# Incentive compatibility and procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method: Distinguishing strategic behaviour from the anchoring heuristic

Ian J. Bateman[1,*], Brett H. Day[1], Diane P. Dupont[2], and Stavros Georgiou[3]

1. Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia, Norwich, NR4 7TJ, United Kingdom.
2. Department of Economics, Brock University, 500 Glenridge Avenue, St. Catharines, Ontario, Canada, L2S 3A1.
3. Environment Policy Economics, Department for Environment, Food and Rural Affairs (DEFRA), London, UK.

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23-26, 2006.





**Acknowledgements:** Funding for this research was provided by the Commission of the European Community (CEC) through the CLIME project, Framework V Ref. No. EVK1-CT-2002-00121, by the Economics for the Environment Consultancy (EFTEC) and by the PEDM which is funded by the UK Economic and Social Research Council (ESRC). We would like to thank Graham Loomes and the participants at *Heartland Conference*, University of Iowa, September 2003, the *Envecon 2004: Applied Environmental Economics Conference*, Royal Society, London, March 2004, and the *Thirteenth Annual Conference of the European Association of Environmental and Resource Economists (EAERE)*, Institute of Environmental Sciences, Budapest University of Economic Sciences and Public Administration (BUESPA), Budapest, June 2004, and an anonymous commentator for their very helpful comments. Ian Bateman is also Adjunct Professor in Agricultural and Resource Economics at the University of Western Australia, Perth and is grateful to Lincoln University, Canterbury, New Zealand for hosting him while this paper was completed.

**Abstract**

The contingent valuation method for estimating willingness to pay for public goods typically adopts a single referendum question format which is statistically inefficient. As an alternative, Cooper, Hanemann and Signorello (2002) propose the 'one-and-one-half bound' (OOHB) format allowing researchers to question respondents about both a lower and higher limit upon project costs, thereby securing substantial statistical efficiency gains. These bounds are presented prior to the elicitation of responses thereby avoiding the negative 'surprise' induced by an unanticipated second question. However, this approach conflicts with the Gibbard-Satterthwaite result that only a single referendum format question is incentive compatible. The OOHB method may therefore be liable to strategic behaviour or reliance upon the anchoring heuristic observed in other repeated response elicitation formats. In a first formal test of the method we show that it fails crucial tests of procedural invariance and induces strategic behaviour amongst responses.

More than fifty years ago, Ciriacy-Wantrup (1947, 1952) suggested that “appropriately constructed interviews” are capable of obtaining information about people’s preferences for goods not ordinarily priced in the market. Since then the contingent valuation (CV) method has become the most widely used approach to obtain willingness-to-pay (WTP) values for an assortment of environmental and other non-market public goods (Carson, forthcoming). However, the issue of how WTP questions should be phrased and responses elicited has been and remains one of the most consistent themes in CV research (Cummings et al., 1986; Mitchell and Carson, 1989; Arrow et al., 1993; Bateman et al., 2002).

An ideal WTP elicitation method would have three defining characteristics: (i) incentive compatibility; (ii) statistical efficiency; and (iii) procedural invariance. Although all three criteria are important, it was the issue of incentive compatibility which dominated the landmark NOAA ‘Blue-Ribbon Panel’ report on CV (Arrow et al., 1993) and guided its endorsement of the ‘single bound’ (SB) dichotomous choice elicitation technique. Under the SB approach each survey respondent is presented with a single question asking whether or not they are willing to pay a specified sum (often referred to as the bid amount), $$X$, for the good in question. The sum $$X$ is varied across respondents allowing the analyst to estimate decision-relevant characteristics of the WTP distribution. Strategic behaviour arguments suggest that the dichotomous nature of SB responses makes the approach incentive compatible (Gibbard, 1973; Satterthwaite, 1975). However, compared to other elicitation methods, it is statistically inefficient, only determining whether a respondent’s WTP lies above or below the bid amount offered to them.

Hanemann, Loomis and Kanninen (1991) propose a solution to the inefficiency of the SB format in the form of their ‘double-bound’ (DB) elicitation method. Here, following an initial dichotomous choice question and response, a supplementary ‘follow-up’ question is added to further probe the respondents’ WTP. The elicitation of a second dichotomous choice response yields substantial gains in terms of statistical efficiency[1]. However, the DB approach fails the Gibbard-Satterthwaite criterion for incentive compatibility since the approach undermines the crucial face-value interpretation of the bid amount as being the cost of providing the good in question. While this is credible in the initial question, it is no longer so in the follow-up (Alberini et al., 1997; Carson, Groves and Machina, 2000). Indeed, researchers have noted that respondents are adversely ‘surprised’ by the unanticipated follow-up question to the extent that their first and second bound responses are inconsistent (McFadden, 1994; Cameron and Quiggen, 1994; Carson *et al*., 1994; Herriges and Shogren, 1996; Alberini, Kanninen and Carson, 1997; Bateman et al. 2001; DeShazo, 2002).

Cooper, Hanemann and Signorello (2002; hereafter CHS) propose the ‘one-and-one-half’ bound (OOHB) format “as a means to reduce the potential for response bias on the follow-up bid in multiple-bound discrete choice formats […] while maintaining much of the efficiency gains of the multiple-bound approach…” (p. 742). The OOHB format initially informs survey respondents that the costs of the good in question lie between some lower and higher amount (which we shall denote $*L* and $*H* respectively). This simultaneous presentation of costs is intended to avoid the adverse surprise effects of the unanticipated second question in the DB format. In order to elicit responses regarding each of these amounts, CHS employ a prior random process to assign respondents to either an ‘ascending’ or ‘descending’ presentation sequence. In an ascending sequence (which we shall denote LH), survey respondents are initially asked whether they are prepared to pay the lower cost $*L* for the good. Here a “No” response terminates the WTP elicitation process, while a “Yes” response results in the respondent being asked the second bound question regarding whether they would pay the higher amount $*H*. Conversely in the descending sequence (which we shall denote HL), the first bound question presents the higher amount $*H*. Here a positive response terminates the questioning process while a negative response results in the lower amount $*L* being presented at the second bound.

As CHS demonstrate this simple but highly innovative format yields much of the gains in statistical efficiency afforded by the DB approach without recourse to an unexpected follow-up question. This marks the OOHB format out as one of the most exciting prospects for the elicitation of WTP responses since the introduction of the SB method twenty-five years ago (Bishop and Heberlein, 1979). This potential is reflected in the rapid uptake of the method which has recently been used (or recommended for future use) in valuations of non-market goods for a number of government agencies in locations as diverse as North America, Europe, Asia and Africa (Barreiro, et al., 2005; Bateman et al., 2002[2]; Bradford, et al., 2004; Kerr, et al., 2004; NCEE and NCER, 2000; Scarpa, 2003; Signorello and Cooper, 2002; Sutton, et al., 2004).

Despite this potential and uptake a problem remains. While the OOHB may avoid the surprise associated with the follow-up question of the DB format, its reliance upon two prices still sets it at odds with the incentive compatibility requirements of the Gibbard-Satterthwaite criterion. This sets the scene for the present study. While CHS assert that they can find "no obvious vices" (p. 749) within their application, they acknowledge that their sample size is insufficient for rigorous testing (see p746, footnote 8; indeed much of their paper is taken up by comparing the OOHB to the heavily criticised DB approach; a comparison which given such criticism cannot of itself validate the former method). In contrast, incentive compatibility arguments would suggest that the OOHB provides respondents with ample opportunity to indulge in strategic behaviour and/or may induce framing anomalies or reliance upon the anchoring heuristic (DeShazo, 2002)

The central purpose of this paper is therefore to provide a full and rigorous examination of the procedural invariance properties of the OOHB using what is, to our knowledge, the first dataset designed and sufficient for such analytical purposes. Through such testing we seek to examine whether the incentive compatibility arguments against the OOHB do indeed result in anomalous responses, i.e. are theoretical expectations supported. In the following section we set out a suite of non-parametric, parametric and welfare measure tests are employed. In summary these consistently reject the null hypothesis of procedural invariance within the OOHB, thereby supporting the economic-theoretic expectation that the method's lack of incentive compatibility has resulted in either strategic behaviour (as suggested by standard theory) or in recourse to some anomalous response heuristic, most obviously anchoring (*ibid.*). Accordingly we extend Section II to define further tests with which to discriminate between these competing behavioural models. Section III reports details of the survey instrument design and implementation process while Section IV reports our focal analyses of procedural invariance and response behaviour. Section IV discusses implications and concludes.

**Testing For Procedural Invariance And Response Behaviour In The OOHB Format**

As noted above, the central purpose of this paper is to undertake tests of whether theory driven concerns regarding the incentive compatibility of the OOHB format result in a failure of procedural invariance within the method. In order to formulate such tests it is useful to develop some notation. Consider some given bid amount $\$X$. Respondents might encounter this bid amount in one of four possible arrangements. It may be the first question in an ascending sequence of questions, in which case we label it $\$X_{L1}$ (treatment LOW 1), or as the second question in a descending sequence in which case we label it $\$X_{L2}$ (treatment LOW 2). Likewise, if the bid amount is the high value of the pair it is labelled $\$X_{H1}$ when it is the first question in the descending sequence (treatment HIGH 1) or as $\$X_{H2}$ when it is the second question in the ascending sequence (treatment HIGH 2). Our testable hypotheses concern acceptance rates for $\$X$ under these four treatments. The probability of a respondent accepting a particular bid amount is $P\left[\$X = Yes\right]$ which we label simply as $P_X$. Likewise the acceptance rate for $\$X_{L1}$ is labelled $P_{X_{L1}}$ and the definitions of $P_{X_{L2}}$, $P_{X_{H1}}$ and $P_{X_{H2}}$ follow accordingly.

We adopt three approaches to procedural invariance testing, as follows:

(1) Nonparametric tests: comparing the proportion of respondents saying "yes" to a certain price under different treatments[3];

(2) Parametric tests: examining the coefficient estimates from parametric modelling of the WTP distribution under different treatments;

(3) Tests of mean WTP: comparing estimates of mean WTP derived from models of the WTP distribution under different treatments.

In all cases the null hypothesis of procedural invariance is that there will be no treatment effects. Our nonparametric tests formalise this as our first null hypothesis ($H_0$) in Table 1. Similarly our subsequent parametric testing examines the hypothesis that scale and location parameter for variables describing treatments LOW 1, LOW 2, HIGH 1 and HIGH 2 can be constrained to be equal. Furthermore, our examination of derived welfare measures examines whether each of these treatments yield estimates of mean WTP which are not statistically different. These three forms of testing are advocated by CHS. However, as noted previously, CHS acknowledge that sample size restrictions constrain them from undertaking adequate nonparametric testing, while their parametric tests and examinations of welfare measures principally focus upon the impacts of changing functional form and comparisons between the OOHB and DB approach (which, as indicated above, is of lesser interest to us given well founded theoretical and empirical criticism of the DB method), neither of which provide clear tests of the procedural invariance of the OOHB[4].

As discussed in detail in the subsequent section, all three forms of testing clearly reject procedural invariance in the OOHB method. This confirms the economic-theoretic prediction that the incentive compatibility deficiencies of the method result in some variant of strategic and/or anomalous behaviour. We therefore consider a number of further tests examining the nature of this behaviour. In constructing these tests we draw upon existing literature regarding repeated response CV formats and in particular that regarding procedural invariance in the DB method. Of particular relevance is the work of DeShazo (2002), who advances a "framing" hypothesis and an "anchoring" hypothesis as alternative behavioural models that might induce apparently anomalous responses to DB WTP questions, and Carson, Groves and Machina (2000), who discuss the various strategic behaviour responses to CV elicitation formats which fail tests of incentive compatibility.

The framing model derives from prospect theory and reference-dependent utility theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Bateman et al., 1997, 2005). Key to DeShazo's framing model is the assertion that in accepting an offered price, respondents assume that an informal exchange has been concluded. As a result, a respondent answering "Yes" to the first bid amount forms a 'reference point' that includes the surplus that they expect to enjoy from the exchange made at that price. From this reference point, the subsequent (and perhaps unexpected) presentation of a higher price is regarded by the respondent as precipitating a loss in surplus. DeShazo argues that this negative framing will tend to bias down the rate of acceptance of a bid level when presented as the second (and higher) price in an ascending sequence of questions. In contrast, a respondent answering "No" to the first bid amount effectively refuses the trade when offered at that price. In this case, no new reference point is formed, such that the second (and lower) price offered in the follow-up question is neither negatively, nor positively framed. This account also conforms with a strategic behaviour model derived from standard theory. However, the empirical manifestations of these accounts are identical and cannot be distinguished as both predict a lowering of acceptance rates for bid amounts under the HIGH 2 treatment. This prediction is presented in the second row of Table 1 as our first alternative hypothesis ($H_{A1}$).

The anchoring model opinions that respondents, uncertain of their own valuation of the good being offered them, may interpret the initial bid level as an amount that provides information on the good's 'true' value. As such, the first bid amount is assumed to act as an anchor towards which respondents adjust their valuation. In particular, respondents are assumed to update their valuation of the good to some amount between their original (and uncertain) valuation and the initial bid amount (DeShazo, 2002, Herriges and Shogren, 1996). Clearly, the behaviour hypothesized by this model should not affect a respondent's decision as to whether to accept or refuse the good offered at the initial bid price. The same is not true of decisions concerning the follow-up question. Having moved their valuation in the direction of the initial bid amount, respondents will be more likely to refuse a follow-up question offering the good at a higher price. Likewise, they will be more likely to accept a follow-up question offering the good at a lower price. DeShazo's anchoring model again predicts a lowering of acceptance rates for bid amounts under the HIGH 2 treatment while additionally anticipating inflated acceptance rates for the LOW 2 treatment. This is our second alternative hypothesis ($H_{A2}$) in which second responses are purely anchored to initial bids and has no strategic behaviour equivalent. In his investigation DeShazo finds empirical evidence in DB data that supports the framing model in favour of the null hypothesis and the anchoring hypothesis, though he states that "the anchoring hypothesis raises the most difficulties and most deserves further study" (p. 372).

Of course, the DB and OOHB formats differ fundamentally in the way in which the bid levels are revealed to the respondent. In particular, under the OOHB format respondents are made aware of both high and low bid levels in advance of the valuation questions. As such, anchoring, as envisaged by DeShazo for the DB format, may not carry over to the OOHB format. Given that respondents in the OOHB format are aware of both a high and a low price of provision for the good, it may well be that they do not anchor solely on the first bid amount presented to them. Rather, they may take both pieces of information into account, anchoring on a value somewhere in the range between the high and low bid values.

If we suppose that respondents weight both pieces of information equally then a respondent will form the same anchor independent of the order in which the two prices are presented. We refer to such behaviour, which is based upon literature from the heuristics field of cognitive psychology (Tversky and Kahneman, 1974), as "symmetric anchoring". Such behaviour will act so as to increase acceptance rates for low bid amounts and decrease acceptance rates for high bid amounts. However, symmetric anchoring implies that responses to the low (high) bid amount will be the same whether this amount is presented first or second. We formalize the predictions of the symmetric anchoring model within our third alternative hypothesis ( $H_{A3}$ ).

Alternatively, we might suppose that respondents take account of both pieces of information ($*L* and $*H*) but give added weight to the bid amount they are asked to consider first. This seems a reasonable formalisation of a variant of strategic behaviour in which the ordering of bids conveys information to the respondent; in this case the information being that the amount presented first is more likely to represent the real cost of the programme than is the amount presented second. We refer to such economic-theoretic strategic behaviour as "asymmetric anchoring". In this case, the anchoring point will differ according to the order in which the pair of bids are presented. In particular, not only will acceptance rates be deflated for high bid amounts and inflated for low bid amounts, but also these effects will be more pronounced when bids are presented as the second in the pair of WTP questions. The predictions of the asymmetric anchoring model make up our fourth alternative hypothesis ($H_{A4}$). Comparisons of $H_{A3}$ and $H_{A4}$ respectively provide an insight into whether non-standard or standard strategic behaviour explanations are best supported by our data.

Moreover, we would like to test a second null hypothesis concerning the strength of effects. Put simply, we might expect that the magnitude of any effect may be related to the difference between the high and low bids. Certainly, such an observation would be consistent with the symmetric and asymmetric anchoring models. To illustrate, imagine a particular bid amount presented to a respondent as the low amount in a pair of questions. As usual we shall label this $\$X_L$. Imagine also that the other question in the pair presents the bid amount $\$H^*$. Now according to both the symmetric and asymmetric anchoring models, respondents use some rule to calculate a value in the range between $\$X_L$ and $\$H^*$ on which they anchor. Respondents tend to adjust their valuation of the good towards this anchor and in so doing inflate the acceptance rate for $\$X_L$. Now imagine that $\$X_L$ had been paired with $\$H^{**}$, an amount higher than $\$H^*$. Provided respondents apply the same rule, then we would imagine that their anchoring point for the ($\$X_L$, $\$H^{**}$) pair would be higher than for the ($\$X_L$, $\$H^*$) pair. As such, the acceptance rate for $\$X_L$ when paired with $\$H^{**}$ should be somewhat greater than when it is paired with $\$H^*$. The inverse argument can be made with regards to a high bid amount, $\$X_H$ paired with the two different low bid values $\$L^*$ or $\$L^{**}$ (where $\$L^* < \$L^{**}$).

Again we can formalize these predictions into testable hypotheses. First we denote the acceptance rate for $\$X_L$ as $P_{X_L|H^*}$ when it is paired with $\$H^*$, and as $P_{X_L|H^{**}}$ when it is paired with $\$H^{**}$. The definitions of $P_{X_H|L^*}$ and $P_{X_H|L^{**}}$ follow accordingly. Moreover, it is always the case that $\$H^* < \$H^{**}$ and $\$L^* < \$L^{**}$. Using this notation, the first row of Table 2 describes a second null hypothesis ($H'_0$) in which there is procedural invariance in acceptance rates. The alternative hypothesis ($H'_A$), in which the magnitude of the difference between the high and low bid amounts affects acceptance rates, is defined by the predictions in the second row of Table 2.

The symmetric and asymmetric anchoring hypotheses make different predictions concerning the influence of question order on the strength of anchoring effects. In particular, symmetric anchoring implies that the rule by which respondents form their anchoring point is to choose the midpoint between the high and low bids. Consequently, the strength of the anchoring effect in this case, should not be influenced by the order in which the high and low bid are presented. This hypothesis is formalized in the two predictions making up hypothesis $H'_{A3}$ (row three of Table 2). In contrast, the asymmetric anchoring model implies that the rule used by respondents is to form their anchoring point in the range between the high and low bids but favouring the bid that is presented to them first. In this case, the order of presentation is taken as conveying information and hence will matter in determining the strength of the anchoring effect. In particular, the strength of the effect will be diminished for bid amounts presented first in the pair of OOHB questions, but will be exaggerated for bid amounts forming the second of the pair. The predictions for this hypothesis are given by $H'_{A4}$ in the final row of Table 2. As before symmetric anchoring supports a psychologically based expectation whereas asymmetric anchoring is a consequence of strategic behaviour.

**Survey Instrument Design And Implementation Process**

While CHS choose to test the OOHB on a private good, the method is clearly intended for use within the more usual CV domain of non-market public good valuation. Our empirical case study accordingly focuses upon such a good; the remediation of phosphate induced eutrophication problems affecting nearby rivers and lakes. Such water environment improvements have been a consistent focus of CV research (Desvousges et al., 1987; Sanders et al., 1990; Carson and Mitchell, 1993; Whittington et al., 1994; Goffe, 1995; Day and Mourato, 1998; Georgiou et al., 1998, 2000; Eisen-Hecht and Kramer, 2002) and so provide an ideal testing ground for our study.

The survey questionnaire was designed in accordance with best practice guidelines (Mitchell and Carson, 1989; Arrow et al., 1993; Bateman et al., 2002) and is available from the authors. Extensive use was made of focus groups to refine the description of the good and formulate an appropriate contingent market which was conveyed using a combination of clear and concise text augmented by visual aids. The resultant survey instrument was tested through a pilot survey of some 100 households after which the final survey questionnaire was refined. Both focus group and pilot exercises were also used to define an appropriate vector of bid amounts across which a range of positive and negative responses might be expected.

The survey questionnaire presented respondents with information regarding the nature of the eutrophication problem and details of a proposal to address this issue through the installation of new technology at sewage works so as to remove phosphates from household sewage. Survey respondents were informed that the implementation of such an environmental improvement programme would increase their annual household water bill. This payment vehicle is attractive from a CV perspective as it is effectively universal and unavoidable thereby avoiding the problems associated with discretionary payment vehicles (Mitchell and Carson, 1989).

Survey respondents were informed in advance that the cost to their household of the phosphate removal scheme was between a specified lower and upper bound ($L and $H respectively). An unseen random process was used to allocate respondents to one of thirteen pairs of amounts. Of these, seven described ascending sequences as follows: £10-£50; £25-£100; £50-£100; £75-£100; £100-£150; £100-£200; £48.50-£98.50[5]; these pairings being labelled LH1 to LH7 respectively. The remaining six pairs described descending sequences as follows: £50-£10; £100-£25; £100-£50; £100-£75; £150-£100; and £200-£100 (labelled HL1 to HL6 respectively). These pairs were chosen upon two criteria: first that they all fell within the distribution of bids implied by our focus group and pilot survey investigations; second, that they permitted ready and unambiguous testing of our hypotheses. In particular the repetition of certain bid amounts, such as the £100 bid, across a variety of contexts assists simple non-parametric testing of hypotheses.

The data were collected using a face-to-face interviewing techniques applied to a sample of randomly selected households in and around the city of Norwich, England. Surveying was conducted during a five week period in the summer of 2003. In total 1254 households provided completed questionnaires[6].

**Data And Analyses Of Procedural Invariance And Response Behaviour**

Table 3 reports the resulting acceptance rates for each bid level from each of the thirteen bid pairings describing our various ascending and descending sequences. Given that the central issue of this paper is to undertake procedural invariance testing of the OOHB format, we will begin our presentation of results by presenting first parametric, then welfare measure analyses, followed by nonparametric testing of this issue. These tests conclusively reject the null of procedural invariance and we consequently extend our nonparametric analysis through the various hypotheses of Tables 1 and 2 to examine the origin of problems with the OOHB approach.

In undertaking parametric tests of procedural invariance in the OOHB format we follow the established bivariate parametric testing framework of Cameron and Quiggin (1994). We allow the location parameter of the estimated WTP distributions to be functions of the socioeconomic characteristics of respondents to assure ourselves that differences in the estimated distributions are not being driven by random differences in the characteristics of our subsamples[7]. More fundamentally we specified our unconstrained model to incorporate two crucial aspects necessary to appropriate testing of procedural invariance. First, we allowed both the location parameter and the scale parameter of the WTP distributions to differ across the four treatments (i.e. LOW 1, HIGH 2, HIGH 1, LOW 2). Second, we allowed for non-perfect correlation in the responses of subjects to the first and second question[8]. The parameter estimates for this unconstrained model are reported in Table 4. We found that allowing each treatment to have a different WTP distribution results in an estimate for the correlation in subjects' responses that is almost identically one.

In contrast, our fully constrained model imposed the same distribution of WTP on each of the four treatments. The parameter estimates for the fully constrained model are reproduced in Table 5. In this model, the correlation in subjects' responses is only 0.43 which is significantly different from 1 (p-value $< 0.001$). Clearly, if this model were supported by the data (which it is not), then the implication would be that subjects' use different WTP values in responding to the first and second question. However, a likelihood ratio test comparing the fully constrained model to the unconstrained model confirmed that we can categorically reject the hypothesis of identical WTP distributions (test ratio = 69.85, critical value at 95% confidence = 28.90, p-value < 0.01).

Accordingly, we experimented to see which restrictions might be imposed across distributions and which must be rejected. We present our final parametric model in Table 6. Here, we found that we could not reject the possibility that the WTP distributions for all four treatments have identical scale parameters. Likewise, we could not reject the possibility that the marginal impacts of subjects' socioeconomic characteristics on the location parameters of the WTP distributions are identical across all four treatments. Moreover, our final model indicated that the WTP distributions in response to the first question are effectively the same regardless of whether subjects face the high or low price (LOW 1 = HIGH 1). However, compared to the WTP distribution for the first question, there is a significant upward shift in the implied WTP distribution for those facing a low price as the second question (LOW 2). Moreover, in contrast to the LOW 2 WTP distribution, those facing a high price as the second question (HIGH 2) tend to express significantly lower WTP, and the degree of shift is determined by the absolute difference between this bid and the low bid they received in the first question.

To confirm that this is the preferred model we carried out a likelihood ratio test that compared this constrained model with the fully unconstrained model. This test confirmed that we could not reject the restrictions imposed on the preferred model (test ratio = 23.32, critical value at 95% confidence = 26.30, p-value = 0.11). Moreover, imposing any further restrictions on the preferred model resulted in that model being significantly different from the unconstrained model. We also carried out a likelihood ratio test which confirmed that the preferred model differs significantly from the fully constrained model (test ratio = 46.52, critical value at 95% confidence = 5.99, p-value <

0.01). In summary then, our parametric testing rejects the null hypothesis of procedural invariance in the OOHB format.

Turning to consider tests of welfare measures comparisons were made between estimates of mean WTP using the following data;

(1) Responses to the first bid level.

(2) Responses to the second bid level.

(3) Responses to both first and second bid levels.

(4) Responses to the first bid level and the second bid level when this represents an ascending sequence.

(5) Responses to the first bid level and the second bid level when this represents a descending sequence.

Distribution functions in for each of these five cases were constructed using nonparametric maximum likelihood (NPML) estimates of the distribution function. This method for constructing the nonparametric distribution function is similar to the linear interpolation method described in Boman et al. (1999). The NPML estimator identifies intervals, known as equivalence classes, within which the probability distribution may, but not necessarily does, attribute probability mass (see Day, 2005). Here we assume that this probability mass is uniformly distributed across equivalence classes. Furthermore we truncate the distribution at the highest bid amount offered to respondents. Estimates of mean WTP are calculated as the area under the survivor function of this probability distribution function. Column 2 of Table 7 presents estimates of mean WTP from the 5 cases along with bootstrapped confidence intervals.

We wish to compare estimates of mean WTP from cases 2 to 5 with the base case. To do this we use a simple bootstrap procedure. Our null hypothesis is that mean WTP estimated from 1st bids only is an unbiased estimator of the population WTP. Let us call this estimate $\bar{x}^{1st}$. We draw 1000 bootstrap samples from the data and for each sample estimate mean WTP based on responses to 1st bids. We subtract $\bar{x}^{1st}$ from each bootstrap estimate and square the results. The bootstrap procedure, therefore, provides an estimate of the sampling distribution of squared deviations from mean WTP under the null. To test whether responses to the 2nd bid result in an estimate of WTP that differs significantly from that from the 1st bid, we first subtract $\bar{x}^{1st}$ from the estimate of mean WTP based on responses to 2nd bids and square the result. This statistic can then be compared to the distribution of squared deviations under the null to ascertain the likelihood of observing such a difference by chance.

The same procedure was repeated for each of the cases and the results presented in Column 3 of Table 7. These can be read as describing the probability of a difference in means of the observed size or more occurring purely by chance. For example, we can say that there is less than 0.1% chance of observing a mean value that differs from $\bar{x}^{1st}$ by as much as that calculated from responses to the 2nd bids. In contrast, there is a 70.1% chance of observing a mean value that differs from $\bar{x}^{1st}$ by as much as that calculated from responses to 1st bids and descending 2nd bids.

The analysis again rejects the hypothesis of procedural invariance in the OOHB format. It appears that estimates of mean WTP derived using just 1st bids are significantly larger than those estimated using just 2nd bids and, for that matter, significantly larger than those that use both 1st and 2nd bid information. The analysis also gives some insight into what is driving these differences. It seems that the main bias is to be found in data pertaining to responses to 2nd bids in an ascending sequence. This conforms to the asymmetric anchoring pattern symptomatic of strategic behaviour.

Our final and arguably most informative set of analyses are provided by the nonparametric tests set out in Tables 1 and 2. In Table 1 our null hypothesis of procedural invariance ($H_0$) states that acceptance rates should be invariant to the treatment. However, even a visual inspection of the raw acceptance rates reported in Table 3 strongly suggests that we should reject this null. For example, acceptance rates for the £100 bid level under the LOW 2 treatment reach as high as 61.5% compared with a low of 22.1% under the HIGH 2 treatment.

In total our experimental design allows for 48 pairwise comparisons of response proportions where the same price is presented to separate subsamples under different treatments. To ensure the robustness of nonparametric testing, each subsample had between 90 and 106 subjects. With the exception of just 4 of these 48 comparisons, the response proportions satisfy the asymmetric anchoring ($H_{A4}$) pattern of strategic behaviour namely that $P_{X_{L2}} > P_{X_{L1}} > P_{X_{H1}} > P_{X_{H2}}$. Furthermore, 23 of these differences are statistically significant[9] with the pattern of significance being detailed in Table 8. As can be seen, the consistency of significance is lower where solely first bound responses are compared and highest where we solely second bound responses where the strategic payoff is highest.

The findings reported in Table 8 convincingly reject the null of procedural invariance. However, they also provide evidence regarding the framing and pure anchoring models as formulated by DeShazo for the DB format ($H_{A1}$ and $H_{A2}$ in Table 1). These assert that responses to the initial bid level are unbiased. As such, acceptance rates for the same bid level under treatment LOW 1 should not differ systematically from those under treatment HIGH 1. Alternatively, the modified anchoring models that take account of pre-revelation of both high and low prices ($H_{A3}$ and $H_{A4}$ in Table 1), both predict acceptance rates under the LOW 1 treatment to exceed those under the HIGH 1 treatment. Indeed, making this comparison for the £100 bid level reveals that acceptance rates for the LOW 1 treatment (42.1% and 46.9%) do appear to be somewhat higher than they are for the HIGH 1 treatment (30.5%, 31.1% and 36.6%). In this case, since our alternative hypotheses are directional, we employ a one-tailed z-test to compare these differences in proportions. The £50 bid level is also repeated in both treatments so that there are seven pairwise comparisons that can be made. Of these we find three to be different at the 95% level of confidence, two at the 90% level of confidence and the remaining two not to show statistically significant differences. Thus the weight of evidence rejects equality of acceptance rates over these two treatments and with it DeShazo's framing and anchoring hypotheses ($H_{A1}$ and $H_{A2}$). Rather, the data appear to support the symmetric and asymmetric anchoring models formulated in hypotheses $H_{A3}$ and $H_{A4}$.

Finally, to distinguish between the two symmetric and asymmetric anchoring models, we compare the HIGH 1 treatment with the HIGH 2 treatment and the LOW 1 treatment with the LOW 2 treatment. Since these two models are founded on the assumption that responses are determined in part by the specific values of the pair of bids presented to a respondent, we restrict our testing to comparisons in which the bid pairs are identical but presented in different orders. Again we employ one-tailed z-tests of differences in proportions to test our directional hypotheses.

There are six suitable comparisons that can be made between the HIGH 1 and HIGH 2 treatments and a further six for the LOW 1 and LOW 2 treatments. All twelve of these comparisons were found to conform with the directional expectations of the asymmetric anchoring model. Testing revealed that in two cases differences were significant at the 99% confidence level, a further three at the 95% level and two more at the 90% level[10].

With regards to our first set of testable hypotheses as described in Table 1, we can conclude that our data strongly rejects procedural invariance in the OOHB format. Moreover, the pattern of responses is unlike that described by DeShazo for DB data. Rather we find that our OOHB data most strongly supports a model of asymmetric anchoring in which respondents form an anchor based on both high and low price information but give additional weight to the price level to which they are asked to respond first.

Now let us consider our strength of effect hypotheses as described in Table 2. Once again, the null hypothesis ($H'_0$) is one of procedural invariance; that is, acceptance rates for a particular bid amount under a particular treatment are supposed to be unaffected by the value of the other bid in the pair. Alternatively, the anchoring models suggest that the magnitude of the difference between the pair of bids will have predictable impacts on the rates of acceptance for those bid amounts ($H'_A$).

Our experiment was structured in part so that a single design point, the £100 bid amount, would provide a robust basis of our tests of effect strength. So, examining acceptance rates for this bid level presented as the second in the pair of WTP questions (i.e. in treatments LOW 2 and HIGH 2) provides a ready means of comparing our procedural invariance null hypothesis $H'_0$ with $H'_A$. Within the LOW 2 treatment, acceptance rates for the £100 amount are 48.4% when preceded by the £150 bid but increase sharply to 61.5% when preceded by the £200 bid. Conversely acceptance rates for the £100 bid amount under the HIGH 2 treatment fall from 32.7% when paired with £25, to 23.7% when paired with £50, down to 22.1% when paired with £75. Together these provide four directional hypotheses concerning differences in acceptance rates for which we can employ one-tailed z-tests of differences in proportions. These show that three of these four differences are significant at a 95% confidence level (with one being significant at the 99% level). Given this weight of evidence we reject the null hypothesis of procedural invariance in favour of an alternative in which the acceptance rate for a bid amount is determined in part by the value of the bid amount with which it is paired.

Given this conclusion, we can again extend our analysis to examine whether some variant of anchoring or a strategic behaviour model is supported by our data. To undertake this we consider the differing predictions made by the symmetric and asymmetric anchoring models. As described in hypothesis $H'_{A3}$ in Table 2, the symmetric anchoring model predicts that for a particular bid amount the strength of the effect will be the same independent of the order in which the questions are asked. In contrast, the asymmetric anchoring model predicts that the strength of the effect will be less for that particular bid amount when presented as the first question in the pair (hypothesis $H'_{A4}$ in Table 2).

The bid design is constructed so as to test these alternative hypotheses using the £100 bid amount. In particular, given pairs of bid amounts with £100 as the LOW 1 bid are also employed in reverse presentation order with £100 as the LOW 2 bid. Similarly mirrored pairings are implemented with £100 as either the HIGH 1 or HIGH 2 bid. Under the LOW 2 and HIGH 2 treatments (i.e. where the £100 bid amount was presented second) three of the four possible comparisons returned statistically significant differences in the direction that was to be expected under an anchoring hypothesis. However, making the same set of comparisons for the LOW 1 and HIGH 1 treatments (i.e. where the £100 bid amount was presented first) we find that none of the four comparisons return statistically significant differences (and one even reverses the ordering that would be expected under pure anchoring). It seems that the strength of the anchoring effects is more pronounced for a bid amount forming the second of the pair of questions than it is for that bid amount forming the first of the pair; a pattern of responses that supports the asymmetric anchoring model ($H'_{A4}$) over the symmetric anchoring model ($H'_{A3}$).

In summary, the tests presented in Tables 1 and 2 reject numerous proposed models of behaviour (including procedural invariance) but are compatible with a model in which respondents anchor their WTP on a value in the range between the high and low bid amounts. Furthermore, it appears that responses are influenced by the order of presentation of bid amounts. In particular, the data supports the asymmetric anchoring model which is an indicator of strategic behaviour.

**Discussion and Conclusions**

The OOHB approach is an innovative addition to the armoury of CV elicitation methods. It combines the response simplicity of a dichotomous choice approach with substantial statistical efficiency gains over the more conventional SB approach. Furthermore it explicitly sets out to avoid the adverse ‘surprise’ induced by the unanticipated follow-up question in the DB format. However, it still presents respondents with two possible prices for the good on offer and as such fails the Gibbard-Satterthwaite conditions for incentive compatibility. Standard economic theory would suggest that such a format is likely to induce strategic behaviour within responses. Alternatively the dual price approach of the OOHB may lead respondents to rely upon anchoring heuristics. Both possibilities suggest that the format will fail tests of procedural invariance.

This paper presents the first systematic examination of procedural invariance within the OOHB method. Through a series of nonparametric, parametric and welfare measure comparisons we consistently reject the null hypothesis of procedural invariance.

This is, we contend, an important finding given the rapid uptake of the approach within decision analyses and applied CV work and its recommendation as a preferred method for future applications.

We extend our analysis to examine the nature of the effects induced by the method. Our findings support a strategic behaviour account of responses (as reflected in our asymmetric anchoring model) over reliance upon a pure anchoring heuristic.

These findings underline the importance role which incentive compatibility can play in the valuation of preferences for non-market public goods. However, by the same argument it should be that the OOHB approach is less vulnerable to procedural variance problems within a real trading private goods environment (which is not dissimilar from that employed within the original CHS example) or in an experimental setting within which such conditions can be maintained.

## References


Alberini, A. and Carson, R.T., (2001) Yea-Sayers, Nay-Sayers or Just Plain Confused? Mixture of Populations in Contingent Valuation Survey Responses, paper presented at *The Eleventh Annual Conference of the European Association of Environmental and Resource Economists,* University of Southampton, June 28th - June 30th 2001.

Alberini, A., Kanninen, B. and Carson, R.T., (1997) Modeling response incentive effects in dichotomous choice contingent valuation data, *Land Economics*, 73: 309-324.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H., (1993), Report of the NOAA panel on contingent valuation, *Federal Register,* 58: 4601-4614.

Barreiro, J., Sánchez, M. and Viladrich-Grau, M. (2005) How much are people willing to pay for silence? A contingent valuation study, *Applied Economics*, 37(11): 1233-1246.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroğlu, E., Pearce, D.W., Sugden, R. and Swanson, J. (2002) *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar Publishing, Cheltenham.

Bateman, I.J., Day, B.H., Dupont, D.P. and Georgiou, S. (2004) OOH la la: Bid range and direction effects in the one-and-one-half bound dichotomous choice approach, *CSERGE Working Paper EDM 04-06,* Centre for Social and Economic Research on the Global Environment, University of East Anglia.

Bateman, I.J., Day, B.H., Dupont, D., Georgiou, S., Matias, N.G.N. Morimoto, S. and Subramanian, L. (forthcoming) Cost-benefit analysis and the prevention of eutrophication, in Pearce, D.W. (ed.) *Valuing the Environment in Developed Countries: Case Studies*, Edward Elgar, Cheltenham, UK.

Bateman, I.J., Kahneman, D., Munro, A., Starmer, C. and Sugden, R. (2005) Testing competing models of loss aversion: An adversarial collaboration, *Journal of Public Economics*, 89:1561–1580.

Bateman, I.J., I.H. Langford, A.P. Jones, A.P. and G.N. Kerr, (2001) Bound and path effects in double and triple bounded dichotomous choice contingent valuation, *Resource and Energy Economics,* 23(3): 191-213.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997) A test of the theory of reference-dependent preferences, *Quarterly Journal of Economics,* 112(2): 479-505.

Boman, M., G. Bostedt and B. Kristrom (1999) Obtaining welfare bounds in discrete-response valuation studies: A nonparametric approach, *Land Economics*, 75, pp. 284-294.

Bradford, W. David, James Zoller, Gerard A. Silvestri, (2004) Estimating the Effect of Individual Time Preferences on the Demand for Preventative Health Care, *CHEPS Working Paper 007-04*, Center for Health Economic and Policy Studies, Medical University of South Carolina.

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(3): 218-234.

Carson, Richard T. (forthcoming), *Contingent Valuation: A Comprehensive Bibliography and History*. Northampton, MA: Edward Elgar.

Carson, R., Groves, T. and Machina, M. (2000) Incentive and Informational Properties of Preference Questions, *Manuscript.* University of California San Diego, February.

Carson, R.T. and Mitchell, R.C. (1993) The value of clean water: the public's willingness to pay for boatable, fishable and swimmable quality water, *Water Resources Research*, 29(7): 2445-2454.

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.

Ciriacy-Wantrup, S.V. (1952) *Resource Conservation Economics and Policies*, Berkeley, University of California Press.

Ciriacy-Wantrup, S.V. (1947).Capital Returns from Soil-Conservation Practices *Journal of Farm Economics* 29:1181-96.

Cooper, J., W.M. Hanemann, and G. Signorello, (2002) One-and-One-Half-Bound Dichotomous Choice Contingent Valuation, Review of Economics and Statistics 84(4): 742-750.

Cummings, R.G., Brookshire, D.S. and Schulze, W.D. (eds.) (1986) *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Method*, Rowman and Allenheld, Totowa, N.J.

Day (2005) Distribution-free estimation with interval-censored contingent valuation data: Troubles with Turnbull?, *CSERGE Working Paper*, EDM 05-07, University of East Anglia, Norwich, UK.

Day, B., and Mourato, S. (1998) Willingness to pay for water quality maintenance in Chinese rivers, *CSERGE Working Paper WM 98-02*, Centre for Social and Economic Research on Global environment, University of East Anglia, and University College of London.

DeShazo, J.R., (2002) Designing Transactions without Framing Effects in Iterative Question Formats, *Journal of Environmental Economics and Management*, 43, 360-385.

Desvousges, W.H., Smith, V.K. and Fisher, A. (1987) Option price estimates for water quality improvements: a contingent valuation study of the Monongahela River*, Journal of Environmental Economics and Management*, 14: 248-267.

Eisen-Hecht, J.I. and Kramer, R.A. (2002) A Cost-Benefit Analysis of water quality protection in the Catawba basin, *Journal of the American Water Resources Association*, 38(2): 453-465.

Georgiou, S., Langford, I.H., Bateman, I.J. and Turner, R.K. (1998) Determinants of Individuals' Willingness to Pay for Perceived Reductions in Environmental Health Risks: A Case Study of Bathing Quality, *Environment and Planning A,* 30: 577-594.

Georgiou, S., Bateman, I.J., Langford, I.H. and Day, R.J. (2000) Coastal bathing water health risks: Assessing the adequacy of proposals to amend the 1976 EC Directive, *Risk, Decision and Policy*, 5: 49-68.

Gibbard A., (1973) Manipulation of Voting Schemes: A General Result, *Econometrica* 41, 587-601.

Goffe, P.L. (1995) The benefits of improvements in coastal water quality: a contingent approach, *Journal of Environmental Economics and Management*, 45:305-317.

Hanemann, W.M., J. Loomis, and B. Kanninen, (1991) Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation, *American Journal of Agricultural Economics* 73(4), 1255-1263.

Herriges, J. and J. Shogren (1996) Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning, *Journal of Environmental Economics and Management*, 30, 112-131.

Kahneman, D. and A. Tversky, (1979) Prospect Theory: An Analysis of Decisions Under Risk, *Econometrica,* 47: 263-291.

Kerr, John M., Pender, John L, and Suyanto (2004) Property Rights, Environmental Services and Poverty in Indonesia, *Mimeo*, Michigan State University.

Langford, I.H., Bateman, I.J. and Langford, H.D. (1996) A multilevel modelling approach to triple-bounded dichotomous choice contingent valuation, *Environmental and Resource Economics*, 7(3):197-211.

McFadden, D., (1994) Contingent valuation and social choice, American Journal of Agricultural Economics, 76: 689-708.

Mitchell, R.C. and Carson, R.T. (1989) Using Surveys to Value Public Good: The Contingent Valuation Method, Resources for the Future, Washington, DC.

NCEE and NCER (2000) Stated Preference: What do we know? Where do we go? Proceedings from a Workshop Sponsored by the US Environmental Protection Agency (Session Four, Panel Discussion), *The NOAA Panel and the Seven Year Itch*, National Center for Environmental Economics and National Center for Environmental Research, Washington, DC.

Sanders, L.D., Walsh, R.G. and Loomis, J.B. (1990), 'towards empirical estimation of the total value of protecting rivers', *Water Resources Research*, 26(7):1345-1357.

Satterthwaite, M.A. (1975) Strategy-proofness and Arrow's conditions: Existence and correspondence theorems of voting procedures and social welfare functions, *Journal of Economic Theory*, 10: 187- 217.

Scarpa, R. (2003) Social & Environmental Benefits of Forestry Phase 2: The Recreation Value of Woodlands, *Report to the UK Forestry Commission*, Centre for Research in Environmental Appraisal & Management, University of Newcastle.

Scarpa, R. and Bateman. I.J. (2000) Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete choice CV studies, *Land Economics*, 76(2): 299-311.

Signorello, Giovanni and Cooper, Joseph C., (2002) Farmer Premiums for the Voluntary Adoption of Conservation Plans, *FEEM Working Paper No. 27.2002*, Fondazione Eni Enrico Mattei, Milan, Italy.

Sutton, W.R., Larson, D.M. and Jarvis, L.S. (2004) A new approach for assessing the costs of living with wildlife in developing countries, *DEA Research Discussion Paper 69*, Directorate of Environmental Affairs, Ministry of Environment and Tourism, Namibia.

Tversky, Amos and Daniel Kahneman (1974) Judgment under uncertainty: Heuristics and biases, *Science*, 185(4157): 1124–1131.

Tversky, Amos and Daniel Kahneman (1991) Loss aversion in riskless choice: a reference-dependent model, *Quarterly Journal of Economics*, 106: 1039-1061.

Tversky, Amos and Daniel Kahneman (1992) Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5: 297-323.

Whittington, D.G., Amaral, C.D., McClelland, Wang, H. and Poulos, C. (1994) The economic value of improving the environmental quality of Galveston Bay, *GBNEP – 38*, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill and Galveston Bay National Estuary Program.

**Table 1: Predicted Relationships Between Acceptance Rates For Bid Amounts Under Different Hypotheses**

| Hypothesis | Predicted Relationships between Acceptance Rates by Treatment | | | | | | |
|---|---|---|---|---|---|---|---|
| | LOW 2 | | LOW 1 | | HIGH 1 | | HIGH 2 |
| $H_0$: CHS (procedural invariance) | $P_{X_{L2}}$ | $=$ | $P_{X_{L1}}$ | $=$ | $P_{X_{H1}}$ | $=$ | $P_{X_{H2}}$ |
| $H_{A1}$: Framing | $P_{X_{L2}}$ | $=$ | $P_{X_{L1}}$ | $=$ | $P_{X_{H1}}$ | $>$ | $P_{X_{H2}}$ |
| $H_{A2}$: Pure Anchoring (on initial bid only) | $P_{X_{L2}}$ | $>$ | $P_{X_{L1}}$ | $=$ | $P_{X_{H1}}$ | $>$ | $P_{X_{H2}}$ |
| $H_{A3}$: Symmetric Anchoring (on both bids) | $P_{X_{L2}}$ | $=$ | $P_{X_{L1}}$ | $>$ | $P_{X_{H1}}$ | $=$ | $P_{X_{H2}}$ |
| $H_{A4}$: Asymmetric Anchoring (on both bids) | $P_{X_{L2}}$ | $>$ | $P_{X_{L1}}$ | $>$ | $P_{X_{H1}}$ | $>$ | $P_{X_{H2}}$ |

**Table 2: Predicted Relationships Concerning The Strength Of Framing Or Anchoring Effects Under Different Hypotheses**

| Hypothesis | Predicted Relationships between Conditional Acceptance Rates ($\$H^{*} < \$H^{**}$ and $\$L^{*} < \$L^{**}$) |
|---|---|
| $H_0'$: CHS (procedural invariance) | $P_{X_L \mid H^{*}} = P_{X_L \mid H^{**}}$ and $P_{X_H \mid L^{*}} = P_{X_H \mid L^{**}}$ |
| $H_A'$: Procedural Variance (e.g. pure anchoring) | $P_{X_L \mid H^{*}} < P_{X_L \mid H^{**}}$ and $P_{X_H \mid L^{*}} < P_{X_H \mid L^{**}}$ |
| $H_{A3}'$: Symmetric Anchoring (on both bids) | $\left(P_{X_{L1} \mid H^{*}} - P_{X_{L1} \mid H^{**}}\right) = \left(P_{X_{L2} \mid H^{*}} - P_{X_{L2} \mid H^{**}}\right)$ and $\left(P_{X_{H1} \mid L^{*}} - P_{X_{H1} \mid L^{**}}\right) = \left(P_{X_{H2} \mid L^{*}} - P_{X_{H2} \mid L^{**}}\right)$ |
| $H_{A4}'$: Asymmetric Anchoring (on both bids) | $\left(P_{X_{L1} \mid H^{*}} - P_{X_{L1} \mid H^{**}}\right) < \left(P_{X_{L2} \mid H^{*}} - P_{X_{L2} \mid H^{**}}\right)$ and $\left(P_{X_{H1} \mid L^{*}} - P_{X_{H1} \mid L^{**}}\right) < \left(P_{X_{H2} \mid L^{*}} - P_{X_{H2} \mid L^{**}}\right)$ |

**Table 3: Comparison Of Bid Acceptance Rates Across Treatments[1]**

| Ascending Sequence (LH) | | | | | | Descending Sequence (HL) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | LOW 1 | | HIGH 2 | | | | HIGH 1 | | LOW 2 | | |
| | First bound | | Second bound | | | | First bound | | Second bound | | |
| **Label** | **Initial (Lower) bid amount ($L)** | **Acceptance rate for $L (%)** | **Follow-up (Higher) bid amount ($H)** | **Acceptance rate for $H[2] (%)** | **Absolute difference ($H - $L)** | **Label** | **Initial (Higher) bid amount ($H)** | **Acceptance rate for $H (%)** | **Follow-up (Lower) bid amount ($L)** | **Acceptance rate for $L[††] (%)** | **Absolute difference ($H - $L)** |
| LH1 | 10 | 90.1% | 50 | 46.5% | 40 | HL1 | 50 | 59.4% | 10 | 90.1% | 40 |
| LH2 | 50 | 55.7% | 100 | 23.7% | 50 | HL2 | 100 | 36.6% | 50 | 65.6% | 50 |
| LH3 | 100 | 42.1% | 150 | 26.3% | 50 | HL3 | 150 | 38.9% | 100 | 48.4% | 50 |
| LH4 | 100 | 46.9% | 200 | 9.2% | 100 | HL4 | 200 | 28.1% | 100 | 61.5% | 100 |
| LH5 | 25 | 82.1% | 100 | 22.1% | 75 | HL5 | 100 | 30.5% | 25 | 85.3% | 75 |
| LH6 | 75 | 41.6% | 100 | 32.7% | 25 | HL6 | 100 | 31.1% | 75 | 42.2% | 25 |
| LH7 | 48.50 | 52.6% | 98.50 | 18.6% | 50 | | | | | | |

[1] Total sample size = 1254 households. Sample sizes within each treatment vary from a minimum of 90 to a maximum of 106 households. Bid amounts are in GB pounds. [2] Acceptance (rejection) rates for follow-up questions include as ‘yes’ (‘no’) responses those respondents who were not asked the second question because they had implicitly accepted (rejected) this amount in their initial response.

**Table 4: Fully Unconstrained Model: WTP Distributions Allowed To Differ For First And Second Bids And For High And Low Bids; LOW1 ≠ LOW2 ≠ HIGH1 ≠ HIGH2 For Location And Scale Parameters**

| **Variable** | **Coefficients** | **Std. Err.** |
|---|---|---|
| **Equation 1: Responses to First Question:** | | |
| **Respondents receiving low bids:** | | |
| Ln PRICE | -0.7368 | 0.0721*** |
| INCOME | 0.0205 | 0.0037*** |
| AGE | 0.0008 | 0.0029 |
| FEMALE | -0.1205 | 0.1089 |
| VISITOR | 0.0006 | 0.0006 |
| CONSTANT | 2.6446 | 0.3247*** |
| **Respondents receiving high bids:** | | |
| Ln PRICE | 0.1651 | 0.0677** |
| INCOME | -0.0130 | 0.0055** |
| AGE | -0.0052 | 0.0043 |
| FEMALE | -.2635 | 0.1598* |
| VISITOR | 0.0000 | 0.0010 |
| ABS DIFF FROM 1st PRICE | -0.0051 | 0.0023** |
| **Equation 2: Responses to Second Question:** | | |
| **Respondents receiving low bids:** | | |
| Ln PRICE | -0.7023 | 0.0796*** |
| INCOME | 0.0154 | 0.0037*** |
| AGE | -0.0045 | 0.0031 |
| FEMALE | 0.1084 | 0.1142 |
| VISITOR | 0.0021 | 0.0006*** |
| CONSTANT | 3.2306 | 0.3592*** |
| **Respondents receiving high bids:** | | |
| Ln PRICE | -0.0850 | 0.0707 |
| INCOME | -0.0033 | 0.0057 |
| AGE | -0.0026 | 0.0046 |
| FEMALE | 0.0488 | 0.1691 |
| VISITOR | -0.0019 | 0.0010* |
| ABS DIFF FROM 1st PRICE | -0.0850 | 0.0707 |
| ρ | 1.0000 | 0.0000*** |
| Log likelihood | -1057.234 | |

Confidence levels are: *** 99% ; ** 95% ; * 90%

**Table 5: Fully Constrained Model: WTP Distributions Constrained To Be Identical: LOW1 = LOW2 = HIGH1 = HIGH2 For Location And Scale Parameters**

| **Variable** | **Coefficients** | **Std. Err.** |
|---|---|---|
| Ln PRICE | -0.7783 | 0.0584*** |
| INCOME | 0.0133 | 0.0024*** |
| AGE | -0.0040 | 0.0019** |
| FEMALE | 0.0652 | 0.0680 |
| VISITOR | 0.0009 | 0.0004** |
| CONSTANT | 3.0917 | 0.2777*** |
| $\rho$ | 0.4263 | 0.1134*** |
| Log likelihood | -1092.157 | |

Confidence levels are: *** 99% ; ** 95% ; * 90%

**Table 6: Preferred Model: WTP Distributions Have Fully Constrained Scale Parameters (LOW1 = LOW2 = HIGH1 = HIGH2) And Equal Covariate Effects Whilst Location Parameters Are Held Equal For First Price Responses But Allowed To Differ For Second Price Responses ((LOW1 = HIGH1) ≠ LOW2 ≠ HIGH2)**

| Variable | Coefficients | Std. Err. |
|---|---|---|
| **Equation 1: Responses to First Question:** | | |
| CONSTANT | 2.8975 | 0.2509*** |
| **Equation 2: Responses to Second Question:** | | |
| CONSTANT | 3.1106 | 0.2371*** |
| ABS DIFF FROM 1st PRICE | -0.0109 | 0.0015*** |
| **Common coefficients constrained to be same for both equations:** | | |
| Ln PRICE | -0.7312 | 0.0504*** |
| INCOME | 0.0142 | 0.0025*** |
| AGE | -0.0034 | 0.0020* |
| FEMALE | 0.0470 | 0.0711 |
| VISITOR | 0.0009 | 0.0004** |
| $\rho$ | 0.9991 | 27.9580 |
| Log likelihood | -1068.895 | -1068.895 |

Confidence levels are: *** 99% ; ** 95% ; * 90%

**Table 7: Estimates Of Mean WTP Derived From Combining Responses To Different Bounds Of The OOHB Format**

| Case | Mean WTP (95% Confidence Interval) | Probability that this is equal to $1^{st}$ bids Mean |
|---|---|---|
| (1) $1^{st}$ bids | 96.98 (90.72 to 101.90) | |
| (2) $2^{nd}$ Bids | 86.09 (79.64 to 92.12) | <0.001 |
| (3) $1^{st}$ and $2^{nd}$ Bids | 88.47 (84.17 to 92.74) | 0.002 |
| (4) $1^{st}$ Bids and Ascending $2^{nd}$ Bids | 87.06 (82.53 to 91.58) | <0.001 |
| (5) $1^{st}$ Bids and Descending $2^{nd}$ Bids | 96.11 (90.83 to 100.79) | .701 |

**Table 8: Pairwise Comparisons Of Response Proportions For Same Price Under Different Treatments In Separate Subsamples**

| Treatments Compared | Number of cases | Cases following the asymmetric anchoring pattern ($H_{A4}$) | Responses portions are different with 95% confidence (two-tailed test) |
|---|---|---|---|
| LOW 2 vs. HIGH 2 | 7 | 7 | 7 |
| LOW 2 vs. HIGH 1 | 7 | 7 | 5 |
| LOW 1 vs. HIGH 2 | 7 | 7 | 5 |
| LOW 2 vs. LOW 1 | 8 | 7 | 2 |
| LOW 1 vs. HIGH 1 | 7 | 6 | 2 |
| HIGH 1 vs. HIGH 2 | 12 | 10 | 2 |

**Endnotes**

---

[1] These gains may be extended through the addition of multiple bounds (Langford et al., 1996; Scarpa and Bateman, 2000).

[2] Our recommendations to the UK Department for Transport (Bateman et al., 2002) pre-date the present study.

[3] For example, we test whether the proportion of subjects (drawn from independent sample) responding “yes” to a price of $100 does not differ whether that price is offered as the first or second price in a OOHB question (tested via Pearson’s chi-squared test for independent samples). We assume only that a respondent who refuses to pay some lower initial price (say $50) is implicitly indicating that they will respond “no” to the follow-up price of $100. Accordingly, under the null, the “no” response proportion for the sample facing $100 as a follow-up price can be calculated as the ratio of the sum of actual *and implied* “no” responses to the sample size. Likewise, under the null, the “yes” response proportion can be taken as the ratio of the actual “yes” responses to the sample size. Marginal probabilities calculated in this way are equally valid irrespective of whether the price of interest (here $100) is presented as either the first or second price in the OOHB format. We contend that this approach provides the most straightforward and unequivocal nonparametric testing. However, by contrast, CHS choose only to examine the responses of those that *actually* respond to the $100 price. Of course, only those with a WTP in excess of $50 get to actually respond to the $100 bid as a follow-up price. Accordingly, CHS observe that under the null, the proportion of actual “yes” responses to $100 as a follow-up price to $50 should exceed the proportion of actual “yes” responses to $100 as an initial price. The test employed by CHS, therefore, differs from our test in comparing a conditional probability to a marginal probability, rather than comparing two marginal probabilities. Interestingly CHS report that “in most instances” (p746, footnote 8) they observe that the marginal probability exceeds the conditional probability, i.e. procedural invariance is *not* supported. However, CHS note that their sample size was insufficient to undertake adequate tests of the significance of these findings.

[4] CHS do report one test in their footnote 13 (p.748) where responses to a SB question format (actually the first question of a DB format, though from the point of view of the subject these are identical) are compared with responses to the first question of an OOHB

question format. Their test rejects the null hypothesis that parameters are drawn from the same distribution raising questions regarding the procedural invariance of CHS's own dataset (likelihood ratio test score of 52.76 compared to a critical test score of 9.49 at the 95% confidence level). However, further parametric testing of procedural invariance is not undertaken, tests instead focussing on functional form and OOHB versus DB issues.

[5] Note that the £48.50-£98.50 pair was only used in the ascending sequence to provide a side analysis comparison with the £50-£100 pair examining whether the implied greater accuracy of the former pair resulted in any significant impact upon acceptance rates. A comparison of acceptance rates for the LH bid pairs (£48.50, £98.50) and (£50, £100) in Table 3 reveals no clear evidence that use of the former 'more accurate' pair results in any substantial impact upon acceptance rates.

[6] A further 1067 households were approached but declined to take part in the survey giving a response rate of 54%. The most common reasons for refusing to take part were time constraints and a lack of interest in any survey (respondents were unaware of the subject matter of the study at the outset of the survey). Arguably this may mean sample values might differ from those held across the population. This is a common problem for survey research and is not the focus of the present study, being discussed in Bateman et al., (forthcoming).

[7] In this case the variables INCOME (respondents household income), AGE (respondents age), FEMALE (=1 if respondent is female; zero otherwise) and VISITOR (frequency of use of the non-market good being valued). Other variables in our parametric models are the natural logarithm of the bid level (Ln PRICE) and, for models of second bound responses, the absolute difference between the first and second bid amount (ABS DIFF FROM 1st PRICE).

[8] In contrast CHS assume that subjects' implied WTP is perfectly correlated across the two questions, thus making it impossible to perform the Cameron and Quiggin procedural invariance tests. Moreover, CHS do not compare estimated parameters across different models to test the null hypothesis that the distribution of WTP values is invariant as to whether that distribution is estimated from the first price data, second price data or some combination of the two. Furthermore, although CHS themselves do state (p.748) that the

appropriate range for WTP in their data is only over non-negative values, they model WTP using a univariate logistic distribution, which allows for negative values. All of the models reported in the present paper conform with economic theory by estimating WTP distributions that are only defined for non-negative values.

[9] It should be noted that not all of the tests in Table 8 are independent; the same price might appear more than once in any one treatment such that multiple comparisons can be made for a subsample facing that same price under a different treatment.

[10] Note that the significance of differences was generally stronger for the HIGH 1 versus HIGH 2 comparisons. This may reflect that fact that the highest bid level (£200) can only appear in this comparison set (while the lowest bid level of £25 only appears in the LOW 1 versus LOW2 comparison). This may suggest that the degree of strategic behaviour (and hence asymmetric anchoring) is positively associated with the absolute level of bids. This seems plausible if high bid amounts are perceived as indicators of an authority which is attempting to capture the consumer surplus associated with provision of a public good.