Comparing Continuous and Discrete Contingent Valuation Questions*

BENGT KRISTRÖM
Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden

Abstract. This paper explores two commonly used methods to elicit an individual's willingness to pay (WTP) for a public good in contingent valuation studies. Currently, the most preferred method is the "take-it-or-leave" valuation question, or discrete valuation question (DVQ), where the respondent accepts or rejects a suggested cost for the good. The traditional method, the continuous valuation question (CVQ), simply asks an individual to state his WTP for the suggested change in the provision of a public good like cleaner air. We introduce a simple way to compare the results from these two methods. We also test the anchoring behavior suggested in the psychological literature on choice under uncertainty. The results do not support the anchoring hypothesis, but suggest the hypothesis that people perceive the two tested valuation questions differently.

Key words. Contingent valuation, valuation questions, anchoring, non-parametrics

Introduction

The contingent valuation approach to eliciting an individual's valuation of a public good like cleaner air has become very popular in recent years. At the heart of these experiments lies the valuation question, wherein the environmental good to be traded is described and the market for the good is established. Over the past decade we have seen a shift from the conventional continuous valuation question (CVQ) to the discrete valuation question (DVQ). These questions differ essentially by how much leeway the respondent is given when he is responding to the scenario presented. In the CVQ case he is asked to state his willingness to pay (WTP), while in the DVQ case he is "only" able to reject or accept a suggested amount. Researchers are increasingly using the DVQ in various forms. One of the main reasons why researchers prefer the DVQ method is its closer resemblance to our ordinary notion of "buying-not buying" a private market good. Also, Hoehn and Randall (1987) claims that the DVQ is less vulnerable to strategic behavior.

This paper explores whether subjects perceive the DVQ and the CVQ differently and sheds some light on the anchoring hypothesis. The paper expands on previous investigations in two ways. First, earlier investigations simply compare the estimated means, where the mean has been estimated using some specific distributional assumption in the DVQcase. This paper shows how comparisons can be done without invoking any specific distributional assumption. Second, our survey design permits a test of a simple

Environmental and Resource Economics **3**: 63–71, 1993.

version of the anchoring hypothesis. According to this hypothesis, developed largely by Kahnemann *et al.* (1982), a respondent will use any exogenous information given to him in an unusual valuation situation as a base. We test whether the suggested cost given in the DVQ affects the answers people provide in the CVQ by using a split-sample approach detailed below. Our results suggest that they do not.

The paper begins by briefly considering earlier results on the observed disparity between the CVQ and DVQ. We then define a measure of disparity between these two valuation questions (in terms of the responses received). The third section introduces our application and describes the experimental design. In section four we discuss the disparity test. The fifth section contains a test of the anchoring hypothesis.

Earlier Literature

The existing literature does not give a coherent picture as to the existence of a disparity (in the mean) between the DVQ and the CVQ. Sellar *et al.* (1985) present evidence implying that the DVQ generates higher estimate of mean WTP than the CVQ approach. In contrast, Kealy *et al.* (1988) find no significant difference. A similar conclusion is obtained by Bishop and Heberlein (undated). The above studies have exclusively used the (log)-logistic model in the discrete case in that the unobservable WTP is assumed to follow a log-logistic distribution. The tests for disparity are therefore only valid if that particular model happens to fit the data. The meta-analysis of Walsh *et al.* (1989), where they explored the results of a large number of contingent valuation experiments, provides some additional support for the hypothesis that the discrete valuation design may elicit higher mean values. They regressed reported mean values from hypothetical market experiments against a set of variables including the particular design of the studies. Their results suggests that the DVQ tend to give higher values compared to the CVQ (Walsh *et al.* (1989, Table 3, p. 189)).

It is important to recognize that economic theory offers no explanation as to why a disparity between the DVQ and CVQ may exist. Psychologists have developed the hypothesis of anchoring, which essentially implies that a respondent will tend to use any information given to him in an unusual valuation situation. We will test this anchoring hypothesis below. We begin by discussing the definition of differences.

Definitions of Differences

A natural way to define differences between two valuation question designs is the difference in the mean, although difference in median seems as plausible. Note that even though we might find that two distributions generate about the same mean, it is certainly not true that these two distributions are the

same. An individual may interpret the DVQ and CVQ very differently even though we obtain the same estimated means.[1]

One way to compare the designs is to draw the corresponding empirical survival distributions, using a non-parametric estimator in the discrete case. If we denote the empirical distribution functions by F_d and F_c for the discrete and continuous data, we can search for an appropriate measure of distance between these two distributions. We discuss this in more detail below.

Experimental Design

Our application is a study of the value Swedes place on preserving a selected set of forest areas in Sweden. A sample of 1 100 Swedes were sent a questionnaire in the spring of 1987.[2] The questionnaire contained a set of questions regarding household characteristics such as income and household size, and several attitudinal questions. In addition, the respondents were asked their WTP to preserve the suggested areas. Two types of valuation questions were asked, a CVQ and a DVQ.[3] There was a 67 percent response rate and a routine analysis of sample representativeness implied that our sample is a representative sub-sample of the Swedish households. For details, see Kriström (1990a).

The sample was split into two parts, sample *A* and *B* with 900 and 200 observations respectively. Sample *A* received both types of questions, sample *B* only the continuous question. The questionnaires were identical in all other respects. We received 599 questionnaires from sample *A*, and 133 from sample *B*. The response rates are almost exactly the same. Table I reports characteristics of sample *A* and *B*.

Table I. Characteristics of sample *A* (received both valuation questions) and sample *B* (received only continuous version)

	Mean *A*	Mean *B*	*t*-test	Mann-Whitney
Sex	0.52	0.52	0.03	N.R. ($\alpha = 0.98$)
Byear	43.2	41.6	1.12	N.R. ($\alpha = 0.24$)
House	2.7	2.7	−0.36	N.R. ($\alpha = 0.63$)
WTP	968.0	1209.0	−0.67	N.R. ($\alpha = 0.70$)
Visits	7.4	8.6	−0.65	N.R. ($\alpha = 0.32$)
Expect	2.8	2.7	1.62	N.R. ($\alpha = 0.10$)
Income	110.2E + 3	106.6E + 3	0.73	N.R. ($\alpha = 0.16$)

Notes: *t*-test computed without assuming equal variances. N.R. = cannot reject hypothesis of equal medians at $\alpha = 0.05$, where α is the significance level of the test. Sex = male (1), female (0); Byear = respondent's birthyear, House = Number of household members, WTP = willingness to pay, Visits = Number of visits made to the areas, Expect = expected number of visits to the areas in the next ten years (1 = at least 10, 2 = 5—10, 3 = 1—5, 4 = zero). Income = net household income.

Table I suggests there is no difference between the sample in terms of mean and medians of observable characteristics. The hypothesis of unequal means is rejected in all cases, and this is also true for the case of the median. These statistics lead us to claim that we have an ideal situation for performing the two-sample split test for disparity between the CVQ and the DVQ.[4]

Test of the CVQ vs. the DVQ

A test for the presence of a disparity between the DVQ and the CVQ can be easily developed. Recall that earlier investigations have employed a specific distributional assumptions for the DVQ and have simply compared the means. In this section we show how such a test can be performed without invoking any distributional assumption for the discrete response data. Our test is not restricted to a particular point of the distribution such as the mean. To obtain such a test it is natural to look for distance test of the Kolmogorov-Smirnov type. To use it we need to find the empirical survival function for the discrete response data. This can be done using a theorem of Ayer *et al.* (1955), as shown in Kriström (1990b).

Figure 1 shows the two empirical survival functions.

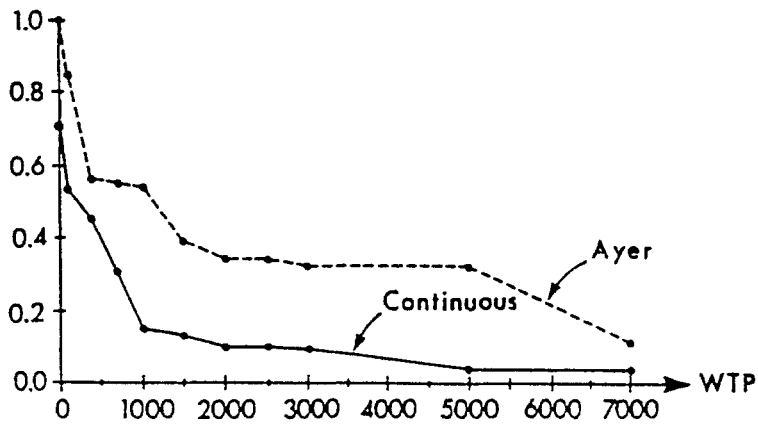

Fig. 1. Empirical survival functions for the discrete and the continuous valuation question.

It seems as if the discrete response question generates a distribution that is shifted outward compared to the distribution for the continuous data. The discrete response question will therefore produce a higher mean estimate as well as a higher estimate of the median. The largest difference is at the lowest suggested cost $A_i = 100$ ($i = 1, 2 \ldots 10$). About 47% of the respondents reported a WTP in the range zero to 100. In contrast 15% rejected the offer to pay 100 SEK. The difference between the distributions is smallest at a

cost of 400 when about 55% of the population claimed a WTP in the range zero to 400, while 44% rejected the cost in the discrete case. One interpretation of these result is that an individual who rejects to pay anything when asked for his WTP might accept to pay a small cost when presented with a low cost for preservation.

It is of some interest to develop a test for the hypothesis that the replies to both valuation questions are generated by the same distribution. One possible approach is to calculate the maximum distance between the empirical distribution functions and use a Kolmogorov-Smirnov type of statisic. This approach is complicated by the fact that the data are grouped. An appropriate Kolmogorov-Smirnov goodness-of-fit test with discrete and grouped data is developed by Petitt and Stephens (1977), but they do not give the appropriate values for the relevant number of degrees of freedom in our case (although a computer program is available from the authors). Therefore, we use a simpler approach, but are aware that distance-type test may be applied to compare discrete and continuous data. A simple chi-square test has been used instead. We compute the difference between expected and observed number of no-answerers in each group. For each A_t we compute the number of respondents having reported a WTP marginally less than A_t (they would have rejected A_t). We then use the proportion of no-answers to A_t to obtain the expected number of rejecters in the continuous case, assuming the respondents perceive the questions symmetrically. The relevant statistic is then

$$\sum (O_t - E_t)^2/E_t) = 113.71 \tag{1}$$

where O_t is the observed number of people having a WTP less than A_t and E_i is the expected value. This quantity may be compared with a chi-square distribution with nine degrees of freedom. The critical value is 16.9 at the 95% level. Hence, we reject the hypothesis that the DVQ and CVQ answers are generated from the same distribution. It might be argued that there is a bias towards accepting the nullhypothesis in this case. This is because subjects in sample A may have been more inclined to answer the CVQ and the DVQ "consistently". If this "in-sample" bias exists, then our conclusion is all the more powerful.

The Anchoring Hypothesis

Anchoring is a psychological concept developed largely by Kahnemann *et al.* (1982). According to the theory of anchoring behavior, a respondent will use any exogenous information given to him in an unusual valuation situation as a base. Anchoring might be viewed as a more general type of starting point bias, which has been analyzed to some extent in the literature. Starting point bias arises in a bidding game when respondents are sensitive to the starting

bid. One explanation of this phenomena might be that the individual views the starting bid as an implied value cue. The anchoring hypothesis tested in this paper is similar in spirit to the starting point bias issue, since we test whether or not the suggested cost functions as an implied value cue. If our version of the anchoring hypothesis is true, it may explain an eventual difference between WTP in sample *A* and *B*. This difference will, however, be closely tied to the design of the bid-vector, at least with the simple hypothesis that WTP is centered around the suggested bid. Mean WTP in sample *A* would then approach the mean of the bid-vector. The bid-vector used was (100, 400, 700, 1000, 1500, 2000, 2500, 3000, 5000, 7000), which has a mean of 2320. A *t*-test of the hypothesis that the mean of WTP is equal to 2320 in sample *A* gave a value of −12.38 (that is, a strong rejection of the hypothesis tested).

There is no a priori direction of anchoring bias. This is because anchoring could affect WTP in any direction depending on whether the suggested bid is higher or lower than the individuals "true" WTP. If anchoring exists, then a bid lower than WTP would tend to bias the answers downwards and vice versa. Accordingly, the magnitude of anchoring bias depends on the location of the bid-vector in the sample. A bid-vector placed in the left tail of the distribution of WTP would tend to have a downward effect on the responses to the CVQ. The hypothesis we want to test in light of this discussion is then whether the distribution of WTP in sample *A* is affected (relative to sample *B*) by the suggested cost. Table I provides rough evidence against such a hypothesis. First, we find no significant difference between mean WTP in sample *A* and *B*. Second, the non-parametric (Mann-Whitney) test of equal medians implies that the null hypothesis cannot be rejected. A general test of the anchoring hypothesis can be developed by computing the empirical survival functions for the continuous data on WTP in sample *A* and *B*. The empirical survival functions are reported in Figure 2.

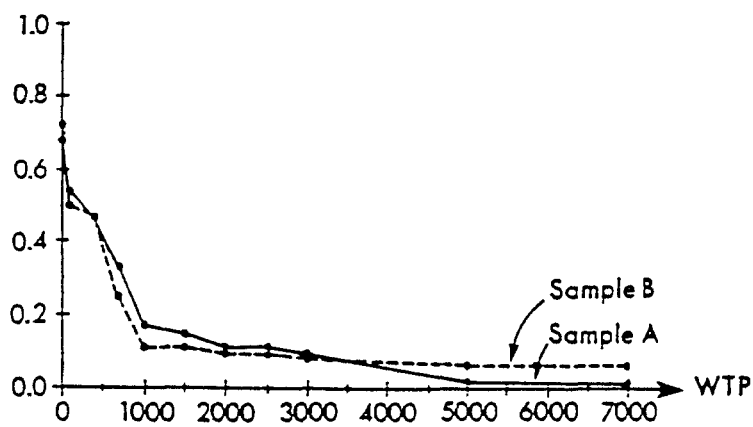

Fig. 2. Empirical survival functions for WTP in sample *A* and *B*.

Computing a Smirnov test, i.e. the maximum distance between the empirical distribution functions gave a value of 0.09. Using Table VII in Lindgren (1969, p. 487) we find that the critical value is about 0.14 at the 95% level. Hence, the hypothesis that $F_c^A = F_c^B$ is not rejected. We therefore conclude that the hypothesis of anchoring is not accepted.[5]

Final Remarks

This paper has explored two different ways of asking people to reveal how much they are willing to pay for a public good. The dataset provides evidence that people interpret different valuation questions differently. This finding is consistent with some earlier research on this issue. Earlier investigations have focused on a parameter of the distributions such as the mean. Our approach shows a simple way to obtain a more informative picture by computing the empirical survival distributions. We have also suggested a simple way to test for the anchoring hypothesis. Our data does not support this hypothesis, even though our "in-sample" comparisons are arguably biased in the direction of supporting the hypothesis.

An important question not answered in this paper is why there might be a disparity between different valuation question designs. This question can probably not be answered within conventional microeconomic theory. It is still an important question given the fact that the DVQ is currently a popular choice in contingent valuation studies. The important question why the DVQ should be preferred cannot be answered within the context of the analysis in this paper. It remains an issue for future research.

Finally, we note that the approach suggested in this paper extends naturally to analyses of other question formats chosen in a contingent valuation study. The question of anchoring is, of course, not limited to the simple referendum type question that we have analyzed in this paper. The suggested approach is easily extended to cover other question formats, such as the double-bounded. In this format, people are given an additional DVQ, where the new amount suggested depends on the reaction to the initial amount.

Appendix 1. The Willing-to-Pay Questions

The questionnaire contained an introductory letter (not reproduced here, see Kriström (1990a)) that explained why the investigation was made. It also gave some other background information. A map detailed the areas under investigation. The respondents were guaranteed anonymity. The valuation questions were introduced with a "box" that flagged the nature of the question to come and why it was asked. The WTP-question for sample A was as follows:

"There are many pristine woodlands, for example virgin forests and sub-

alpine woodlands, that may be of considerable interest to us and future generations. Other woodlands harbor animals, plants and birds threatened with extinction. Suppose we would abstain from forestry in the fragile areas marked on the map that you find at the end of this questionnaire. Would you be willing to contribute a once and for all sum of SEK . . . to cover the costs of preserving the areas? (yes/no).

How much would you maximally be willing to contribute? I would be willing to contribute a once-for-all sum of maximally about . . . SEK."

The subjects in sample *B* received an identical questionnaire, except that they did not receive a (yes/no) question.

Notes

* I would like to thank Prof. Glenn Harrison, University of South Carolina, Prof. P-O Johansson, Stockholm School of Economics, Prof. Jason Shogren, Iowa State University, and two referees for helpful comments. The usual disclaimer applies.

[1] Even if two distributions have the same moments, we cannot even show in general that they must be the same. For a proof, see Shiryayev (1984, p. 292).

[2] The questionnaire is available from the author on request.

[3] See Appendix 1. A stated WTP of zero was interpreted as a valid bid.

[4] As is shown below in Figure 1 and Figure 2, the distribution of WTP is markedly skewed in both samples. One may then want to use a non-parametric test instead to test differences in the mean. Note, however, that we have a relatively large number of observations. Therefore we can appeal to the central limit theorem. In his discussion of this theorem Lindgren (1969, p. 245) writes "Because of this dependency on the rate of approach to normality on the underlying distributions of the summands, it is not easy to specify the size of n for a good approximation. However, in most cases an n of 25 or more is adequate for approximations to two decimal places."

[5] It should be noted that the Kolmogorov-Smirnov test is known to have low power. See Fisz (1963).

References

Ayer, M., H. D. Brunk, G. M. Ewing, and E. Silverman (1955), 'An Empirical Distribution Function for Sampling with Incomplete Information', *Annals of Mathematical Statistics* **26**, 641—647.

Bishop, R. C. and T. A. Heberlein (Undated), 'The Contingent Valuation Method', Mimeo. Department of Agricultural Economics, University of Wisconsin, Madison.

Fisz, M. (1963), *Probability Theory and Mathematical Statistics*, Wiley, New York.

Hoehn, J. P. and A. Randall (1987), 'A Satisfactory Benefit Cost Indicator from Contingent Valuation', *Journal of Environmental Economics* **14**, 226—247.

Kahnemann, D., P. Slovic, and A. Tversky (1982), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.

Kealy, M. J., J. F. Dovidio, and M. L. Rochel (1988), 'Accuracy in Valuation Is a Matter of Degree', *Land Economics* **65**, 158—171.

Kriström, B. (1990a), 'Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis', Diss. Umeå economic studies, 219, 1990.

Kriström, B. (1990b), 'A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies', *Land Economics* **66**(2), 135—139.

Lindgren, B. W. (1969), *Statistical Theory*, MacMillan, Toronto.
Petitt, A. N. and M. A. Stevens (1977), 'The Kolmogorov-Smirnov Goodness-of-Fit Statistics for Discrete and Grouped Data', *Technometrics* **19**, 205–210.
Sellar, C., J. R. Stoll, and Chavas, J.-P. (1985), 'Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques', *Land Economics* **61**, 156–175.
Shiryayev, S. (1984), *Probability*, Springer-Verlag, New York.
Walsh, R. G., D. M. Johnson, and J. R. McKean (1989a), 'Issues in Nonmarket Valuation and Policy Application: A Retrospective Glance', *Western Journal of Agricultural Economics* **14**(1), 178–188.