Bias and Efficiency of Single vs Double Bound Models for Contingent Valuation Studies: a Monte Carlo Analysis

PINUCCIA CALIA, Dpt. of Economics, University of Cagliari, calia@unica.it

ELISABETTA STRAZZERA, DRES and CRENoS, University of Cagliari, strazzera@unica.it

Keywords: Contingent Valuation, Single Bound Estimator, Double Bound Estimator, Bias, Statistical Efficiency.

JEL: C15, C35, D60, H41, Q26.

We would like to thank Trudy Ann Cameron for helpful suggestions, and Rossella Diana for assistance in the early stage of this research.

SUMMARY

The *Dichotomous Choice Contingent Valuation Method* can be used either in the single or double bound formulation. The former is easier to implement, while the latter is known to be more efficient. We analyse the bias of the ML estimates produced by either model, and the gain in efficiency associated to the double bound model, in different experimental settings. We find that there are no relevant differences in point estimates given by the two models, even for small sample size, and no estimator can be said to be less biased than the other. The greater efficiency of the double bound is confirmed, although differences tend to reduce by increasing the sample size. Provided that a reliable *pre-test* is conducted, and the sample size is large, use of the single rather than the double bound model is warranted.

Address for correspondence*:* E. Strazzera, DRES, *University of Cagliari, Via Fra Ignazio 78, I-09123 Cagliari, Italy. E-mail: strazzera@unica.it*

NON TECHNICAL SUMMARY

The *Dichotomous Choice Contingent Valuation Method* (DC-CVM) is increasingly used as a method to value nonmarket goods. One of the reasons is that it was explicitly recommended by the panel of experts, chaired by Nobel laureates Kenneth Arrow and Robert Solow, appointed by the *National Oceanic and Atmospheric Administration (NOAA)* of the United States to assess the validity of the contingent valuation method. The value that people give to some public good is elicited by asking them if they would be willing to pay a given amount for its provision: the individual has just to answer YES or NO. It is like a market situation, where for each good the price is given, and consumers decide whether to accept it or not.

The method is used in two variants: the *single bound,* if only one question is posed to each individual; and the *double bound,* where a second bid is offered, higher than the first if the answer was positive, and lower otherwise.

Either version has its own advantages and disadvantages. For any given sample size, survey costs tend to be higher for the double bound model, since the interactive procedure requires that the interview is made on the spot, either face-to-face or over the telephone. Also, the double bound method can be affected by bias in responses that are due to the introduction of the follow-up. On the other hand, the single bound provides less information than the double bound model, and produces less precise estimates for the willingness to pay (*wtp*).

Some authors have argued that point estimates from the double bound are also less biased than those produced by the single bound: this contention has not been proved, though, and the opposite might be true. If we take into account another recommendation of the NOAA panel, i.e. that the analyst should prefer more prudent estimates, it might be safer to take the estimate with a larger confidence interval, unless we are sure that the more precise estimate is also less biased.

The discrepancy between the estimates produced by the single and the double bound method has been extensively discussed in the literature, but, as far as we know, no simulation study has been conducted to assess gains or losses in precision and unbiasedness from using either model for contingent valuation. This is the scope of our paper: we carry on a Monte Carlo analysis to compare the statistical performance of the two estimators under different experimental situations. Our results confirm the theoretical findings about the efficiency of the double bound model: it produces more precise point estimates of parameters and central tendency measures of wtp, with narrower confidence intervals around mean or median wtp. Differences though tend to reduce by increasing the sample size, and are often negligible for medium size samples. On the contrary no relevant differences can be found in point estimates of parameters and central tendency measures between the two models, even for small sample size, and no estimator can be said to be less biased than the other. Our results warrant the use of the single bound model whenever the sample size is large enough, and a pre-test conducted on a small population sample is thought to give a good *a priori* for the bid design of the survey. If instead the sample size is very small, or the *pre-test* survey is not much reliable, it is advisable to use the double bound model: in these circumstances the gain in efficiency is so large that indeed may overwhelm other possible costs associated to the use of the double bound model.

1. Introduction

The *Dichotomous Choice Contingent Valuation Method* (DC-CVM) has been in the last years the most popular technique among practitioners of contingent valuation. One of the reasons is that it was explicitly recommended by the panel of experts, chaired by Nobel laureates Kenneth Arrow and Robert Solow, that was appointed by the *National Oceanic and Atmospheric Administration (NOAA)* of the United States to assess the validity of the contingent valuation method. The value that people give to some public good is elicited by asking them if they would be willing to pay a given amount (*bid*) for its provision: the individual has just to answer YES or NO. It is like a market situation, where for each good the price is given, and consumers choose whether to accept it or not. Since people are familiar with this valuation procedure, response distortions should be reduced to a minimum.

The method is used in two variants: the *single bound,* if only one question is posed to each individual; and the *double bound,* where a second bid is offered, higher than the first if the answer was positive, and lower otherwise.

Either model has its own advantages and disadvantages. For any given sample size, survey costs probably tend to be higher for the double bound model, since the interactive procedure requires that the interview is made on the spot, either face-to-face or over the telephone. If a specific member of the household is the target of the interview (for example, the head of the household), contacts may be difficult and expensive, both in terms of time and money. Furthermore, Herriges and Shogren (1996) found that the response rate decreases when follow-up questions are introduced in the survey. It might be possible, they argue, that "*the additional complexity of the questionnaire may discourage survey response, directly reducing* *the efficiency gains from follow-up questioning and increasing the potential for nonresponse bias"* (Herriges and Shogren, cit., p.130). In addition, lack of time to think might have an impact on the validity of the answers obtained through the double bound process. In an experimental study about willingness to pay for water services in Nigeria, Whittington et al. (1992) found that giving respondents time to think had a clear influence on their answers, producing consistently lower estimates. As the same authors point out, these findings do not necessarily transfer to developed economies (given the substantial differences in education and demographic characteristics). Yet, there is a concrete possibility that some "yea-effect" is produced when the bid question requires that an answer is given on the spot. It would be probably more reasonable to allow individuals to take a price, think about it, and then decide - just as they usually do before buying, say, an appliance or other kind of durables.

From this point of view, the single bound model would be more suitable. Unlike the double bound approach, the single bound option allows mailing of the questionnaires together with the relevant informative material. Respondents can take their time to answer, which should help to decrease the nonresponse rate: subjects read and fill the questionnaire at their own convenience, and then can leave it to be read on the phone by any member of the household.

Yet, the double bound seems now preferred by CV analysts to the single bound method. The main reason is that the double bound DC-CVM is asymptotically more efficient than the single bound model, as proved by Hanemann, Loomis and Kanninen (1991). Empirical applications confirm this property also for finite samples.

Granted that confidence intervals are larger for the single bound, it still remains to be seen if point estimates from the double bound are also less biased than those produced by the single bound. If we take into account another recommendation of the NOAA panel, i.e. that the analyst should prefer more prudent estimates, it might be safer to take the estimate with a larger confidence interval, unless we are sure that the more precise estimate is also less biased. Indeed, point estimates for the central tendency measures of *wtp* produced by the two models are quite different (cfr. Hanemann, Loomis and Kanninen (1991), McFadden and Leonard (1993), León (1995), Herriges and Shrogren (cit.)). In some cases the difference between the measures produced by the two estimators has been attributed to distortions brought into the data by the follow-up question. In the aforementioned study, Herriges and Shrogren investigate the existence of an anchoring effect caused by the first bid, and conclude that it affects, at least in part, the estimates. Controlling for the anchoring effect resulted in a significant reduction of the efficiency gains from the follow-up question. Other sources of disturbance on the data arising from the follow-up question are analysed by Cameron and Quiggin (1994) and Alberini, Kanninen and Carson (1997), that propose different econometric specifications to correct for these flaws[1].

Alternatively, the discrepancies between the single bound and the double bound estimates can be interpreted in the sense that the double bound model produces not only more efficient but also less biased estimates than the single bound. In fact, Hanemann, Loomis and Kanninen (1991) suggest that since the double bound model allows for correction of a poor choice of the initial vector of bids, it should also produce less biased estimates. The same contention is also purported by Kanninen (1995): with real data and assuming that the *wtp* distribution of the population is a Logistic, she calculates the bias of the double and the single model estimates,

[1] Alberini (1995) compares the performance of the bivariate probit (suggested by Cameron and Quiggin) and the univariate probit for the double bound model.

finding out that the latter is larger. This can be hardly thought to be a definitive answer, though, given the small sample (100 observations) considered in her study, and, more fundamentally, that her assumption about the true *wtp* model might have been incorrect. However, if the hypothesis is correct, the superiority of the double bound method in terms of the statistical properties of the estimator would be very strong indeed.

Some more research should be done to investigate on the properties of efficiency <u>and</u> unbiasedness of the two estimators. If the double bound model does not produce substantial gains when both criteria are taken into account, use of the single bound model may be warranted, especially if we consider the aforementioned drawbacks of the double bound model.

A Monte Carlo analysis conducted by Kling (1997) using a travel cost model combined with a contingent valuation model does not confirm the cited hypothesis of less bias of the double bound model, at least when combined with travel cost data, since mixed results are obtained. We are not aware, however, of any simulation study specifically aimed to assess gains in precision and unbiasedness from using the double bound model rather than the single bound when only contingent valuation data are considered.

This is the scope of the present study. In order to consider only the econometric performance of the two estimators, we generate a "clean" dataset, assuming no response bias in the follow-up question. The performance of either estimator is then analysed under different experimental situations.

After a quick overview of the two models (section 2), we present in the following sections the experimental setting (section 3) and the results (section 4); section 5 concludes the paper.

2. The wtp models

We adopt the censored econometric model proposed by Cameron and James (1987) and Cameron (1988), which, unlike the utility differential model of Hanemann (1984), produces separate estimates for the standard deviation of the *wtp* and the parameters of the model. This allows us to easily compute the confidence intervals for the central tendency measures of *wtp*: as described later, estimates of the standard errors of the coefficients are directly plugged in an analytical formula. It is worth to mention that only recently confidence intervals (either derived through analytical calculus or through bootstrapping methods[2]) for the *wtp* estimate are being included in contingent valuation studies[3].

Assuming a linear functional form for the *wtp*, the econometric model is the following:

$$(2.1) \qquad Y_i = x_i'\beta + \varepsilon_i$$

where Y_i is the true individual willingness to pay, which is assumed to depend on individual socioeconomic characteristics contained in the vector x_i. The error term ε_i is distributed with c.d.f. $F(\varepsilon_i)$ with zero mean and variance equal to v^2. In this model Y_i is considered a latent continuous censored variable: the observed variable is the answer YES or NO to the question regarding whether or not the individual would be willing to pay a given amount t_i.

For a given sample of *n* independent observation, the log-likelihood function is:

[2] For a comparison between different methods cfr. Cooper (1994).

[3] Cfr. Cameron (1991) and Park, Loomis and Creel (1991).

(2.2) $$LogL = \sum_{i=1}^{n} \{ I_i \log [1 - F((t_i - x_i'\beta)/v)] + (1 - I_i) \log [F((t_i - x_i'\beta)/v)] \},$$

where I_i is a dummy variable assuming value one if the answer is positive, zero otherwise. Since $1/v$ is the coefficient of the bid t_i and bids are varied among individuals, β and v can be estimated separately, so we have a direct estimate of the standard deviation of *wtp*.

When the double bound model is chosen instead, we observe two dichotomous variables, i.e. the answers to the first question and its follow up.

The log-likelihood function for the double-bound is:

(2.3) $$\begin{aligned} LogL = \sum_{i=1}^{n} \{ & I_i \, I_i^u \log [F((t_i^u - x_i'\beta)/v)] + \\ & + I_i (1 - I_i^u) \log [F((t_i^u - x_i'\beta)/v) - F((t_i - x_i'\beta)/v)] + \\ & + I_i^l (1 - I_i) \log [F((t_i - x_i'\beta)/v) - F((t_i^l - x_i'\beta)/v)] + \\ & + (1 - I_i)(1 - I_i^l) \log [F((t_i^l - x_i'\beta)/v)] \} \end{aligned}$$

Here t_i stays for the bid offered in the first question; t_i^u is the follow up if the answer to the first question has been positive; t_i^l is the follow up when the answer to the first question has been negative. I_i, I_i^u, I_i^l are dichotomous variables with value one if the answer to the first bid or the corresponding follow-up has been positive, and zero otherwise.

Once the parameters of either model are estimated, through Maximum Likelihood procedure, estimation of the mean *wtp* is straightforward: it suffices to calculate

(2.4) $$E(Y) = \bar{x}'\hat{\beta}$$

where $\bar{x}$ is the vector of sample averages of the regressors and $\hat{\beta}$ is the vector of ML estimates of the parameters. Another measure of interest in contingent valuation studies, especially when the *wtp* distribution is asymmetric, is the estimate of the median *wtp*, whose analytical form depends on the *wtp* distribution.

It is useful to calculate also confidence intervals for the mean or median *wtp*. Only recently researchers have begun to include confidence intervals in their reported fitted *wtp* measures, either using refinements of the bootstrap method (Krinsky and Robb (1986); McLeod and Bergland (1989)) or using the analytical formula proposed by Cameron (1991). For the model in eq. (2.1), Cameron demonstrated that an interval for $E(Y)$ at significance level α can be calculated as follows:

(2.5) $$CI_{1-\alpha}\left[E(Y)\right] = \bar{x}'\hat{\beta} \pm t_{\alpha/2}\sqrt{\bar{x}'\Sigma_{\beta}\bar{x}}$$

where Σ_{β} is an estimate of the variance-covariance matrix of the parameter estimates. In a paper by Cooper (1994) it is shown that either method to calculate confidence intervals performs quite well, the relative ranking depending on sample size and specification of the *wtp* model. Given the simplicity of Cameron's method, we use her analytical formula to calculate confidence intervals for the fitted mean (median) values of *wtp* in our experiments.

3. The Monte Carlo Study

We consider two specifications for the *wtp* among the most commonly used in CVM studies. The first one is a linear equation for the latent variable:

(3.1) $$Y = x'\beta + \varepsilon$$

The second specification is a logarithmic function:

(3.2) $$\ln Y = \lambda + \delta' \ln x + \eta$$

The variables ε and η are error terms with zero mean and variance σ^2 and τ^2 respectively. In designing the *Monte Carlo* analysis we assume, for specification (3.1), that *wtp* has two different distributions with mean $E(Y) = x'\beta$ and variance equal to σ^2: the first is Normal, the second is a mixture of two Normal which resembles an asymmetric distribution.

For specification (3.2), *wtp* is assumed to have a lognormal distribution (so that the error term η has a Normal distribution) with mean $E(Y) = exp(\lambda + \delta' \ln x + 0.5\tau^2)$, median $M(Y) = exp(\lambda + \delta' \ln x)$ and variance:

$$V(Y) = exp(2(\lambda + \delta' \ln x + \tau^2)) - exp(2(\lambda + \delta' \ln x) + \tau^2).$$

This specification is particularly suited to account for asymmetries in the *wtp* distribution, often observed in real data.

For each specification we generate 200 samples with four different size: 100, 250, 400 and 1000 observations.

3.1. The linear specification

The *wtp* data is generated according to the model:

(3.1.1) $$Y_i = \alpha + \beta\, x_i + \varepsilon_i$$

with α = 20, β = 0.1 and values of the regressor x drawn from a Uniform distribution in the range 40-750. In a first set of experiments, the error term ε is a Normal variable with zero mean and standard deviation σ = 10. In another experiment, we consider a situation where the error term has a mixture distribution:

(3.1.2) $$f(\varepsilon) = p\, f_1(\varepsilon) + (1-p) f_2(\varepsilon)$$

where $f_1(\varepsilon) = N(\mu_1, \sigma_1^2)$ and $f_2(\varepsilon) = N(\mu_2, \sigma_2^2)$. By setting different values of p, μ_1, μ_2, σ_1^2 and σ_2^2, $f(\varepsilon)$ is allowed to assume different forms (either symmetric or asymmetric, unimodal or bimodal). We choose p = 0.4 , μ_1 = -8, μ_2 = 5.33, σ_1 = 3 and σ_2 = 15.22, in order ε to have mean equal to zero and standard deviation equal to 10; besides, the resulting distribution is almost bimodal and asymmetric with a heavy right tail (Figure 1).

******Insert fig. 1 here**

We report results for two bid designs: in the first (bid design A), the chosen bids are the quartiles of the *wtp* empirical distribution in a small independent sample (50 observations), simulated through eq. (3.1.1) with no error term. In the second experiment instead (bid design B), three bid values (10, 20, 30) are selected such that only the left tail of the *wtp* distribution is covered (less than 15% for both *wtp* distributions).

In both experiments, the selected values are then randomly assigned to the individuals of the sample and compared with the corresponding Y_i in order to create the dichotomous variable for the first answer. The follow-ups required for the double bound model are obtained from the first bid by increasing or decreasing it by 25% of its amount: whenever the first bid is lower than Y_i, the bid is reduced; otherwise, it is increased. The dichotomous dependent variable, I_i assumes value zero if the true *wtp* is lower or equal to the assigned bid; otherwise it assumes value one. For the double bound model, we have two dependent variables: the first is generated, as before, by comparing each Y_i to the assigned first bid; the second is obtained analogously, matching Y_i with the second bid.

A Gauss-386i Aptech Maximum Likelihood routine is employed to maximise the log-likelihood functions (2.4) and (2.6). The Normal c.d.f. Φ is plugged into the log-likelihood functions in place of the generic c.d.f. *F*: the model is therefore correctly specified in both the deterministic and the stochastic part when the *wtp* distribution is normal, while we are allowing for misspecification in the stochastic part of the model when the *wtp* distribution is asymmetric.

3.2. The loglinear specification

The *wtp* data for the loglinear model is simulated according to the following equation:

(3.2.1) $$\ln Y_i = \lambda + \delta \ln x_i + \eta_i$$

where l = 1.05, d = 0.35 and *x* is a Uniform regressor in the range 2500-125000[4].

The disturbance h is simulated from a Normal distribution with mean zero and standard deviation t = 1.48.

The bids are selected as percentiles (5th, 10th, 20th, 45th, 75th, 95th) of the *wtp* (obtained as exp(ln (*Y*))) empirical distribution in a small independent sample. Analogously to the experiment with the linear model, the follow-up bid is created by increasing, or decreasing, the first bid by 25% of its amount. The dichotomous dependent variables are then created by comparing ln(Y_i) to the logarithm of the first bid and, sequentially, to that of the appropriate follow up.

For this experiment we assume a correct specification of the model, so that for estimation of the single and the double bound models the normal c.d.f. F is substituted for *F* in the log-likelihood equation (2.4) and (2.6) respectively.

Given the asymmetric shape of the *wtp* distribution generated by eq. (3.2.1), the median rather than the mean value can be indicated as an appropriate measure of central tendency. In such a case the calculus of the confidence intervals follows two steps: in the first step we calculate the limits of the interval around *E* (ln(*Y*)); then, we transform these values by taking the anti-log. This is a correct confidence interval for the median (cfr. Greene (1991, pag.168) for OLS estimates and Cameron (1991) for ML estimates of the loglinear model parameters); the results, though, are not entirely satisfactorily, as it will be seen in the next section.

4 These values for the parameters and the regressor are taken from Jordan and Elnagheeb (1994).

4. Results

The results of our experiments are summarised in tables 1 through 5. All experiments confirm that the double bound model is more efficient than the single bound: the standard deviations of the estimates from the double bound are always smaller than those obtained from the single bound. It should be noted, though, that differences in efficiency are especially relevant for the small sample size (100 observations). Point estimates from both models get more precise when the number of observations increases, and for small-medium size samples (250 obs.) and larger, the differences in precision of the two estimators are often negligible.

Results about the bias of the estimates obtained from the two models instead are not so clear cut. The central tendency measures are in some cases estimated more accurately by the single bound model, even though the opposite holds more often. Anyway, as we can see from the results in the following tables, there are no substantial differences in bias for the relevant measure of *wtp* between the two models.

More remarkable differences can be found in the estimates of confidence intervals: as it can be expected, the double bound model gives narrower intervals (about half the length of corresponding interval of the single model). As a consequence of this, and since the bias of the estimated mean or median *wtp* for the two models is quite similar, the double bound model produces also intervals with lower empirical confidence level in almost all experiments. It can be noticed that in general, for the smallest dimension, the estimated confidence intervals are not much reliable: empirical levels close to the nominal are associated to wide intervals. This problem becomes less serious for larger samples, where we find narrower intervals and empirical levels closer to the nominal confidence level of 90%.

A comparison of table 1 with table 2 shows that a wrong bid design (design B) affects to some extent the performance of both models: estimates are more biased and less precise, in particular for small sample size. Especially severe is the increase in the standard deviation of mean *wtp* estimates, which is reflected also in the marked increase, for small sample size, of the width of the confidence intervals.

******Insert table 1 here**

******Insert table 2 here**

Table 3 shows the results from the experiment where we consider a possible misspecification of the econometric model: we assume that the *wtp* is normally distributed while instead it is not.

It can be noticed that the two models are quite robust to misspecification, giving, in general, good point estimates of the parameters and mean *wtp*. The exception is the estimate of the standard deviation of *wtp*, which is always overestimated by both models for all sample sizes. Anyway, comparison with the results reported in table 1 shows that misspecification affects in particular the precision of estimates, resulting in higher standard deviations.

******Insert table 3 here**

The results of the experiment with the asymmetric distribution and bid design B are reported in table 4. When misspecification and bad bid design combine, the optimisation algorithm fails to converge several times, particularly for the smallest sample. In this case we also found that abnormal values for point estimates of the parameters are produced in many replications by the single bound model, while the double bound is more robust. In calculating the summary statistics, the replications with such abnormal values are dropped off the sample.

******Insert table 4 here**

It is quite clear that for this experimental design the double bound performs better. Especially for the smallest sample size, the double bound secures a relevant gain in efficiency, while, as usual, differences tend to decrease when working with more observations.

This effect can also be observed by looking at the average width of the confidence intervals: the proportion of the single bound interval width with respect to the corresponding double bound interval is about 3.5 for the sample size of 100 observations, and falls to 2 for greater sample sizes. It is also interesting that for this experiment design the confidence levels associated to the double bound intervals are always better than the single bound.

Differences in bias instead are not so significant. Taking into account the misspecification and the very poor bid design, we can say that both estimators perform reasonably well in giving point estimates for the parameters and the mean wtp, at least for sample sizes 250 and over. We signal that in this experiment some replications are dropped off, either because of non-convergence, or because abnormal values were produced. This introduces some sampling variability, as a different number of replications are dropped off for each sample size.

Finally, table 5 reports the results of the experiment with the loglinear specification. Notwithstanding the correct specification and good bid design (such that most of the *wtp* distribution is covered) we observe large bias and standard deviation values, in particular if compared with the analogous experimental design for the linear model. This casts some doubt about the widespread use of log specifications in contingent valuation studies: other transformations (e.g. Box-Cox) might be more suitable[5]. As usual, moving from a sample size of 100 to a sample size of 250 has a dramatic effect on the performance of both estimators.

Our application of Cameron's analytical formula to the loglinear model is not very satisfying, presenting extremely large average widths. This result is in line with Cooper's (1994) finding that when the distribution is asymmetric Cameron's technique is not much reliable, and bootstrap methods for calculating confidence intervals should be preferred.

******Insert table 5 here**

5. Conclusion

The single bound method presents some attractive features with respect to the double bound. It requires less information, it is easier to implement at data collection and estimation stages, and avoids systematic bias in responses that are due to the introduction of the follow-up (for example, the so called "anchoring effect"). On the other hand, it is well known that the double bound is more efficient than the single bound estimator. It is therefore interesting to compare their behaviour in terms of bias of the ML estimates produced by either model, and to analyse the gain in efficiency associated to the double bound model, in different experimental settings.

Our results confirm the theoretical findings about the efficiency of the double bound model. It produces more precise point estimates of parameters and central tendency measures of *wtp*, as

[5] As suggested by Trudy Ann Cameron in a private communication.

well as narrower confidence intervals around mean or median *wtp*. The differences, though, tend to reduce by increasing the sample size, and are often negligible even for small-medium size samples. On the contrary, no clear-cut results are obtained for the point estimates given by the two models, even for small sample size, so that neither estimator can be said to be less biased than the other.

Granted that no other sources of systematic bias arise, and the sample size is large enough, huge differences in point estimates between the two models observed in some applications should be ascribed to misspecification of the model, or poor bid design, or, more probably, both. Generally, Contingent Valuation surveys are preceded by a *pre-test* survey on a small population sample, that allows to gather information about the *wtp* distribution. If the pre-test is conducted correctly, it gives a good *a priori* for the bid design of the survey: in such a case, use of the single bound model should be warranted. If instead the sample size is small, or the *pre-test* survey is not much reliable, it is advisable to use the double bound model: in these circumstances the gain in efficiency is so large that may overwhelm indeed other possible costs associated to the use of the double bound.

References

Alberini, A. (1995) "Efficiency vs Bias of Willingness to Pay Estimates: Bivariate and Interval-Data Models". *Journal of Environmental Economics and Management* 29, pp. 169-80.

Alberini, A., B.J. Kanninen and R.T. Carson (1997) "Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data". *Land Economics,* forthcoming.

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

Cameron, T. A. (1991) "Interval Estimates of Non-Market Resource Values from Referendum *Land Economics* 67, pp. 413-21.

Cameron, T. A. and M.D. James (1987) "Efficient Estimation Methods for Closed-Ended Contingent Valuation Surveys". *The Review of Economics and Statistics* 69, pp. 269-76.

Cameron, T. A. and J. Quiggin (1994) "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-up' Questionnaire". *Journal of Environmental Economics and Management* 27, pp. 218-34.

Carson, R., W. Hanemann and R. Mitchell (1986) "Determining the Demand for Public Goods by Simulating Referendums at Different Tax Prices". *Department of Economics Working Papers,* University of California, San Diego.

Cooper, J. (1994) "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys". *Land Economics* 70, pp. 111-122.

Greene, W. H. (1991) *Econometric Analysis*. MacMillan Publishing Company;

Hanemann, W.M. (1984) "Welfare Evaluations in Contingent Valuation Experiments with *American Journal of Agricultural Economics* 66, pp. 332-41.

Hanemann, W. M., J. Loomis, and B.J. Kanninen (1991) "Statistical Efficiency of Double-bounded Dichotomous Choice Contingent Valuation". *American Journal of Agricultural Economics* 73, pp. 1255-63.

Herriges, J.A. and J.F. Shogren (1994) “Starting Point Bias in Dichotomous Choice Valuation *Journal of Environmental Economics and Management* 30, pp. 112-31.

Kanninen, B.J. (1995) “Bias in Discrete response Contingent Valuation”. *Journal of Environmental Economics and Management* 28, pp. 114-25.

Kling, C.L. (1997) "The Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods". *Land Economics* 73, pp. 428-39.

Krinsky, I. and A. Robb (1986) “Approximating the Statistical Properties of *Review of Economics and Statistics* 68, pp. 715-19.

Jordan, G.L. and A.H. Elnagheeb (1994) "Consequences of Using Different Question Formats in Contingent Valuation: A Monte Carlo Study". *Land Economics* 70, pp. 97-110.

León, C.J. (1995) “El metodo dicotomico de valoración contingente: una aplicación a los espacios naturales en Gran Canaria". *Investigaciones Economicas* 19.

McFadden, D. and G. Leonard (1993) "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis", in J. Hausman *(ed.) Contingent Valuation: a Critical Assessment,* North-Holland.

McLeod, D. and O. Bergland (1989) “The Use of Bootstrapping in Contingent Valuation W.P., Dpt. of Agricultural Economics. Oregon State University, Corvallis

Park, T., J.B. Loomis and M. Creel (1991) "Confidence Intervals for Evaluating Benefits Estimates form Dichotomous Choice Contingent Valuation Studies". *Land Economics* 67, pp. 64-73.

Whittington, D. V.Kerry Smith, A. Okorafor, A. Okore, Jin L. Liu, and A. McPhail (1992) "Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application", *Journal-of-Environmental-Economics-and-Management* 22(3), pp. 205-25.

Fig. 1. Density function of the mixture distribution

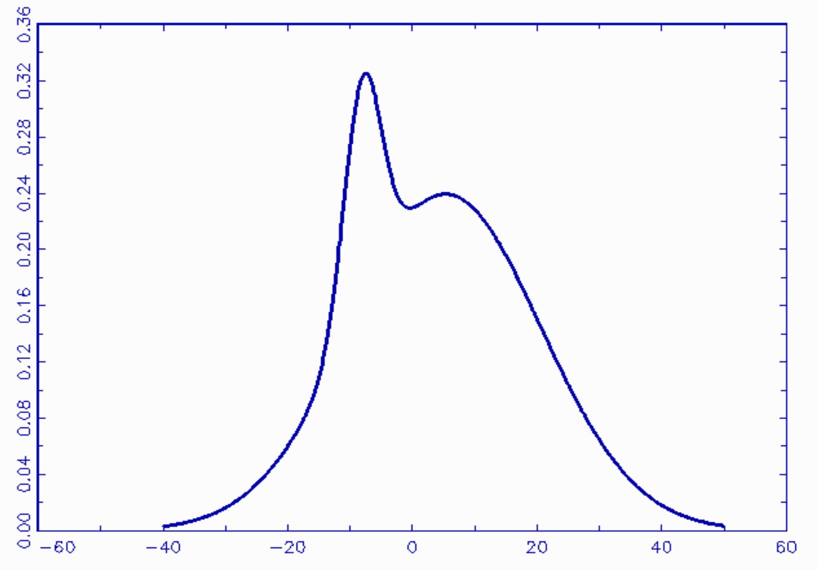

Table 1. Linear model (bid design A): summary statistics on estimated mean *wtp* across 200 replications

Estimates	Sample size			
	100	250	400	1000
E(Y) (59.5)				
Single	63.184 [a]	61.014	60.687	59.820
	(2.255) [b]	(1.181)	(1.034)	(0.641)
Double	63.228	61.002	60.649	59.860
	(1.476)	(0.842)	(0.663)	(0.473)
Bias (E(Y))				
Single	3.684	1.513	1.187	0.320
Double	3.727	1.501	1.149	0.360
Conf. Level				
Single	44.7 [c]	64.5	63.5	85.5
Double	18.1	47.0	48.5	78.0
Average width				
Single	6.976 [d]	4.211	3.335	2.115
Double	4.629	2.868	2.280	1.442

[a] Average of estimated mean *wtp* . [b] Standard deviation of estimated mean *wtp* . [c] Empirical confidence levels: percentage of inclusion of true mean *wtp* into the confidence intervals . [d] Mean difference between upper and lower limits.

Table 2. Linear model (bid design B): summary statistics on estimated mean *wtp* across 200 replications

Estimates	Sample size			
	100*	**250**	**400**	**1000**
E(Y) (59.5)				
Single	63.914 [a]	61.609	60.840	59.982
	(7.806) [b]	(3.776)	(2.461)	(1.516)
Double	63.648	61.114	60.731	59.898
	(3.545)	(1.850)	(1.440)	(0.899)
Bias (E(Y))				
Single	4.414	2.109	1.339	0.482
Double	4.148	1.614	1.231	0.398
Conf. Level				
Single	91.7 [c]	87.0	94.0	91.5
Double	76.5	81.5	84.0	88.0
Average width				
Single	22.453 [d]	11.795	8.830	5.363
Double	11.162	6.071	4.682	2.926

* Two replications giving abnormal values have been dropped off from the results of the single bound model.. [a] Average of estimated mean *wtp* . [b] Standard deviation of estimated mean *wtp* . [c] Empirical confidence levels: percentage of inclusion of true mean *wtp* in the confidence intervals . [d] Mean difference between upper and lower limits.

Table 3. Linear model (bid design A, asymmetric mixture distribution): summary statistics on estimated mean *wtp* across 200 replications

Estimates	Sample size			
	100	250	400	1000
E(Y) (59.5)				
Single	63.491[a]	60.974	60.609	59.924
	(2.641) [b]	(1.615)	(1.210)	(0.760)
Double	63.283	60.821	60.560	59.722
	(1.901)	(1.118)	(0.884)	(0.548)
Bias (E(Y))				
Single	3.991	1.474	1.109	0.424
Double	3.783	1.321	1.060	0.222
Conf. Level				
Single	51.5 [c]	73.5	77.5	85.5
Double	34.0	66.5	66.0	88.0
Average width				
Single	8.187 [d]	5.066	4.005	2.559
Double	5.820	3.668	2.898	1.843

[a] Average of estimated mean *wtp.* [b] Standard deviation of estimated mean *wtp* .[c] Empirical confidence levels: percentage of inclusion of true mean *wtp* in the confidence intervals . [d] Mean difference between upper and lower limits.

Table 4. Linear model (bid design B, asymmetric mixture distribution): summary statistics on estimated mean *wtp* across 200 replications

Estimates	Sample size			
	100 [i]	250 [ii]	400 [iii]	1000 [iv]
E(Y) (59.5)				
Single	69.038 [a]	61.307	59.490	57.063
	(37.742) [b]	(16.092)	(9.238)	(5.400)
Double	67.172	59.691	59.938	58.168
	(14.870)	(6.172)	(4.715)	(2.545)
Bias (E(Y))				
Single	9.537	1.807	-0.010	-2.437
Double	7.672	0.191	0.438	-1.332
Conf. Level				
Single	96.3 [c]	85.9	86.8	74.4
Double	93.2	89.3	89.8	84.5
Average width				
Single	163.55 [d]	44.83	30.44	16.41
Double	44.36	19.34	15.46	8.78

[i] 21 replications are deleted because of failure to convergence and 14 for the single bound and 3 for the double bound model due to abnormal parameter values.[ii] 3 replications are deleted because of failure to convergence and 5 for the single bound model due to abnormal parameter values.[iii] 1 replication is deleted because of failure to convergence and 2 for the single bound and 1 for the double bound model due to abnormal parameter values.[iv] 1 replication for the single bound model is deleted due to abnormal parameter values.

[a] Average of estimated mean *wtp*.[b] Standard deviation of estimated mean *wtp*. [c] Empirical confidence levels: percentage of inclusion of true mean *wtp* in the confidence intervals. [d] Mean difference between upper and lower limits.

Table 5. Loglinear model: summary statistics on estimated median *wtp* across 200 replications

Estimates	Sample size			
	100	250	400	1000
M(Y) (125.87)				
Single	132.525 [a]	127.850	128.401	126.155
	(31.230) [b]	(16.490)	(14.632)	(9.307)
Double	132.231	128.283	128.944	126.194
	(24.730)	(13.170)	(11.026)	(7.157)
Bias (M(Y))				
Single	6.655	1.980	2.531	0.285
Double	6.361	2.413	3.074	0.234
Conf. Level				
Single	89.0 [c]	94.9	89.4	87.5
Double	84.0	92.4	88.4	88.5
Average width				
Single	100.770 [d]	60.212	47.809	29.310
Double	77.553	46.668	37.379	22.898

[a] Average of estimated median *wtp*. [b] Standard deviation of estimated median *wtp.* [c] Empirical confidence levels: percentage of inclusion of true median *wtp* in the confidence intervals. [d] Mean difference between upper and lower limits.