Survival curves, that is, the estimated Pr(Yes|**x**), are then employed to derive estimates of median and expected WTP which are compared to their parametric counterparts. The nonparametric estimator employed here is the generalized self-consistent algorithm (GSCA) of An and Ayala (1996), which is a generalization of the Turnbull estimator for interval-censored data able to deal with across-interval censored data such as those characterizing this study (Turnbull, 1974,1976). This is a ML estimator yielding a lower bound, conservative estimate of welfare change.

Does a third bound help? Analytical methodology

Under the set of assumptions that underlie the interval data approach, it is natural to question whether successive iterations can be expected to be beneficial in the overall economy of the estimation exercise. A similar question was investigated by Cooper and Hanemann (1995) who employ Monte Carlo techniques to conclude that, while a second bound is worthwhile, simulation results indicate that a third bound does not deliver justifiable efficiency gains. Although Monte Carlo studies are informative, these findings are often specific to the given data generating process employed. The present study employs an empirical approach to the same issue, but our final conclusions are similar.

Because the prevailing reason to carry out CV studies is welfare evaluation we have framed our analysis around this issue. In particular we find it useful to focus attention on the differences between point and interval estimates of conventional welfare measures obtained from both parametric and nonparametric models.

Depending on the choice of model specification and the purpose of the analysis, the economic analyst can have an interest in different welfare measures. Using the established notation in the literature (Hanemann, 1984, 1989) we define the following :

- the median, $C^* = A^*$ s.t. $\Pr(Yes|A) = 0.5$
- the mean, $C^+ = \int_{-\infty}^{\infty} \Pr(Yes \mid A)dA$
- the mean over the non negative orthant, $C^\bullet = \int_0^{\infty} \Pr(Yes \mid A)dA$
- the non-negative truncated mean, $C^\sim = \int_0^{A^{\max}} \Pr(Yes \mid A)dA$

It is noteworthy that $C^\bullet = C^+$ for the log-logistic model and that $C^* = C^+$ for the linear-logistic, while $C^\bullet$ bounds from above all the $C^\sim$ measures. The measure $C^\sim$ is of particular interest in cost-benefit analysis (Duffield and Patterson, 1991) because it is robust to the fat tail problem which often arises when the logistic specification is employed. It also excludes *a-priori* the possibility of negative WTP for the assessed program, which is unlikely to share the same error structure as the positive WTP, given the asymmetry often perceived between WTA and WTP. Furthermore, like the C^+ measure, $C^\sim$ allows for aggregation across the relevant population of interest to derive aggregate benefit measures. However, $C^{\sim}$ is clearly sensitive to the choice of upper tail truncation point A^{max}. For this reason in this paper we investigate two different decision rules for the selection of truncation points.

The first decision rule is "no extrapolation". This corresponds to assuming that the stochastic specification of choice holds only along the empirical support of A. It is therefore truncated above by available income on a per annum basis. However, when bids are smaller than income, as in the case of most CV studies, the highest observed bid A^{max} represents the upper limit of the relevant range. In our study these value are £500, £1000 and £2000 for the SB, DB and TB. For comparative purposes across differently bound interval data we also provide estimates of $C^{\sim}$ for A^{max} = £1000 and A^{max} = £2000 for the SB and DB data, even though for the SB A^{max} = £500 and for DB A^{max} = £1000. Hence for these models some truncation points are beyond the range of observed data. The second truncation point rule is "upper percentile limited extrapolation". This is equivalent to assuming that the estimated survival curve can be extrapolated with confidence only up to a defined upper percentile of the probability distribution. Here we use arbitrarily defined amounts of 1, 5 and 10 percent.

Both rules have decision making relevance. For example, if an application shows that the empirical probability at A^{max} is proximate to zero, then one might reasonably assume to have identified the choke price for the program. This would allow the use of measure $C^{\bullet}$. Conversely, one might find relatively high empirical density of a positive response at A^{max}, and hence think plausible that the estimated survival curve approximates well the way probability dies out on the upper tail, and so rely on an upper percentile limited extrapolation rule. This correspond to using the measure $C^{\sim}$ where A^{max} is chosen to be an upper percentile of the distribution. Finally, one may hold a conservative stand and decide that no extrapolation is warranted off the support of the investigated bid range. In this case the truncation of the mean at the highest bid value used in the survey would be a plausible choice to make, an approach adopted by, amongst others, Bishop and Heberlein (1979) and Sellar, Chavas and Stoll (1985).

Given a choice of welfare measure, although it is plausible to expect changes in magnitude of the estimates across SB, DB and TB because of the additional information that each extra bound provides, however, observing large differences in magnitude within a given bound due to small changes in model specification is at odd with theoretical expectation.

Parametric estimation is not necessary to derive welfare measures, but it does allow an investigation of the structural parameters of the underlying theoretical construct. Comparison with parametric estimators is provided through application of the An and Ayala GSCA approach to yield a non-parametric, robust estimates of the probability of a yes response from the interval data. From these estimated probability masses we integrate discretely over the non-negative bid space to find $C^{\sim}$ and identify the interval within which falls C^{*}. The only *a priori* constraint imposed on this probability step function is for it to be non-increasing on the bid.

All the welfare measures obtained are then compared across models, decision rules and interval-data bounds.

The data

The data-set is from a large sample CV study employing face-to-face survey techniques to ask respondents for their WTP in respect of a saline flooding prevention program in the Norfolk Broads, a major freshwater wetland area of National Park status located in the East Anglian region of the UK (a detailed description of the study is given in Bateman *et al.*, 1995). The survey collected discrete choice responses which were iterated three times, to produce the triple-bounded dataset. Here, when the respondent was initially presented with a bid-amount *A* the subsequent amounts were doubled if the obtained response was positive, else it was halved. The third iteration was carried out only for those respondents who replied Yes,Yes or No,No to the first two iterations in which case the bid amount was either doubled or halved from the second bound level as appropriate. However Yes,No and No,Yes responses were included in both the DB and TB estimations, hence the observed sample censoring is exclusively due to incomplete responses. Due to the random assignment of initial bids and respondents being *a-priori* unaware of the iterated design, the discrete choice dataset offers the possibility of deriving welfare estimates by using SB, DB and TB interval-data analyses.

Out of a total of 2070 interviews, after a standard treatment for outliers and missing data, there are 1727 useable responses to the SB, 1660 for the SB and 1217 for the TB. This implies censoring due to non-response of 16.6%, 19.8% and 41.2% respectively. It is notable that, while censoring increases by only approximately 3% when going from the SB to DB, it rises by 20% when going from the DB to TB. This might have non-trivial consequences in inference and must be accounted for when choosing a multiple bound survey design. In particular this effect must be weighted against the potential efficiency gains afforded by successive bounds.

Choice of model specification and estimation.

A conventional parametric specification is the linear-logistic model. The latent variable determining the Yes-No response is assumed to have the following structure:

$$y = \mathbf{b}'\boldsymbol{x} + u$$

where $\mathbf{b}$ is a vector of parameters to be estimated and $\boldsymbol{x}$ is a vector of variables (including a constant) determining the respondent's choice. The vector $\mathbf{b}'\boldsymbol{x}$ can be interpreted as the underlying valuation function (Cameron, 1987) or as the utility difference function (Hanemann, 1984). In either case the statistical properties of the model are the same, with the exception that the valuation function can conceptually accommodate a larger set of conditioning variables than the utility difference function. To achieve identification this specification must be "closed" by assuming a distribution for the unobserved component u. Many CVM studies derive welfare measures from model estimates based on the assumption that the unobservable component of the alternative utility states is i.i.d. extreme value type I, and hence u is i.i.d. logistic. Given the general lack of theoretical guidance, this assumption is no worse than others and affords quite convenient computational properties. For these reasons we invoke this assumption throughout our parametric estimation. Given this, the conditional probability of a Yes response given $\boldsymbol{x}$ is therefore:

$$\Pr(\text{Yes}|\boldsymbol{x}) = \Lambda(\Delta v \geq u|\boldsymbol{x}) = \Lambda(\mathbf{b}\ \boldsymbol{x}) = [1+\exp(-\mathbf{b}'\boldsymbol{x})]^{-1}.$$

Let $\alpha(\beta_{-A},\boldsymbol{x}_{-A})$ be the inner product between all the conditioning variables with the exception of the bid A and its relative coefficient vector β_{-A} . We now focus our attention on two familiar logistic models:

a) the linear-logistic model where $\Delta v = \mathbf{b}'\boldsymbol{x} = \alpha + \beta A$ as detailed by Hanemann (1984,1989), and;

b) the log-logistic model $\Delta v = \mathbf{b}'\boldsymbol{x} = \alpha + \beta\ln(A)$, first used by Bishop and Heberlein, (1979).

Notice that both these models can be derived from a random utility model (Hanemann and Kanninen, 1996) and that they are consistent with maximization of economic preferences. For the sake of parsimony and without loss of generality for our analysis, we narrow down the choice of dependent variables to bid amount (A) and a wealth index (INC) derived from respondents' reported levels of annual gross household income. The role of covariates is not essential for probability density estimation, and income and bid amount are arguments of both the indirect utility and demand functions as conventionally treated in microeconomic theory (Mas-Colell *et al.*, 1995). They are also used in that they are likely to be the major determinants of willingness-to-pay as measured by equivalent variation, which is the relevant measure of welfare change in this study. In the one case $\boldsymbol{x}_{-A}$ is equal to the unity vector (constant term), while in the other it also includes our wealth proxy INC.

The ML estimates were obtained with Limdep v.7.0 using the log-likelihood functions described in Table 1 and yielded the estimates in Table 2. We also estimated a specification including the Box-Cox transformation of A and its relative coefficient λ. Although likelihood ratio tests of the adequacy of the linear versus the logarithmic specification (adjusted for the number of observations) are inconclusive at the conventional significance levels, the log-linear specification estimates produce log-likelihood values much lower and closer to those of the Box-Cox transformation than those of the linear specifiation. As shown in Table 2, we find λ estimates that are very close to zero for all models, which corroborates the assumption of a logarithmic specification. We conclude that the log-linear specification tends to fit the data better in all cases. However, we report both linear and logarithmic estimates in the light of results from Monte Carlo experiments conducted by Alberini and Cooper (1995) where they conclude:

> "that postulating that utility is a linear function of log income is a "high risk" assumption, in the sense that, should that assumption not be valid, highly biased estimates of welfare would be obtained. A linear-in-income specification of utility appears to be a somewhat safer choice"
>
> (Alberini and Cooper, 1995, page 320).

We then test the H_o that the correct specification includes the proxy for wealth in both log-logistic and linear-logistic forms. In the SB and DB models the inclusion of INC always improves the fit as shown by the decrease in the per-observation likelihood function values, these improvement are not, however statistically significant in likelihood ratio tests. In the TB models the per-observation likelihood function values are higher in the models with INC showing that the censoring due to attrition (incomplete third bound responses) introduces some structural problem. For example, when we estimate the model without INC from the same sample (N=1172) we find a per-observation log-lik. value of -2.25, while the one with INC gives a -2.17, the value for the entire set of complete TB responses instead (N=1217) is -1.67, providing a much better fit without income.

Estimation of welfare measures and choice of truncation point.

From the ML parameter estimates of the linear-logistic and log-logistic models we derived point and interval estimates of the following four well known welfare estimates:

a) C^* , that is the median WTP, which is computed using the observation that if $\Delta v = \alpha + \beta A$ then the value setting $\Delta v = 0$ can be derived by solving for A where $\Delta v = 0$, which leads to $C^* = -\alpha/\beta$. Analogously, for $\Delta v = \alpha + \beta \ln A$ then $C^* = \exp(-\alpha/\beta.)$. These are the results presented by Hanemann (1984) and they apply to every linear specification when the expected value of the error term is zero and its distribution is symmetric;

b) C^+, the expectation of WTP over the real line, which is equal to $C^* = -\alpha/\beta$ for the linear logistic model and which, when $|\beta| \geq 1$, is defined in closed-form as follows:

$$C^+ = \exp(-\alpha/\beta)\,\Gamma(1+1/\beta)\Gamma(1-1/\beta) = \exp(-\alpha/\beta)\,\pi/\beta\,[\sin(\pi/-\beta)]^{-1} = C^*\,\pi/\beta\,[\sin(\pi/-\beta)]^{-1}$$

Notice that this quantity, when defined, is always greater than C^* since $\pi/\beta\,[\sin(\pi/-\beta)]^{-1} \geq 1$.

c) $C^{\sim}$, the expectation over the non-negative orthant of the real line, truncated at some upper value A^{max}. While this quantity can only be numerically approximated in the log-logistic case, it can be expressed in closed-form for the linear logistic specification as follows (Hanemann, 1989, eq. 10):

$$C^{\sim} = 1/\beta\,\ln(1+\exp(\alpha))(1+\exp(\alpha + \beta\,A_{max}).$$

d) $C^{\bullet}$, which is the untruncated expectation over the entire non-negative orthant of the real line, and it is equal to C^+ for the log-logistic model, for which $\ln A$ spans [0, infinity), while it is defined as $1/\beta\,\ln(1+\exp(\alpha))$ for the linear-logistic model (Hanemann, 1989, eq. 11).

Interval estimates were derived by resampling 10,000 times the estimated asymptotic distribution (Krinsky and Robb, 1986) of the ML parameters, which is known to be normal centered on the population parameters, and with variance-covariance matrix $\mathbf{\Omega}$, which we approximate with their respective ML estimates $\boldsymbol{\beta}_{ML}$ and $\mathbf{\Sigma}_{ML}$. Cooper, (1994) shows via simulation that this method tends to underpredict the true confidence intervals of various data generating processes, particularly those derived from large samples. Although it is not known how general this result is, such a finding does suggest that we should be cautious about the interval estimates reported here and treat them as lower bound estimates of the real confidence intervals.

Nonparametric welfare measure from GSCA probability estimates

Survival curves for censored and interval grouped data can also be estimated using nonparametric techniques. This avoids the strong assumptions concerning distribution and model specification inherent in parametric approaches and thereby yields relatively robust estimates of WTP. However, nonparametric estimates are relatively sample inefficient and depend heavily upon the choice of intervals between bids (McFadden, 1994), a property which can cause problems where sample size is small. Computationally, nonparametric estimates can be derived by means of various estimators. The first step of the process is the estimation of a survival function which defines the probability of a positive response at a given bid amount.

Some theoretical restrictions are normally included, such as weak monotonicity. This simply incorporates the intuitive fact that, the probability of a positive response will not increase as the bid amount rises. One algorithm that ensures this property is the so called PAVA (pool-adjacent-violator-algorithm), by which all the bids at which the sample frequency is higher than the one recorded for the adjacent lower bid are pooled in a unique bid group and their frequency is assigned at the lowest of the two bids.

Once these survival curve estimates are derived, expected WTP is computed by integrating discretely under the curve. Since only the probability mass at each given bid can be estimated, any interpolation between these points is arbitrary. Kriström (1990), for example, uses linear interpolation while Scarpa *et al.* (1998) use a weaker concept of local continuity and implemented with kernel estimation. Conversely, Carson *et al.* (1992) produce conservative, lower-bound estimates of WTP by placing the probability mass at the lower bid amount[i].

In a CV study the first bound discrete response provides a left or right-censored observation on the underlying WTP of the respondent. Given a bid amount b_i a positive response will place $WTP_i \geq A_i$; while a negative one will place $WTP_i < A_i$. Given a positive first bound response then the second bound bid amount A_i further restricts the latent variable WTP_i to either the interval $[A_i, A_i)$ in case of a Yes-No sequence, or to $WTP_i \geq A_i$ in the case of a Yes-Yes sequence, where $A_i > A_i$. Similarly, an iteration following a negative first bound response further restricts WTP_i to the interval $[A_i, A_i)$ for the No-Yes sequence or it places $WTP_i < A_i$ for a No-No sequence, where $A_i < A_i$. Consequently the DB design yields WTP observations which are left or right-censored (as in the SB case) as well as interval-censored. The latter can be further divided into two groups of observations. The first case is when the sequence of responses is designed so as to perfectly partition the final WTP range into non overlapping intervals. Alternatively, these intervals may not be a perfect partition of the bid space and the sequence of responses may imply for some WTP to be within intervals that include other bid amounts. In this case some observations are across-interval-censored, that is, the censoring takes place somewhere within one of the overlapping intervals within which the bid support is partitioned.[ii]

When the grid of bids is designed so as to obtain a perfect partitioning of the sequence of responses into non-overlapping WTP intervals, then the algorithms proposed by Turnbull (1974,1976) provide ML self-consistent probability estimates of the survival curve without recourse to any parametric specification. This approach uses a reallocation mechanism to reduce the multiple-censored problem to a single-censored one, to which the Kaplan-Meier (1958) estimation technique can be applied to obtain consistent and readily tractable probability estimates. However, when we are faced with a choice of bids that do not define a complete partitioning of the bid range, then the strict inequality conditions upon which the Turnbull reallocation algorithms rely no longer apply (Ayala, 1995). The An and Ayala (1996) GSCA algorithm overcomes this limitation to yield readily computed ML estimates. However, this approach does not permit the estimation of standard errors[iii]. Nevertheless, the ease of computation of An and Ayala's GSCA is particularly rewarding when dealing with large datasets, such as the one employed in this study, where sample size ranges from 1217 to 1727.

Results

Tables 3 and 4 present welfare measure estimates for our linear and log-logistic models respectively across all three bounds while Table 5 presents Turnbull probability estimates and associated welfare measures for our nonparametric models again across all three bounds. Figures 1 to 3 illustrate the estimated survival curves for each bound for the linear-logistic, log-logistic and nonparametric models respectively. Consideration of these results shows that estimates of any given welfare measure decrease as one moves from SB to DB formats. This result is consistent across all parametric and nonparametric measures and models. As we move form SB to DB we also register a marked tightening of the 95% confidence intervals across all measures and models. Changes in the width of 95% confidence intervals around the point estimates of these welfare measures are reported in Table 6.

Moving from DB to TB format brings about a further decrease for all welfare measures, except the nonparametric GSCA mean WTP, for which differences are small and therefore likely to be insignificant. Notice that in this case the confidence interval around the TB point estimates of the welfare measures is not much tighter than in the DB case for the welfare measures estimated from the model which uses only bid as a conditioning variable, indicating that most of the efficiency gains are realised by a DB specification, in accordance to the results of Cooper and Hanemann (1995). For TB welfare estimates derived from models that also used INC as a covariate we report an efficiency loss in the linear-logistic estimates, as illustrated by the widening of the c.i. with respect to those registered in the DB case. This is not the case for the TB estimates from the log-logistic with wealth case, for which we do observe the expected efficiency gains. This, however gave welfare estimates much lower than the remaining set.

Across the various model specifications and bounds, welfare measures are reasonably stable for our linear-logistic models but vary substantially for our log-logistic specifications. For the latter, it appears that the estimated C^{+} is implausibly high. This suggests that, although the log-logistic specification fits the data better according to log-likelihood tests, derived welfare measures should be treated with caution. As shown in Figure 2, this problem is due to a particularly fat tail in the log-logistic density (Figure 1 shows this not to be a particular problem for the linear-logistic models). In this case the truncated means $C^{\bullet}$ and $C^{\sim}$ are certainly more reasonable than the mathematical expectation C^{+}. Furthermore, the choice of truncation point seems to matter. For example, an upper percentile extrapolation yields estimates that are implausibly high, while the expectations truncated at the maximum bid in the range (shown in bold in Tables 4 and 5) are within £40 across different interval-bounds. Inspection of these results suggests that measures based upon extrapolations beyond the range of available data are not advisable, particularly in log-logistic specifications and where conservative estimates are required.

Returning to the linear-logistic models, while welfare measures are relatively stable within bounds, they vary considerably across bounds such that in some cases SB models yield estimates which are more than £100 higher than those derived from DB models. While the latter models do provide a good approximation to their corresponding nonparametric estimates, this instability across bounds is cause for some concern.

Considering our parametric models, the general effect of the wealth proxy INC is to reduce the size of the implied welfare measures. This is as expected. However, the observed reduction in measures for the TB case is quite sizable in both models, but particularly for the log-logistic model, which suggests that these common parametrizations may fail to adequately capture some substantive feature of the probability density of a positive response and to be particularly sensitive to variable choice in model specification at the third bound.

If one takes the nonparametric measures described in Table 5 as conservative benchmarks, one would expect the population expected WTP to lie somewhat above £120. In this case the C^{*}, $C^{\sim}$ and $C^{\bullet}$ measures of the DB linear-logistic (and to a somewhat lesser degree the log-logistic) model all perform quite well.

Previous published work on this data yields a number of interesting comparisons with the present results. Many of the SB discrete choice models estimated by Kerr and Graham (1996) are not comparable to ours as they trim the sample of responses over various bid levels (£100, £200 and £500) to simulate the effect of using restricted bid vectors. However, the comparable untrimmed parametric medians do not differ from ours. A similar conclusion can be drawn contrasting our results with those obtained via the multi-level modelling approach adopted by Langford *et al.*, (1996) and Bateman et al (forthcoming). This uses a log-logistic link function without a wealth proxy variable to produce estimates of median WTP falling from £103 to £94 as we move from the first to the third bound. Hence, results from our present study seem to fall in the general range predicted by previous work.

Conclusions

How many bounds are enough? This study tends to support in an empirical setting what Cooper and Hanemann (1995) found in a Monte Carlo simulation context, which is that the costs of a third bound may well exceed its benefits. In this sample, the third bound is accompanied by a 20% increase in censoring probability due to incomplete responses. This *per se* substantially curtails the inferential power of the model estimates. This cost is not counter-balanced by the relatively low increase in efficiency of the parametric welfare estimates produced by the linear and log-logistic specifications using only bid as covariate. In fact, the estimated confidence intervals around the parametric welfare estimates are quite tighter in the DB than in the SB models (Table 3), while no significant reduction in their width is noticed when moving from DB to TB. When a proxy for wealth is included the efficiency gains are negative in the linear-logistic model and still sizeable in the log-logistic, but accompanied by a marked change in the magnitude of the estimates of welfare measures. So, TB data do not produce robust estimates of welfare measures with respect to inclusion of theoretically important wealth variable. A similar argument holds for the third bound effect in the nonparametric probability and welfare estimates, which, while altering substantially between the first and second bound, are relatively similar between the second and third bound.

In conclusion, the addition of a third bound brought a substantial increase of censoring probability and a sizeable sensitivity of estimates of welfare measure to variable inclusion. The expected increase in efficiency was also sensitive to model specification and choice of covariate.

Judged by the standard of nonparametric estimates, the third bound parametric welfare measures were either too conservative in the models with the wealth proxy, or conservative and unstable across choice of welfare measure in the case without this covariate. These empirical results can be added to previous, simulation based arguments (Cooper and Hanemann, 1995) in supporting the use of no more than two bounds in the application of CV discrete choice studies with follow-up questions.

Bibliography

Alberini, A., 1995, Efficiency versus bias of WTP estimates: bivariate and interval-data models, *Journal of Environmental Economics and Management*, 29:169-180.

An Mark Yuying and Ayala Roberto A., 1996, nonparametric estimation of a survivor function with across-interval-censored data, *Manuscript*, Duke University.

Ayala Roberto A., 1995, Econometric issues in valuation of nonmarketable goods, *Ph.D. Dissertation*, Duke University.

Ayer Miriam, Brunk H. D., Ewing G. M. and Reid W. T., 1955, An empirical distribution function for sampling with incomplete information, *Annals of Mathematical Statistics*, 26:357-365.

Bateman Ian J., Langford Ian H., Turner Kerry R., Willis Ken G., Garrod Guy, D. 1995, Elicitation and truncation effects in contingent valuation studies, *Ecological Economics*, 12:161-179

Bateman, I.J., Langford, I.H. and Rasbash, J. (forthcoming) Willingness-to-pay question format effects in contingent valuation studies, in Bateman, I.J. and Willis, K.G. (eds.) (forthcoming) *Contingent Valuation of Environmental Preferences: Assessing Theory and Practice in the USA, Europe, and Developing Countries*, Oxford University Press, in press.

Bishop, R. C. and Heberlein, T. A.,1979, Measuring values of extramarket goods: Are indirect measures biased?, *American Journal of Agricultural Economics*, 61(5): 926-930.

Cameron Trudy Ann and Quiggin John, 1994, Estimation using contingent valuation data from a dichotomous choice with follow-up questionnaire, *Journal of Environmental Economics and Management*, 27:218-234.

Carson R. T., R. C. Mitchell, W. M. Hanemann, R. J. Kopp, S. Presser, P. A. Ruud, 1992, A CV study of lost passive use values resulting from Exxon Valdez oil spill. Report submitted to the Attorney General of the State of Alaska. Reprinted by Natural Resource Damage Assessment.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S. and Ruud, P.A. (1994) Contingent valuation and lost passive use: damages from the Exxon Valdez, *Discussion Paper 94-18*, Resources for the Future, Washington, D.C.

Cooper Joseph C. and Hanemann W. M., 1995, Referendum contingent valuation: how many bounds are enough? *Manuscript.*University of California, Berkeley.

Cummings Ronald G. , Elliott Steven, Harrison Glenn W. and Murphy James, 1997, Are hypothetical referenda incentive compatible?, *Journal of Political Economy,* June, 105(3).

Duffield, J.W. and Patterson, D.A., 1991. Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. *Land Economics*, 67:225-239.

Haab T.C. and K.E. McConnell, and (1996) Referendum models and negative WTP: Alternative solutions, *Journal of Environmental Economics and Management*, 32(2):251-270.

Hanemann Michael W. and Kanninen Barbara, 1996, The statistical analysis of discrete-response CV data, Working Paper n. 798, Giannini Foundation of Agricultural Economics, University of California at Berkeley.

Hanemann Michael W., Loomis John and Kanninen Barbara, 1991, Estimation efficiency of double bounded dichotomous choice contingent valuation, *American Journal of Agricultural Economics*, November, 73(4):1255-1263

Hanemann W. M., 1984, Welfare evaluations in contingent valuations experiments with discrete responses, *American Journal of Agricultural Economics*, 66 :332-341.

Hanemann W. M., 1989, Welfare evaluations in contingent valuations experiments with discrete response data: a reply, *American Journal of Agricultural Economics*, 71:1057-1061.

Kerr Geoffrey N. and Graham Andreas W. K., 1996, Elicitation and truncation effects in contingent valuation studies: comment, *Ecological Economics*, 19:261-264.

Krinsky I. and A. Robb, 1986, Approximating the statistical properties of elasticities, *Review of Economics and Statistics*, 68: 715-719.

Kriström, B., 1990, A non-parametric approach to the estimation of welfare measures in discrete response valuation studies, *Land Economics*, 66(2):135-139.

Langford Ian H., Bateman Ian J. and Langford Hugh D., 1996, A multilevel modelling approach to triple-bounded dichotomous choice contingent valuation, *Environmental and Resource Economics*, 7:197-211.

Mas-Colell Andreu, Whinston Michael and Green Jerry R., 1995, *Microeconomic Theory*, Oxford University Press;

McFadden Daniel, 1994, Contingent valuation and social choice, *American Journal of Agricultural Economics*, November, 76:689-708.

McLeod Donald M. and Bergland Olvar, 1997, WTP estimates using the DB-DC CV format: a test of validity and precision in a Bayesian framework, *Land Economics*, in press.

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

Seller Christine, Stoll John R. and Jean-Paul Chavas, 1985, Validation of empirical measures of welfare change: a comparison of nonmarket techniques, *Land Economics*, 61(2):156-175

Turnbull Bruce W., 1974, Nonparametric estimation of a survivorship function with doubly censored data, *Journal of the American Statistical Association*, 69:169-173.

Turnbull Bruce W., 1976, The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society*, 290-295.

Endnotes

[i] Haab and McConnell (1995) note that such an approach also remedies the so called "fat-tail" problem of survival functions based on common parametric specifications, which tends to produce high welfare estimates when the integration is extrapolated beyond the bid vector.

[ii] An example from the data-set at hand is the following: the interval between 5 and 10 pounds has interval censored observations because all three bounds identify observations in this interval. However, the interval between 20 and 50 pounds contains across-interval-censored data, because the second bound adds three more intervals 20-25, 25-40 and 40-50. This is visible from Table 4.

[iii] These can be approximated using bootstrapping techniques, although these are not applied here.

Tables

Table 1. Likelihood functions

Single Bound	Double Bound	Triple Bound
$\pi_i^{y} = 1 - F(\beta' x)$	$\pi_i^{yy} = 1 - F(\beta' x_u)$	$\pi_i^{yyy} = 1 - F(\beta' x_{uu})$
$\pi_i^{n} = F(\beta' x)$	$\pi_i^{yn} = F(\beta' x_u) - F(\beta' x)$	$\pi_i^{yyn} = F(\beta' x_{uu}) - F(\beta' x_u)$
	$\pi_i^{ny} = F(\beta' x) - F(\beta' x_d)$	$\pi_i^{yn} = F(\beta' x_u) - F(\beta' x)$
	$\pi_i^{nn} = F(\beta' x_d)$	$\pi_i^{ny} = F(\beta' x) - F(\beta' x_d)$
		$\pi_i^{nny} = F(\beta' x_d) - F(\beta' x_{dd})$
		$\pi_i^{nnn} = F(\beta' x_{dd})$

$F(\cdot)$ is the c.d.f.; subscripts u and d refer to an increase (up = u) or decrease (down = d) in bid values; π indicates the contribution of a observation to the likelihood function and its superscripts refer to the various possible sequence of discrete responses (where y = 'yes' response and n = 'no' response).

Table 2: Estimated logistic models.

	Linear-logistic		Box-Cox		Log-logistic	
	A only	*A* and wealth	*A* only	*A* and wealth	*A* only	*A* and wealth
			Single Bound			
N	1726	1660	1726	1660	1726	1660
L-Lik./obs.	-0.450	-0.440	-0.396	-0.386	-0.396	-0.386
Constant	1.9589 (0.0813)	1.3274 (0.1919)	5.7105 (0.6413)	5.4080 (0.7398)	5.1892 (0.2549)	4.7290 (0.3129)
A or ln(*A*)	-0.0081 (0.0004)	-0.0086 (0.0004)	-1.3394 (0.3437)	-1.4653 (0.3752)	-1.0484 (0.0547)	-1.0936 (0.0578)
Income	———	0.1316 (0.0356)	———	0.1231 (0.0372)	———	0.1255 (0.0379)
λ	———	———	-0.0597 (0.0623)	-0.0710 (0.0622)	———	———
			Double Bound			
N	1660	1660	1660	1660	1660	1660
L-Lik./obs.	-1.417	-1.409	-1.166	-1.158	-1.169	-1.161
Constant	1.6898 (0.0814)	0.9979 (0.1681)	4.6219 (0.3213)	3.8810 (0.3640)	5.3359 (0.1655)	4.6308 (0.2181)
A or ln(*A*)	-0.0128 (0.0003)	-0.0129 (0.0003)	-0.8565 (0.1302)	-0.8581 (0.1298)	-1.2144 (0.0386)	-1.2285 (0.0387)
Income	———	0.1380 (0.0286)	———	0.1506 (0.0284)	———	0.1485 (0.0284)
λ	———	———	0.0853 (0.0362)	0.0877 (0.0361)	———	———
			Triple Bound			
N	1217	1172	1217	1172	1217	1172
L-Lik./obs.	-1.678	-2.173	-1.373	-1.375	-1.378	-1.380
Constant	1.5513 (0.0888)	-0.9324 (0.1748)	4.5081 (0.2941)	-0.7507 (0.1220)	5.3165 (0.1677)	-0.6635 (0.1155)
A or ln(*A*)	-0.0147 (0.0003)	-0.0123 (0.0002)	-0.8525 (0.1208)	-0.7724 (0.0464)	-1.2609 (0.0402)	-0.9066 (0.0333)
Income	———	0.4870 (0.0469)	———	0.9501 (0.0184)	———	0.9365 (0.0461)
λ	———	———	0.0982 (0.0348)	0.0624 (0.0458)	———	———

Table 3. Estimates of welfare measures from linear-logistic models (£/household/year).

Welfare Measures	Single Bound		Double Bound		Triple Bound	
	A only	*A* and wealth	*A* only	*A* and wealth	*A* only	*A* and wealth
$C^{*}= C^{+}$	241.84	234.21	131.35	132.36	104.85	32.55
	221.64-265.77	214.49-256.77	122.71 -139.80	123.74- 140.83	95.83-113.39	18.59-45.91
$C^{\bullet}$	258.13	248.76	144.51	145.06	117.84	74.22
	237.30-284.04	228.19-272.69	137.74-151.35	138.33 -151.97	111.08-124.45	66.30- 82.34
$C^{\sim}$, (0, *A*max=0.01q)	256.88	247.59	143.7325	144.29	117.15	73.39
	236.57-281.39	227.56-270.35	137.00-150.47	137.56 -151.13	110.45 -123.75	65.62- 81.38
$C^{\sim}$, (0, *A*max=0.05q)	248.41	242.80	140.53	141.11	112.61	70.04
	232.71-271.84	224.47-262.76	134.04-147.00	134.58-147.70	107.67-120.59	62.80-77.45
$C^{\sim}$, (0, *A*max=0.10q)	241.97	236.51	136.32	136.95	108.99	65.65
	227.56-261.55	220.18-253.77	130.05-142.51	130.73 -143.19	104.16-116.53	58.99-72.32
$C^{\sim}$, (0, *A*max=500)	**241.87**	**237.50**	143.83	144.40	115.87	73.95
	227.48-261.40	**220.89-255.14**	137.09- 150.58	137.66-151.26	110.88-124.25	66.08- 82.02
$C^{\sim}$, (0, *A*max=1000)	254.31	248.59	**144.50**	**145.05**	116.05	74.20
	237.15-283.30	228.13-272.34	**137.73-151.33**	**138.32-151.96**	111.07- 124.44	66.29-82.33
$C^{\sim}$, (0, *A*max=2000)	254.51	248.75	144.50	145.05	**116.05**	**74.21**
	237.29-284.03	228.18-272.69	137.75-151.19	138.32-151.96	**111.07 -124.44**	**66.29-82.33**

Table 4. Estimates of welfare measures from log-logistic models (£/household/year)

	SB		DB		TB	
	A only	*A* and wealth	*A* only	*A* and wealth	*A* only	*A* and wealth
C^{*}	141.12	137.50	80.95	81.49	67.79	8.13
	124.40-161.23	120.98-156.56	74.33-88.39	74.71-88.98	61.87-74.31	7.37-8.90
$C^{+} = C^{\bullet}$	2925.97	1486.92	397.65	377.75	279.07	Non convergent
	871.62-36502.01	150.30-1810.09	294.66-621.57	290.43-632.82	213.87-396.08	
$C^{\sim}$, (0, Amax=0.01q)	568.71	496.39	230.20	226.33	178.87	51.31
	445.92-653.60	396.00-600.41	203.13-263.89	199.97-257.81	157.55-204.38	42.83-61.60
$C^{\sim}$, (0, Amax=0.05q)	386.64	349.77	174.88	173.18	139.04	29.99
	329.28-428.25	301.11-397.08	160.20-191.64	158.86-189.35	127.06-152.56	26.72-33.59
$C^{\sim}$, (0, Amax=0.10q)	303.79	279.84	145.53	144.68	117.14	21.89
	268.98-329.97	250.09-308.80	135.76-156.24	135.01-155.15	108.91-125.91	20.02-23.88
$C^{\sim}$, (0, Amax=500)	**211.82**	**207.34**	146.07	145.91	126.85	39.63
	195.36-225.25	**191.60-222.62**	136.22-156.89	136.06-156.62	117.02-137.59	34.23-45.79
$C^{\sim}$, (0, Amax=1000)	288.01	277.23	**178.92**	**178.07**	151.05	48.03
	256.96-311.79	248.07-305.69	**163.46-196.72**	**162.82-195.45**	136.54-167.71	40.48-57.06
$C^{\sim}$, (0, Amax=2000)	368.27	348.19	208.38	206.63	**171.88**	**57.07**
	316.41-406.21	299.94-395.06	186.64-234.61	185.34-231.76	**152.56-194.92**	**46.89-69.81**

Table 5. Turnbull probability estimates and relative Mean and Median (Bold) WTP estimates

A	Pr(Yes|*A*) (SB)	*A*	Pr(Yes| *A*) (DB)	*A*	Pr(Yes| *A*) (TB)
0	1.000	0	1.000	0	1.000
1	0.977	0.5	0.999	0.5	0.999
5	0.940	1	0.999	1	0.995
10	0.906	2	0.977	2	0.991
20	0.698	2.5	0.977	2.5	0.981
50	0.592	5	0.972	4	0.950
100	0.423	10	0.924	5	0.932
200	0.217	20	0.822	10	0.890
Mean	176.368	25	0.815	12.5	0.824
St.Error	9.440	40	0.641	20	0.819
		50	0.641	25	0.648
		100	0.462	40	0.622
		200	0.252	**50**	0.534
		250	0.252	**80**	0.405
		400	0.067	100	0.313
		500	0.067	125	0.197
		1000	0.005	200	0.197
		Mean	116.610	250	0.057
				400	0.056
				500	0.016
				800	0.015
				1000	0.005
				Mean	118.173

Table 6. Width of confidence intervals in double bound and third bound estimates as percent of single bound interval estimates.

	Linear-logistic w/o wealth			Linear-logistic with wealth		
	Percent of SB c.i.			Percent of SB c.i.		
Welfare measure	DB	TB	Difference	DB	TB	Difference
$C^* = C^+$	38.84%	39.91%	-1.07%	40.42%	64.62%	-24.20%
$C^\bullet$	28.96%	28.45%	0.51%	30.65%	36.04%	-5.39%
$C^{\sim}$, (0, Amax=0.01q)	30.61%	30.23%	0.39%	31.71%	36.83%	-5.12%
$C^{\sim}$, (0, Amax=0.05q)	34.11%	34.00%	0.11%	34.09%	38.06%	-3.98%
$C^{\sim}$, (0, Amax=0.10q)	37.76%	37.48%	0.27%	37.09%	39.68%	-2.59%
$C^{\sim}$, (0, Amax=500)	39.68%	39.32%	0.35%	39.71%	46.54%	-6.83%
$C^{\sim}$, (0, Amax=1000)	29.57%	29.07%	0.50%	30.85%	36.28%	-5.43%
$C^{\sim}$, (0, Amax=2000)	28.60%	28.45%	0.15%	30.64%	36.04%	-5.39%

	Log-logistic w/o wealth			Log-logistic with wealth		
	Percent of SB c.i.			Percent of SB c.i.		
Welfare measure	DB	TB	Difference	DB	TB	Difference
C^*	38.65%	33.06%	5.59%	40.11%	4.30%	35.81%
$C^+ = C^\bullet$	0.92%	0.51%	0.40%	20.63%	N.A.	N.A.
$C^{\sim}$, (0, Amax=0.01q)	29.33%	22.60%	6.73%	28.30%	9.18%	19.11%
$C^{\sim}$, (0, Amax=0.05q)	32.32%	25.25%	7.07%	31.77%	7.16%	24.61%
$C^{\sim}$, (0, Amax=0.10q)	37.04%	31.48%	5.56%	34.30%	6.57%	27.73%
$C^{\sim}$, (0, Amax=500)	70.00%	66.67%	3.33%	66.28%	37.27%	29.01%
$C^{\sim}$, (0, Amax=1000)	61.82%	56.36%	5.45%	56.63%	28.77%	27.85%
$C^{\sim}$, (0, Amax=2000)	54.44%	45.56%	8.89%	48.80%	24.10%	24.71%

Figure 1: Estimated probabilities of a positive response from logistic models

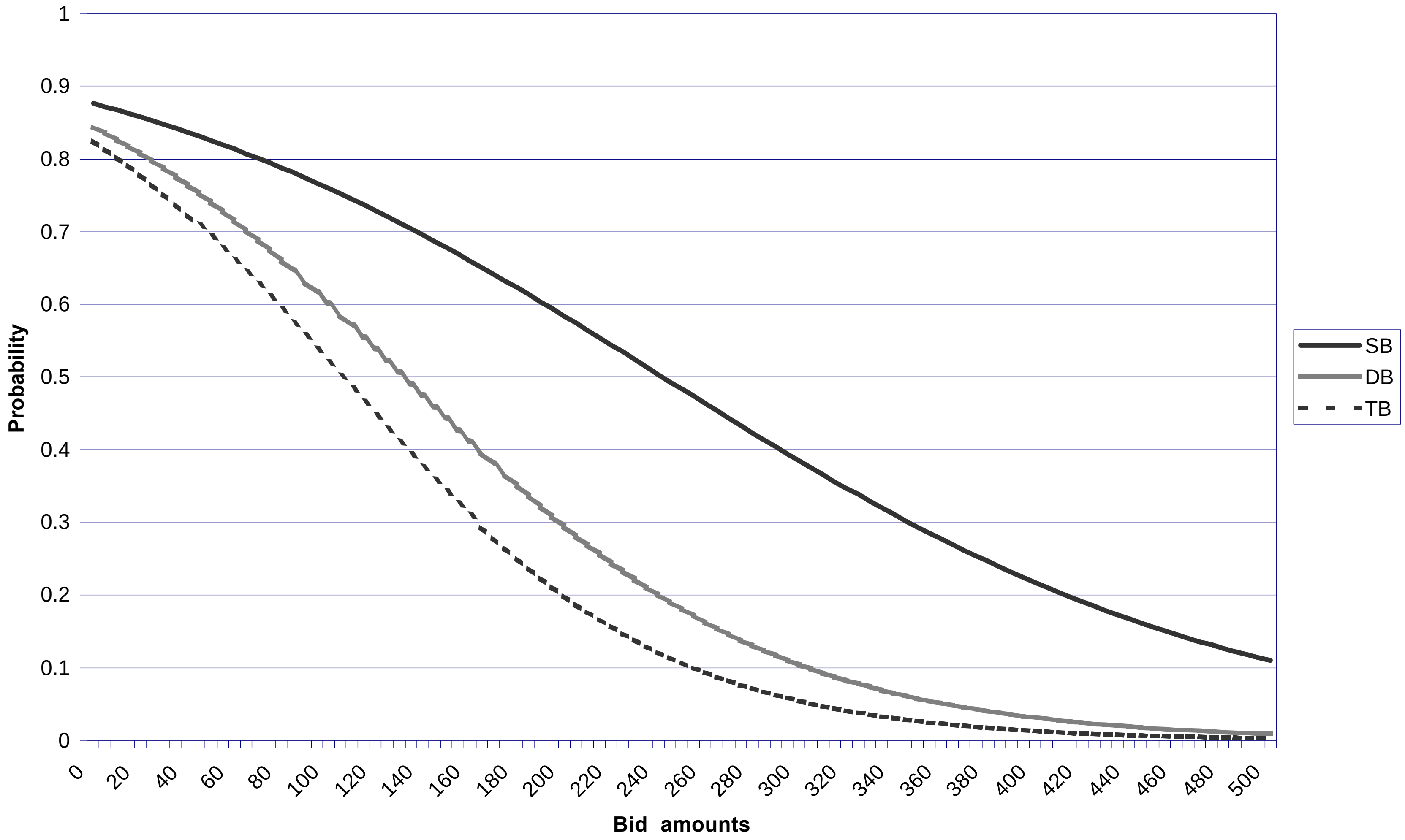
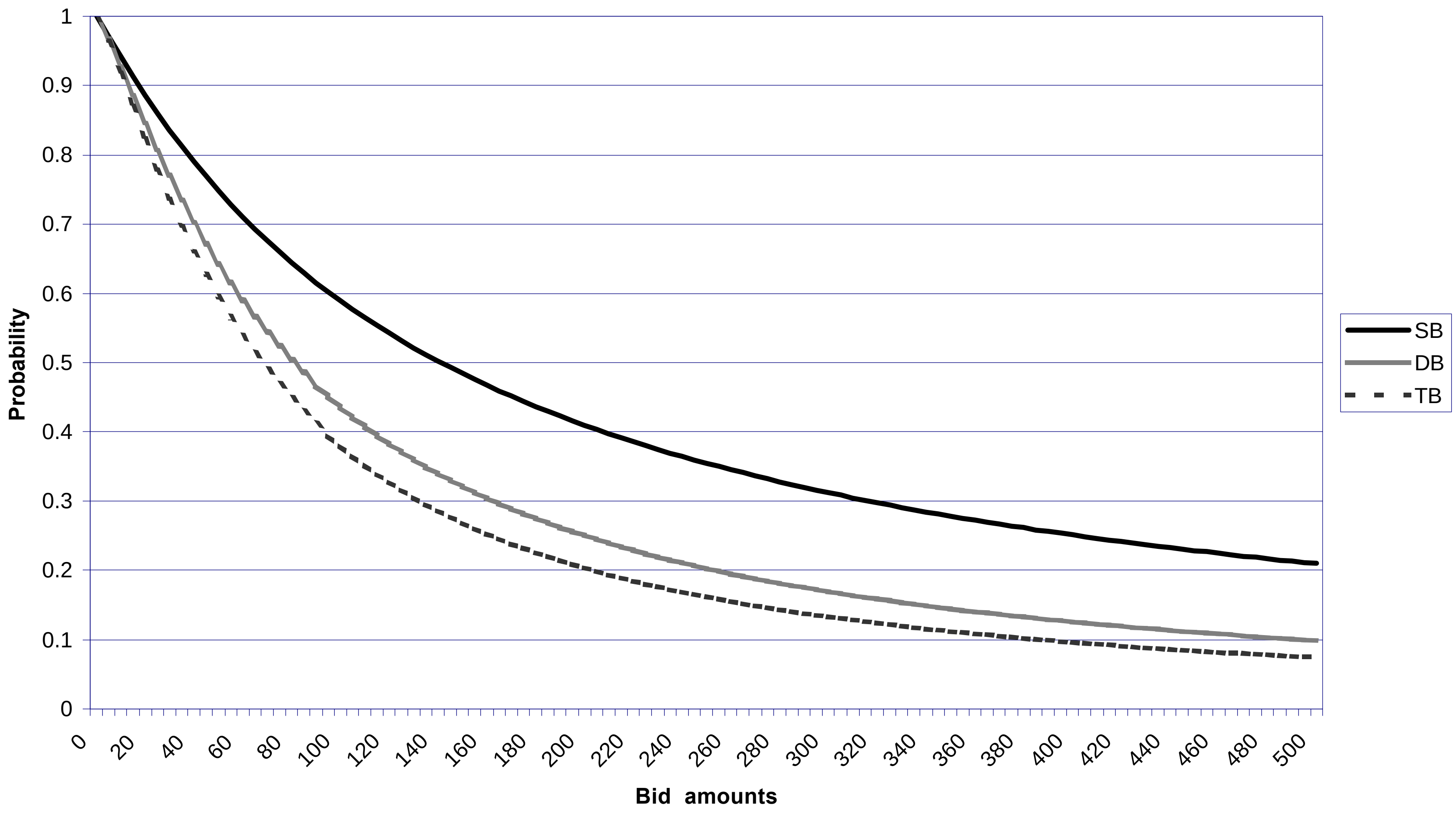


Figure 2: Estimated probability of a positive response from log-logistic models



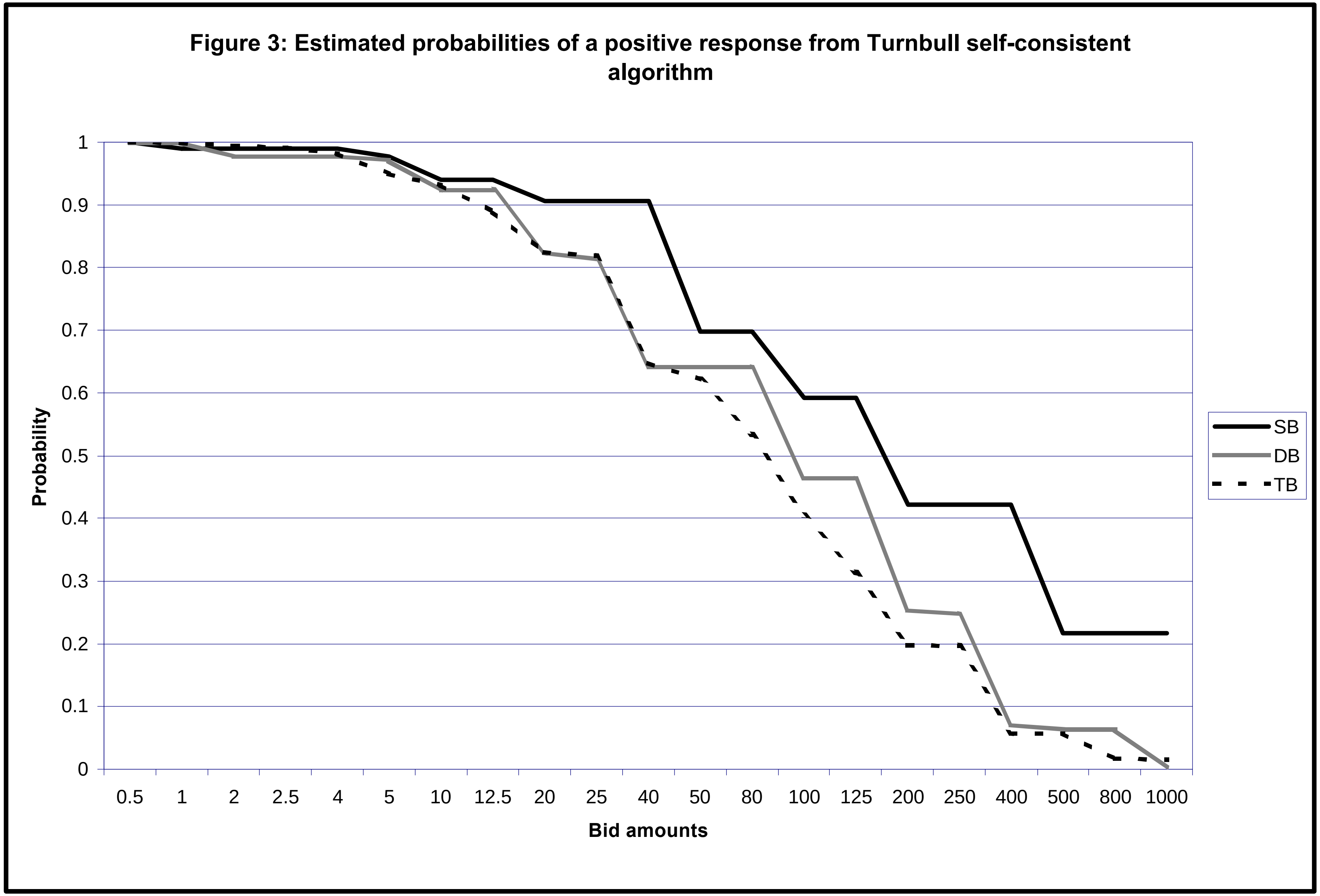


Figure 3: Estimated probabilities of a positive response from Turnbull self-consistent algorithm
Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.5
1
2
2.5
4
5
10
12.5
20
25
40
50
80
100
125
200
250
400
500
800
1000
Bid amounts
SB
DB
TB