

Identification and Estimation with Partial Respondents and Anchoring Effects

Bertrand Melenberg ^a, Arthur van Soest ^b and Rosalia Vazquez-Alvarez ^{c,*},

^a Centre For Economic Research, (CentER), Tilburg University, The Netherlands

^b RAND Corporation, Santa Monica and Centre for Economic Research, (CentER), Tilburg University

^c Swiss Institute for International Economics and Applied Economic Research (SIAW), University of St.Gallen



This version: February, 2008^{*}

Abstract

Household surveys often suffer from nonresponse on variable such as income, savings or wealth. The work by Charles F. Manski in the 1990s shows how bounds on conditional quantiles on the variable of interest can be derived, allowing for any type of non-random item nonresponse. The width between these bounds can be reduced using follow up questions in the form of unfolding brackets for initial non-respondents. However, evidence from both the psychology and economic literature suggest that such design is vulnerable to anchoring effects. In this paper Manski's bounds are extended to incorporate information provided by bracket respondents allowing for three alternative nonparametric assumptions on anchoring. The new bounds are applied to earnings in the 1996 wave of the Health and Retirement study. The results show that introducing categorical questions in the form of unfolding designs can be useful to increase the precision of the bounds, even if anchoring is allowed for.

Keywords: Identification, item nonresponse, survey design, anchoring effects, unfolding bracket design

JEL – Classification : C13, C14, C81, D31

* Address for correspondence: SIAW, University of St.Gallen, Bodanstrasse, 8, CH-9000, Rosalia.Vazquez-Alvarez@unisg.ch

Email: b.melenberg@uvt.nl, avas@uvt.nl, Rosalia.Vazquez-Alvarez@unisg.ch

We are grateful to the ISSC policy research group (Dublin University College) and seminar participants at the University of St.Gallen. We are also indebt to Michael Lechner for providing very valuable advice. This paper draws from an initial discussion paper entitled 'Nonparametric bounds in the presence of item nonresponse, unfolding brackets and anchoring', CentER discussion paper series, 01-57.

1 Introduction

Household surveys are often plagued by item nonresponse on certain variables relating to quantities. For example, in the 1996 wave of the Health and Retirement Study (HRS), a USA panel often used to study socio-economic behaviour of the elderly, 12% of the weighted sample that declare a status of paid employee do not declare their earnings when asked for an exact amount. One way to deal with nonresponse is to use information from full respondents (i.e., assuming nonresponse exogeneity). The problem is that biased estimates of the population parameters may result if nonresponse is related to the variable under study thus leading to the classic selection problem (see the seminal work by Manski, 1990, 1994 or more recent surveys in item nonresponse by Mason et al., 2002). Alternative methods to that of exogeneity (e.g., imputation procedures such as those suggested in Rubin, 1987) still require the use of alternative distributional assumptions that cannot be tested with the same data used in the process of estimation.

Manski (as leading examples, see Manski 1989, 1990, 1994, 1995, 1997, 2000) shows how bounds on conditional quantiles of the variables of interests can be derived, allowing for any type of non-random nonresponse behaviour. Manski's framework is intuitively appealing, easy to apply, and very flexible, but has the drawback that the resulting bounds are often too wide to draw meaningful economic conclusions. In Manski's framework the precision with which features of the distribution (e.g., its quantiles) can be determined, i.e., the width between the bounds, depends on the probability of nonresponse. If item nonresponse is substantial, the approach cannot lead to accurate estimates of the parameters of interest without additional information or additional assumptions. A combination of informative data sets and weak data-driven assumptions are often effective at tightening the bounds (for empirical illustrations that follow this approach, see for example, Ginther (2000), Manski and Pepper (2000), Pepper (2000) and Pepper (2003)).

Data collection strategies can also reduce the problem of item nonresponse even before the data reaches the user. For example, the use of an unfolding bracket designs to elicit partial information from initial item non-respondents. These designs drive individuals through a sequence of questions that sort individual's undisclosed amount inside a particular category in the support of the variable that suffers from nonresponse. This is done by prompting initial non-respondents to react to a sequence of 'bids'. At each bid the respondent has to reveal if they perceive the otherwise undisclosed amount to be greater (or not) than the bid. The inclusion of bracket designs is an effective way to reduce the problem of initial item nonresponse. Following with the example above (HRS, 1996) 73.4% of the weighted initial non-respondents to earnings answer the unfolding bracket design thus effectively reducing full nonresponse rate from 12% to 3% in the complete weighted sample.

In general, the use of categorical questions is motivated because it is thought that rather than the characteristics of the question, it is personal characteristics and cognitive factors (e.g., confidentiality reasons) what drives nonresponse behaviour (for issues relating to survey response behaviour see for example, Borgers and Hox, 2001, Juster and Smith, 1997, Mason et al., 2000, Rabin, 1998, or Tourangeau et al., 2000). The problem with unfolding bracket designs is that these may induce ‘anchoring effects’, a phenomenon well documented in the psychology literature (see, for example, Tversky and Kahneman, 1974, Swizer and Snizek, 1991, Jacowitz and Kahneman, 1995). In the context of the present paper, anchoring effects can be explained in the presence of bids from the unfolding design. Individuals may or may not be uncertain about the undisclosed amount (e.g., some individuals genuinely do not know the amount and other prefer not to disclose it), but different sequences of bids can induce different sets of answers if the bids act as ‘anchors’ and responses are driven in the direction of the bids. Hurd et al. (1998) show the existence of anchoring on questions relating to wealth and consumption using an experimental module from the 1996 wave of the HRS. Other experimental studies show that even if the anchor is arbitrary and uninformative this can have large effects on the responses behaviour of the sample as is the case in the experimental set up of Jacowitz and Kahneman (1995). Alternative parametric models for anchoring are introduced by Cameron and Quiggin (1994) and Herriges and Shogren (1996).

This paper extends the approach by Manski, incorporating the information provided by the bracket respondents and allowing for anchoring effects with three distinct nonparametric assumptions on anchoring. These assumptions draw from Hurd et al. (1998), Jacowitz and Kahneman (1995) and Herriges and Shogren (1996). The bounds are applied to earnings in the 1996 wave of the Health and Retirement Study. The results show that the categorical questions can be useful to increase precision of the bounds, even if anchoring is allowed for. They also help to improve the power of statistical tests for equality of earnings quantiles in sub-populations. This is shown by comparing bounds on the quantiles of earnings between genders. The bounds which take account of bracket information can detect differences which are not revealed by estimated bound based upon full respondents’ information only. We also use the bounding interval to validate the assumptions underlying an imputation procedure on earnings.

The paper is organized as follows. Section 2 derives five sets of bounds: the first ignores the presence of partial respondents; the other four allow for partial respondents but differ in the interpretation of anchoring effects. Section 3 describes the estimation process. Section 4 describes the data used in the empirical example and illustrates the bounds derived in Section 2 testing for earnings equality between genders. Section 5 concludes.

2 Theoretical framework

2.1 Worst case bounds: no bracket respondents

Manski (1989) provides the basic set up for worst case bounds on the distribution of some outcome Y at a given $y \in \mathbb{R}$ conditional on a set of covariates $X = x \in \mathbb{R}^k$. The data we observe is cross-sectional data for a sample of size n representative of the underlying population. We assume all variables in the covariate set to be free from unit and item nonresponse.¹ Furthermore, all reported (exact) values on Y and X are assumed to be correct so that misreporting of exact amounts is not an issue in this analysis. Let FR and NR indicate, respectively, full response and full nonresponse on Y . With this, the conditional distribution function of Y given $x \in X$ can be partition as follows:

$$F(y|x) = F(y|x, FR)P(FR|x) + F(y|x, NR)P(NR|x) \quad (1)$$

The assumptions imply that $F(y|x, FR)$, $P(FR|x)$ and $P(NR|x)$ are identified for all $x \in X$. The sample analogues can be estimated, for example, using some nonparametric regression technique. The identification problem arises because $F(y|x, NR)$ is not revealed by the sampling process. Assuming conditional (on X) exogeneity solves the identification problem because it implies that $F(y|x, FR) = F(y|x, NR)$, i.e., conditional on X response behaviour is independent of Y . In general, however, response behaviour is related to Y and $F(y|x, NR)$ is not identified, so that $F(y|x)$ in (1) is not identified either. Without additional assumptions, all that is known about $F(y|x, NR)$ is that it is naturally bounded between 0 and 1. Applying this to (1) gives,

$$F(y|x, FR)P(FR|x) \leq F(y|x) \leq F(y|x, FR)P(FR|x) + P(NR|x) \quad (2)$$

Expression (2) illustrates Manski's worst case bounds for the conditional distribution function. The width between bounds is $P(NR|x)$, thus, a low nonresponse rate can only improve the identification region for any $x \in X$. Additional weak data-driven assumptions can tighten the bounds as is the case of monotonicity and exclusion restrictions (see Manski, 1994, 1995).

2.2 Partial Information from an Unfolding bracket Design

In this paper bounds in (2) are extended to incorporate information from a follow-up unfolding bracket design. In this type of design initial non-respondents are faced with successive questions each

¹ For an interpretation of bounds with covariates nonresponse see Horowitz and Manski (1995,1998) and for the extension to cover covariates unfolding bracket response with anchoring see Vazquez-Alvarez (2006)

containing different amounts $\$B$ that reflect values in the support of Y . The aim is to elicit partial information by asking initial non-respondents if the otherwise undisclosed amount is greater (or not) than the bid $\$B$.² Presenting successive bids conditional on previous answers helps to classify the unknown quantity in some category in the support of Y . It is possible that in the presence of an unfolding bracket design a fraction of initial non-respondents still provide no information on the undisclosed amount: this is the case when the answer is ‘*don’t know*’ or ‘*refuse to respond*’ at the first bid in the sequence.³ Thus, even in the presence of follow-up questions the probability $P(NR) > 0$ remains. Once a decision is made to answer the first bid, individuals may or may not complete the sequence of bids. Those who complete the sequence are defined as complete bracket respondents (CBR) whereas incomplete bracket respondents (IBR) are those who declare ‘*don’t know*’ or ‘*refuse to respond*’ at some point before the sequence is over. Notice that in both cases – CBR and IBR –, the nature of information they provide is identical, i.e., both types provide partial information that determines a category in the support of Y . The two types differ because their answers define different sets of categories in Y . We want to derive bounding intervals introducing partial information through an unfolding bracket design, and this can be done without loss of generality by assuming that all bracket respondents are CBR. At the end of this section we go back to look at the effect of further introducing information from IBR within the derived bounds. With this, three types of respondents can be distinguished; full respondents (FR), full non-respondents (NR) and bracket respondents (BR) defined as those who complete the sequence of questions in the unfolding bracket design. Allowing for these subgroups of respondents, the following partition applies:

$$F(y|x) = F(y|x, FR)P(FR|x) + F(y|x, BR)P(BR|x) + F(y|x, NR)P(NR|x) \quad (3)$$

The sampling process is informative on $F(y|x, FR)$, $P(FR|x)$, $P(BR|x)$ and $P(NR|x)$ while the assumption $F(y|x, NR) \in [0,1]$ still applies. The new issue is what the answers to an unfolding bracket design say about $F(y|x, BR)$. As described in Section 1, anchoring effects may be induced when individuals’ answers are influenced by the wording in the question. To allow for this possibility we derive bounds on $F(y|x)$ first ignoring and then allowing for potential anchoring effects on $F(y|x, BR)$.

² All questions in an unfolding bracket design follow a unique direction. For example, when the first question in the design is given as ‘is the amount *greater than* B?’, the second and follow up questions also refer to ‘greater than’ although the amounts differ to become above or below B.

³ Survey response often distinguish between types of nonresponse, for example, nonresponse with ‘don’t know’ (that might be due to true inability to retrieve information) and ‘refuse’ to respond, possibly associated with the information that the interviewer tries to elicit (see, for example, Tourangeau et al, 2000). In our case it is important to notice that both types (‘don’t know’ or ‘refuse’) contribute equally to the uncertainty due to item nonresponse, thus to the width between upper and lower bounds. It is therefore not an issue to distinguish between them.

2.3 Partial Information with no anchoring effects

Let κ be a binary outcome so that $\kappa=1$ indicates a ‘yes’ answer and $\kappa=0$ indicates a ‘no’ answer to a given question. An unfolding bracket design implies that a non-respondent to an initial open-ended question faces a sequence of questions of the following nature:

$$\begin{aligned} &\text{Is the amount greater than } B_j | (B_1 = \kappa, B_2 = \kappa, \dots, B_{j-1} = \kappa)? \\ &\text{where} \\ &B_j: j\text{th bid, } j=1, \dots, J \end{aligned} \tag{4}$$

The bid in (4) implies that at each round of the design the bid faced by an individual is conditional on the individual’s previous answers; for CBR the indicator κ can take only two values 0 and 1. Thus, with J bids $\{\min(Y), (B_j | B_1 = \dots = B_{j-1} = \text{NO}), \dots, (B_j | B_1 = \dots = B_{j-1} = \text{YES}), \max(Y)\}$ defines all possible conditional bids in the support of Y . Assuming all bracket respondents face an equal number of bids and all respondents complete the sequence, the result is a set of 2^J categories in the support of the variable that suffers from nonresponse. These 2^J categories are consecutively paired and each pair is associated with either $\kappa=0$ or $\kappa=1$, the two possible answers to the last bid faced by individuals, a bid reflecting an amount that depends on each individual’s previous answers. To simplify notation let B_τ represent $B_\tau = (B_j | B_1 = \kappa, B_2 = \kappa, \dots, B_{j-1} = \kappa)$ for any $j \in J$. Assume that B_τ is the last bid faced by some initial non-respondent that has completed the sequence: answering $\kappa=1$ to B_τ leads to a lower limit in one of the 2^J categories if B_τ is the last bid. Alternatively, if the answer is $\kappa=0$ the same bid acts as an upper limit in the preceding category. With this, we can define the support of Y in terms of resulting categories: $Y = \{\min(Y), B_1, [B_1, B_2), \dots, [B_{\tau-1}, B_\tau), [B_\tau, B_{\tau+1}), \dots, [B_\tau, \max(Y))\}$ with subscript implying ordinal numbers.

Our aim is to interpret the information resulting from a sequence of ‘yes’ ($\kappa=1$) and ‘no’ ($\kappa=0$) answers in the bounding interval’s framework. Let $Q\tau$ indicate the outcome from a *sequence* of κ answers to questions in (4) so that τ implies a route that leads to the j -th bid conditional on some combined set of answers to all other $(j-1)$ preceding bids, $j \in [1, J]$. Thus, τ stands as a sequence of numbers so that $Q\tau, \tau \Rightarrow \{j, \kappa_{j-1}, \kappa_{j-2}, \dots, \kappa_1\}, \kappa_j = \{0, 1\}, j \in [1, J]$ summarizes *all* possible combinations of answers to the conditional bid $B_\tau = (B_j | B_1, B_2, \dots, B_{j-1})$ within the unfolding design. For example, Q311 results from $j=3; \tau \Rightarrow \{3, \kappa_2, \kappa_1\} = \{3, 1, 1\}$ to indicate that individuals face a 3rd bid (B_3) following a ‘yes’ answer to (4) for each of the two sequential bids B_1 and B_2 , thus $Q_\tau = Q311$ implies a question which refers to $B_\tau = B311$. If an answer is such that $Q311=1$, this means that the respondent

perceives the otherwise undisclosed amount to be greater than B_3 conditional on having perceived the undisclosed amount to be greater than B_1 and B_2 (i.e., perceives the amount to be greater than B_{311}). The event $Q_{311} = 0$ becomes the mutually exclusive event to $Q_{311} = 1$ for CBR, i.e., CBR answer either $Q\tau = 1$ or $Q\tau = 0$ at each B_τ bid, thus $P(Q\tau = 1 | BR, Q_{-\tau}) + P(Q\tau = 0 | BR, Q_{-\tau}) = 1$ where $Q_{-\tau}$ stands for the path of answers leading to $Q\tau$ and is mutual to both $Q\tau = 1$ and $Q\tau = 0$. The two mutually exclusive outcomes imply the following representation:

$$\begin{aligned} \text{Let } Y &= \{\min(Y), B_1, [B_1, B_2), \dots, [B_{\tau-1}, B_\tau), [B_\tau, B_{\tau+1}), \dots, [B_{T-1}, B_T)\}, \\ \text{where } B_{\tau-1} &< B_\tau < B_{\tau+1} \end{aligned}$$

$$\text{Let } Q\tau = \kappa, \quad \kappa = \{0, 1\} \text{ be associated with } B_\tau \tag{5}$$

$$\begin{aligned} P(Q\tau = \kappa | x, BR) &= \\ &P(Q\tau = \kappa | Y < B_\tau, x, BR)P(Y < B_\tau | x, BR) \\ &+ P(Q\tau = \kappa | Y \geq B_\tau, x, BR)P(Y \geq B_\tau | x, BR) \end{aligned}$$

Expression (5) allows for a direct interpretation of individual's answers to the bids, that is, we interpret $P(Q\tau = \kappa | \dots)$, the reported fraction, as opposed to directly looking at $P(Y \leq y | \dots)$, the true cumulative fraction. Introducing $P(Q\tau = \kappa | \dots)$ is crucial for us to deal with anchoring effects. This is because the bid $B_\tau = (B_j | B_1, B_2, \dots, B(j-1))$ associated with the two *potential* outcomes $Q\tau = \kappa, \kappa = \{0, 1\}$ may act as an anchor and affect the *actual* outcome. Thus, anchoring implies the possibility that even when the true (but undisclosed) amount is $Y < B_\tau$ ($\geq B_\tau$), the answer becomes $Q\tau = 1$ ($Q\tau = 0$). In the absence of anchoring effects the following applies:

Assumption A0:

Let $Q\tau = \kappa$, $\kappa = \{0,1\}$ be associated with B_τ

No anchoring and the answer is $\kappa = 1 \Rightarrow P(Q\tau = 1 | Y < B_\tau, x, BR) = 0$

from (5) $\Rightarrow P(Q_\tau = 1 | x, BR) = P(Y \geq B_\tau | x, BR)$ (i)

No anchoring and the answer is $\kappa = 0 \Rightarrow P(Q\tau = 0 | Y \geq B_\tau, x, BR) = 0$

from (5) $\Rightarrow P(Y < B_\tau | x, BR) = 1 - P(Q_\tau = 1 | x, BR)$ (ii)

From (i) & (ii), for $y < B_\tau$: $F(y | x, BR) \leq \sum_{i=1}^{r-1} P(Q_i = 0 | x, BR)$

where $i = 1 \Rightarrow$ 'lowest bid' (ordinal) such that

$i < 1 \Rightarrow P(Q_{i < 0} = 0 | x, BR) = 0, \rightarrow \min(Y)$

and $i > T \Rightarrow \sum_{i=0}^T P(Q_i = 0 | x, BR) = 1, \rightarrow \max(Y)$

(6)

Thus, in the absence of anchoring effects (6) suggests an exact link between reported information $P(Q_\tau = \kappa | \dots)$, and the true conditional distribution at the bid B_τ . It is important to notice that the probabilities in (6) do not condition on $Q_{-\tau}$. For example, we consider $P(Q\tau = 1 | Y < B_\tau, x, BR) = 0$ and not $P(Q\tau = 1 | Y < B_\tau, x, BR, Q_{-\tau}) = 0$: the latter implies 'a correct answer' to B_τ conditional on having reached that bid through the path $Q_{-\tau}$, and it is therefore a weaker assumption than the one implied by $P(Q\tau = 1 | Y < B_\tau, x, BR) = 0$; the latter assumes that all bracket respondents answer correctly to all Q_τ questions and for all paths that lead to such question.

We know that each CBR is associated with a category $[B_{\tau-1}, B_\tau)$, $B_\tau > B_{\tau-1}$ thus locating the undisclosed amount between $B_{\tau-1}$ (i.e., $Q_{\tau-1} = 1$) and B_τ (i.e., $Q_\tau = 0$), that is:

Let $[B_{\tau-1}, B_\tau)$ define a category in Y resulting from partial respondents such that $Q_\tau = 0$ and $Q_{\tau-1} = 1$,

From (6): $Q_{\tau-1} = 1 \Rightarrow Q_{\tau-1} = 0$ defined :

$$\begin{aligned} \Rightarrow \quad & \text{for } y < B_{\tau-1} : \quad F(y | x, BR) \leq \sum_{i=1}^{\tau-1} P(Q_i = 0 | x, BR) \\ & Q_\tau = 0 \\ \Rightarrow \quad & \text{for } y < B_\tau : \quad F(y | x, BR) \leq \sum_{i=1}^\tau P(Q_i = 0 | x, BR) \end{aligned} \quad (7)$$

For $y \in [B_{\tau-1}, B_\tau)$

$$L_{\tau-1, \tau} = \sum_{i=1}^{\tau-1} P(Q_i = 0 | x, BR) \leq F(y | x, BR) \leq U_{\tau-1, \tau} = \sum_{i=1}^\tau P(Q_i = 0 | x, BR)$$

where $\min(L_{\tau-1, \tau}) = 0, \quad \max(U_{\tau-1, \tau}) = 1$

The result in (7) implies cumulative bounding interval for the unknown $F(y | x, BR)$ at each of the resulting categories from an unfolding bracket design. Clearly, (7) also defines a range of values if we are looking at some partition $y \geq B$ or $y \leq B$ in the support of Y as long as B is one of the conditional bids in the sequence. Bounds on $F(y | x, BR)$ given by (7) ignore anchoring effects.

2.4 Partial Information allowing for anchoring effects

Anchoring effects implies that answers to (4) could be affected by the information contained in the bids, i.e., the bids act as an ‘anchor’ and information from partial respondents identifies the conditional distribution of Y possibly with bias due to anchoring effects. In practical terms anchoring implies that partial respondents may classify the otherwise undisclosed $y \in Y$ in $[B_{\tau-1}, B_\tau)$ when in fact $y \notin [B_{\tau-1}, B_\tau)$, thus, (i) and (ii) in (6) no longer apply. The following modifies (6):

Let $Q\tau = \kappa, \quad \kappa = \{0, 1\}$ be associated with B_τ

The answer is $\kappa = 1$ but Anchoring may apply:

$$\begin{aligned} & P(Q\tau = 1 | Y < B_\tau, x, BR) \neq 0 \\ \text{from (5)} \Rightarrow & P(Q_\tau = 1 | x, BR) \neq P(Y \geq B_\tau | x, BR) \end{aligned} \quad (8)$$

The answer is $\kappa = 0$ but Anchoring may apply:

$$\begin{aligned} & P(Q\tau = 0 | Y \geq B_\tau, x, BR) \neq 0 \\ \text{from (5)} \Rightarrow & P(Q_\tau = 0 | x, BR) \neq P(Y < B_\tau | x, BR) \end{aligned}$$

Compared to expression (6), expression (8) shows that in the presence of anchoring there is a break down on the direct link between information reported by unfolding bracket respondents, i.e.,

$P(Q\tau = \kappa \mid BR, \dots)$, $\kappa = \{0, 1\}$, and the true conditional distribution of Y , i.e., $P(Y < B_\tau \mid BR, \dots)$ for any B_τ in the unfolding bracket design.

To allow for anchoring effects we need to modify (7) with respect to the implications in (8). In the presence of a single unfolding bracket design (i.e., a single set of bids) anchoring cannot be tested for. Instead, anchoring can be modelled according to assumptions. This paper examines three distinct assumptions to introduce anchoring in the bounding interval approach. The first assumption draws from Hurd et al. (1998) where anchoring enters parametrically as a perception error: we relax the parametric assumption in favour of a semiparametric alternative. The second anchoring assumption takes the finding in the experimental study of Jacowitz and Kahneman (1995) to allow for anchoring effects through weak data-driven assumptions. The third anchoring assumption draws from Herriges and Shogren (1996) where it is assumed that the order of the bids is important to define the effects of anchoring. In what follows each of these three models modifies (7).

A.1 Anchoring effects drawing from Hurd et al., (1998)

Hurd et al. (1998) use an experimental module with data collected during the second wave of the Health and Retirement Study (HRS, 1996) to detect anchoring effects.⁴ The module consists on randomizing different sets of bids (i.e., different unfolding bracket designs) to elicit partial information on earnings and consumption data. According to Hurd et al. (1996) initial non-respondents may face uncertainty about the quantity $y \in Y$ when faced with a question as given in (4) and they resolve this uncertainty by comparing Y to $(B_\tau + \varepsilon_\tau)$ where ε_τ is the ‘perception error’ associated with the bid B_τ .⁵ Hurd et al. (1998) assume that ε_τ is normally distributed with zero mean and independent of Y given X . We modify their parametric set up to allow for a semiparametric median assumption on the perception error:

⁴ The publicly released second wave of the Health and Retirement Study (HRS, 1996) does not include specific module, e.g., does not include the module with randomized unfolding bracket designs. In practice, surveys that rely on unfolding bracket designs (e.g., HRS, SHARE, etc) do not randomize respondents but rely on a single design to cover all initial non-respondents.

⁵ The description also covers individuals who may have exact full knowledge on the missing value but refuse to answer with an exact amount. These individuals are also given the chance to answer question as in (4) that imply a problem solving situation. Once the bid B_τ is presented, this is information they need to take into account (together with the exact knowledge they might have on the unobserved missing value) when solving the problem through an unobserved complex cognitive process. Clearly, the bid B_τ can have an effect even if knowledge on the missing value is exact (e.g., the phenomenon of Yea-saying as explained in Green et al., 1998).

Assumption A.1

$$\begin{aligned}
& \text{Let } Q\tau = \kappa, \quad \kappa = \{0,1\} \text{ be associated with } B_\tau \\
& \Rightarrow \quad Q\tau = 1 \quad \text{if } Y \geq (B_\tau + \varepsilon_\tau) \\
& \quad \quad Q\tau = 0 \quad \text{if } Y < (B_\tau + \varepsilon_\tau) \\
& \text{where, for all } (x, y) \text{ in the support } (Y, X),
\end{aligned} \tag{9}$$

$$\text{median}(\varepsilon_\tau \mid y \in Y, x \in X, BR, Q_{-\tau}) = 0$$

The median assumption implies that the conditional probability that an individual answers question Q_τ correctly (with either yes or no; $\kappa = (1, 0)$) is at least 0.5. Notice that in this case we have conditioned on the event $Q_{-\tau}$ meaning that ‘*answering correctly (or not)*’ is an event revised at each bid in the design (i.e., at each node in the unfolding tree). From (5) the events $P(Q\tau = \kappa \mid Y < B_\tau, x, BR)$ and $P(Q\tau = \kappa \mid Y \geq B_\tau, x, BR)$ are used to understand the ‘no anchoring effect’. But (5) no longer applies because ambiguity may arise if we use these two probabilities to interpret anchoring effects. To see this, we notice from (8) that anchoring effects imply $P(Q\tau = 0 \mid Y \geq B_\tau, x, BR) \neq 0$. However if anchoring occurs prior to Q_τ the assumption $P(Q\tau = 0 \mid Y \geq B_\tau, x, BR) = 0$ can still be compatible with anchoring effects.⁶ Conditioning on the event $Q_{-\tau}$ ensures that anchoring effects are accounted for at each bid in the design so that (8) also apply. The consequence is that we need to consider a new interpretation of (5):

$$\begin{aligned}
& \text{Let } Y = \{[\min(Y), B_1], [B_1, B_2], \dots, [B_{\tau-1}, B_\tau], [B_\tau, B_{\tau+1}], \dots, [B_{T-1}, B_T]\}, \\
& \text{where } B_{\tau-1} < B_\tau < B_{\tau+1} \\
& \text{Let } Q\tau = \kappa, \quad \kappa = \{0,1\} \text{ be associated with } B_\tau \\
& P(Q\tau = \kappa \mid x, BR, Q_{-\tau}) = \\
& \quad P(Q\tau = \kappa \mid Y < B_\tau, x, BR, Q_{-\tau}) P(Y < B_\tau \mid x, BR, Q_{-\tau}) \\
& \quad + P(Q\tau = \kappa \mid Y \geq B_\tau, x, BR, Q_{-\tau}) P(Y \geq B_\tau \mid x, BR, Q_{-\tau})
\end{aligned} \tag{10}$$

Applying (9) to (10) makes Assumption A.1 operational whitening a bounding interval approach while allowing for anchoring effects:

⁶ For example, an individual correctly identifies the event Q_τ with a ‘no’ answer so that $Q_\tau = 0$ if $Y < B_\tau$ is true; however, it may be that answers to some bid $B_{\tau-1} < B_\tau$ where $Y < B_{\tau-1}$ applies is such that $P(Q_{\tau-1} = 1 \mid Y \geq B_{\tau-1}, x, BR) \neq 0$, thus anchoring has been induced previous to Q_τ , and therefore anchoring would be compatible with the event $P(Q_\tau = 0 \mid Y < B_\tau, x, BR) = 0$ that we associate with a no anchoring assumption.

Apply Assumption A.1 to all B_τ :

$\Rightarrow$ Answering $\kappa=1$ and Anchoring:

$$\begin{aligned} P(Q_\tau=1|Y < B_\tau, x, BR, Q_{-\tau}) &= P(\varepsilon_\tau \leq (Y - B_\tau) | (B_\tau - Y) > 0, x, BR, Q_{-\tau}) \leq 0.5 \quad (i) \\ P(Q_\tau=1|Y \geq B_\tau, x, BR, Q_{-\tau}) &= P(\varepsilon_\tau \leq (Y - B_\tau) | (B_\tau - Y) \leq 0, x, BR, Q_{-\tau}) \geq 0.5 \quad (ii) \end{aligned} \quad (11.a)$$

$\Rightarrow$ Answering $\kappa=0$ and Anchoring:

$$\begin{aligned} P(Q_\tau=0|Y < B_\tau, x, BR, Q_{-\tau}) &= P(\varepsilon_\tau \leq (Y - B_\tau) | (B_\tau - Y) > 0, x, BR, Q_{-\tau}) \geq 0.5 \quad (iii) \\ P(Q_\tau=0|Y \geq B_\tau, x, BR, Q_{-\tau}) &= P(\varepsilon_\tau \leq (Y - B_\tau) | (B_\tau - Y) \leq 0, x, BR, Q_{-\tau}) \leq 0.5 \quad (iv) \end{aligned}$$

Conditions in (11.a) are applicable to expression (10):

$\Rightarrow$ Apply (11.a.i) to (10):

$$\begin{aligned} \text{From (11.a.i)} \rightarrow P(Q_\tau=1|x, BR, Q_{-\tau}) &\leq 0.5P(Y < B_\tau | x, BR, Q_{-\tau}) + P(Y \geq B_\tau | x, BR, Q_{-\tau}) \\ &\leq 1 - 0.5P(Y < B_\tau | x, BR, Q_{-\tau}) \end{aligned}$$

$$\begin{aligned} \Rightarrow P(Y < B_\tau | x, BR, Q_{-\tau}) &\leq 2P(Q_\tau=0|x, BR, Q_{-\tau}) \quad (i^*) \\ &\text{(the result is identical if we apply (11.a.iii) to (10))} \end{aligned}$$

(11b)

$\Rightarrow$ Apply (11.a.ii) to (10):

$$\text{From (11.a.ii)} \rightarrow P(Q_\tau=1|x, BR, Q_{-\tau}) \geq 0.5P(Y \geq B_\tau | x, BR, Q_{-\tau})$$

$$\begin{aligned} \Rightarrow P(Y \geq B_\tau | x, BR, Q_{-\tau}) &\leq 2P(Q_\tau=1|x, BR, Q_{-\tau}) \quad (ii^*) \\ &\text{(the result is identical if we apply (11.a.iv) to (10))} \end{aligned}$$

Bounds on $P(Y < B_\tau | \dots)$ and $P(Y \geq B_\tau | \dots)$ given by (i^*) and (ii^*) apply equally to $Q_{-\tau} = \kappa$ for $\kappa = 0, 1$. These, together with a restriction that compares a probability $P(A|C, B)$ to $P(A|C)$ leads to new conditions to substitute those in (6) thus allowing for anchoring with respect to a semiparametric interpretation of Hurd et al. (1998):

From (i*) in (11.b):

$$P(Y < B_\tau | x, BR, Q_{-\tau} = 1) \leq 2P(Q_\tau = 0 | x, BR, Q_{-\tau} = 1) \quad (12.a)$$

$$P(Y < B_\tau | x, BR, Q_{-\tau} = 0) \leq 2P(Q_\tau = 0 | x, BR, Q_{-\tau} = 0) \quad (12.b)$$

From (ii*) in (11.b):

$$P(Y \geq B_\tau | x, BR, Q_{-\tau} = 1) \leq 2P(Q_\tau = 1 | x, BR, Q_{-\tau} = 1) \quad (12.c)$$

$$P(Y \geq B_\tau | x, BR, Q_{-\tau} = 0) \leq 2P(Q_\tau = 1 | x, BR, Q_{-\tau} = 0) \quad (12.d) \quad (12)$$

Also, for either $G = (Y < B_\tau)$ or $(Y \geq B_\tau)$,

$$\begin{aligned} & P(G | BR, Q_{-\tau}, Q_\tau = \kappa) P(Q_\tau = \kappa | BR, Q_{-\tau}) \\ &= P(G | BR, Q_{-\tau}) P(Q_\tau = \kappa | BR, Q_{-\tau}, G) \\ &\leq P(G | BR, Q_{-\tau}) P(Q_\tau = \kappa | BR, Q_{-\tau}) \quad \text{i.e., } P(Q | Z, G) \leq P(A | Z) \end{aligned} \quad (12.e)$$

We use conditions (12.a) – (12.e) to derived bounds on $F(y | BR, x)$, either for some region $y \geq B_\tau$ (or $y < B_\tau$), or for the cumulative distribution function at different categories $y \in [B_{\tau-1}, B_\tau)$. Conditions (12.a) – (12.d) imply restrictions on $F(y | x, BR)$ based on the probability of an answer to Q_τ conditional on having answered to $Q_{-\tau}$. With this, Appendix 1 shows that (12.a) – (12.e) lead to the following bounds on $F(y | x, BR)$:

Let $B_\tau \in Y$ be a bid in the design, and $Q_{-\tau}$ the question that being associated with $B_{-\tau}$ leads to B_τ . The following is always true:

$$\text{if } Q_{-\tau} = 0 \rightarrow B_{-\tau} > B_\tau; \quad \text{if } Q_{-\tau} = 1 \rightarrow B_{-\tau} < B_\tau$$

If $B_{-\tau} > B_\tau$, for $y < B_\tau$:

$$\max[0, 1 - 2P(Q_\tau = 1 | x, BR, Q_{-\tau} = 0)] \leq F(y | x, BR) \leq \min[1, 2P(Q_\tau = 0 | x, BR, Q_{-\tau} = 0)] \quad (13.a)$$

If $B_{-\tau} < B_\tau$, for $y < B_\tau$:

$$\max[0, 1 - 2P(Q_\tau = 1 | x, BR, Q_{-\tau} = 1)] \leq F(y | x, BR) \leq \min[1, 2P(Q_\tau = 0 | x, BR, Q_{-\tau} = 1)] \quad (13.b)$$

Let $l(B_\tau) \leq F(y | x, BR) \leq u(B_\tau)$ & $l(B_{\tau-1}) \leq F(y | x, BR) \leq u(B_{\tau-1})$ represent the identification region (either (13.a) or (13.b)) for both bids $B_{\tau-1}$ and B_τ :

$$\begin{aligned} & \Rightarrow \text{for } y \in [B_{\tau-1}, B_\tau), \quad L_{\tau-1,\tau} \leq F(y | x, BR) \leq U_{\tau-1,\tau} \\ \text{where} \quad & L_{\tau-1,\tau} = \min\{l(B_\tau), l(B_{\tau-1})\} \quad \& \quad U_{\tau-1,\tau} = \max\{u(B_\tau), u(B_{\tau-1})\} \end{aligned} \quad (13)$$

The resulting bounds in (13) substitute those in (7) to allow for anchoring effects according to the median interpretation of the assumptions in Hurd et al.(1998). Any combination between (13.a) and (13.b) is such that $(U_{\tau-1,\tau} - L_{\tau-1,\tau}) > 0$.⁷ However, bounds in (13) are informative if and only if $(U_{\tau-1,\tau} - L_{\tau-1,\tau}) < 1$.

A.2 Anchoring effects drawing from Jacowitz and Kahneman (1995)

Jacowitz and Kahneman (1995) conduct an experimental study to show that inserting either low or high anchors in estimation tasks significantly affects individual answer's in the direction of the anchors. Their empirical findings provide a plausible and intuitive account of anchoring to explain (rather than parametrically model) the existence of a perception error and its direction.

The empirical findings in Jacowitz and Kahneman (1995) suggest that in the presence of a high anchor respondents too often report that the otherwise unknown amount exceeds the anchor; the same study finds that a low anchor drives individual's answers to lower the median estimate of the quantity (when compared to the calibration group who are not subject to anchors). For example, when asked to evaluate the distance between San Francisco and New York, the median for the calibration group was 3,600 miles. Instead, when the group treated with a high anchor (i.e., '*is the distance greater than 6,000 miles?*') were subsequently asked to estimate the distance, they recorded a median of 4,000 miles, while those treated to a low anchor (i.e., *1,500 miles*) recorded a group median of 2,600 miles, (Jacowitz and Kahneman, 1995, page 1163).⁸ In their study the finding suggest that high anchors lead to significantly larger effects than low anchors, but both low and high anchors are effective and significant at pulling individuals' answers towards the anchor. The following set up implies an interpretation of the findings by Jacowitz and Kahneman (1995):

⁷ For $\min[l(B_{\tau-1}), l(B_{\tau})] = l(B_{\tau-1})$ and $\max[u(B_{\tau-1}), u(B_{\tau})] = u(B_{\tau}) \Rightarrow u(B_{\tau}) > u(B_{\tau-1})$, a positive pointwise distance $(U_{\tau-1,\tau} - L_{\tau-1,\tau})$ is guaranteed if $(u(B_{\tau-1}) - l(B_{\tau-1}))$ is positive, which is always true since from either (13.a) or (13.b) it is very easy to see that that $(u(B_{\tau-1}) - l(B_{\tau-1})) > 0$. Similar arguments apply if we examine the case when $\min[l(B_{\tau-1}), l(B_{\tau})] = l(B_{\tau})$ and $\max[u(B_{\tau-1}), u(B_{\tau})] = u(B_{\tau-1}) \Rightarrow u(B_{\tau-1}) > u(B_{\tau})$. For the remaining two cases (i.e., $\min[l(..), l(..)] = l(B)$, $\max[u(..), u(..)] = u(B)$; $B = B_{\tau-v}$, $v = 0, 1$) the distance is unambiguously positive and the bounds are unambiguously informative (that is, the pointwise distance is always below in $(0,1)$).

⁸ The low and high anchor were determined as upper and lower percentiles of the median recorded by the calibration group who were asked to evaluate the quantity in the absence of anchors. The example is one in 15 measurement tasks in page 1163 of Jacowitz and Kahneman (1995).

Let the bid B_{τ} be perceived as a HIGH anchor: Respondents too often will report that the amount exceeds the anchor:

$$\Rightarrow B_{\tau} \text{ 'high': } P(Q_{\tau} = 1 | x, BR, Q_{-\tau}) \geq P(Y > B_{\tau} | x, BR, Q_{-\tau}) \quad (14.a) \quad (14)$$

Let the bid B_{τ_i} be perceived as a LOW anchor: Respondents too often will report that the amount falls below the anchor:

$$\Rightarrow B_{\tau} \text{ 'low': } P(Q_{\tau} = 1 | x, BR, Q_{-\tau}) \leq P(Y > B_{\tau} | x, BR, Q_{-\tau}) \quad (14.b)$$

To make (14) operational one needs to make further assumptions so to distinguish between ‘perceived’ low and high anchors. The following weak data assumption distinguishes between high and low anchors within an unfolding bracket design:

Assumption. A.2

B_{τ} is HIGH and acts as a HIGH ANCHOR if the following is observed:

$$P(Q_{\tau} = 1 | x, BR, Q_{-\tau}) \leq 0.5 \quad (15.a) \quad (15)$$

B_{τ} is LOW and acts as a LOW ANCHOR if the following is observed:

$$P(Q_{\tau} = 1 | x, BR, Q_{-\tau}) \geq 0.5 \quad (15.b)$$

Assumption A.2 shows that only ‘yes’ answers (i.e., $Q_{\tau} = 1$) matter to determine the type of anchor. Together, (14) and (15) can be used to place bounds on $F(y | x, BR)$ with such bounds allowing for potential anchoring effects according to this second alternative explanation of anchoring. However, since $y \in [B_{\tau-1}, B_{\tau})$ is the result of the unfolding bracket design, the effect that anchoring may have on the sample classified in $[B_{\tau-1}, B_{\tau})$ is the combined effect that the two bids have on individual’s answers. We start by looking at what happens when a bid B_{τ} is perceived as a high anchor:

B_r is perceived as a HIGH anchor:

i.e.,

$$\begin{aligned} \text{Observing} \quad & P(Q_r = 1 | x, BR, Q_{-r}) \leq 0.5 \\ \Rightarrow \quad & P(Y > B_r | x, BR, Q_{-r}) \leq P(Q_r = 1 | x, BR, Q_{-r}) \quad \text{from (14.a)} \\ & \text{where } P(Y > B_r | x, BR, Q_{-r}) \text{ is unknown (given anchoring).} \end{aligned}$$

then,

$$\begin{aligned} & P(Y > B_r | x, BR, Q_{-r}) \leq P(Q_r = 1 | x, BR, Q_{-r}) \\ \Rightarrow \quad & P(Y \leq B_r | x, BR, Q_{-r}) \geq P(Q_r = 0 | x, BR, Q_{-r}) \\ \Rightarrow \quad & P(Y \leq B_r | x, BR) \geq P(Q_r = 0 | x, BR, Q_{-r}) \quad \text{because } P(A | Z, C) \leq P(A | Z) \\ \Rightarrow \quad & P(Q_r = 0 | x, BR, Q_{-r}) \leq P(Y \leq B_r | x, BR) \leq 1 \quad (16.a) \end{aligned} \tag{16}$$

Expression (16) shows an upper bound of 1 for the unknown probability $P(Y \leq B_r | PR)$; this simply reflects the fact that the assumption $P(Y > B_r | x, BR, Q_{-r}) \leq P(Q_r = 1 | x, BR, Q_{-r})$ provides no information to bound $P(Y \leq B_r | BR)$ from above. We now look at what happens if B_r is perceived as a low anchor:

B_r is perceived as a LOW bid:

i.e.,

$$\begin{aligned} \text{Observing} \quad & P(Q_r = 1 | x, BR, Q_{-r}) \geq 0.5 \\ \Rightarrow \quad & P(Y > B_r | x, BR, Q_{-r}) \geq P(Q_r = 1 | x, BR, Q_{-r}) \quad \text{from (14.b)} \\ & \text{where } P(Y > B_r | x, BR, Q_{-r}) \text{ is unknown (given anchoring).} \end{aligned}$$

then,

$$\begin{aligned} & P(Y > B_r | x, BR, Q_{-r}) \geq P(Q_r = 1 | x, BR, Q_{-r}) \\ \Rightarrow \quad & P(Y > B_r | x, BR) \geq P(Q_r = 1 | x, BR, Q_{-r}) \quad \text{because } P(A | Z, C) \leq P(A | Z) \\ \Rightarrow \quad & P(Y \leq B_r | x, BR) \leq P(Q_r = 0 | x, BR, Q_{-r}) \\ \Rightarrow \quad & 0 \leq P(Y \leq B_r | x, BR) \leq P(Q_r = 0 | x, BR, Q_{-r}) \quad (17.a) \end{aligned} \tag{17}$$

Expression (17) shows a low bound of 0 for the unknown $P(Y \leq B_r | x, BR)$ because, similar to the case in (16), in the presence of a low bid B_r the assumptions that allow for anchoring effects are only informative with regards to the upper bound on $P(Y \leq B_r | x, BR)$. With (16.a) and (17.a) we can derive bounds on $F(y | x, BR)$:

Let $y \in [B_{\tau-1}, B_\tau]$;

The sampling process is informative on both, $B_{\tau-1}$ and B_τ up to a bounding interval defined in either (16.a) or (17.a), therefore:

$$\begin{aligned} L(B_\tau) &\leq P(A \leq B_\tau | x, BR) \leq U(B_\tau) && \text{by either (16.a) or (17.a)} \\ L(B_{\tau-1}) &\leq P(A \leq B_{\tau-1} | PR) \leq U(B_{\tau-1}) && \text{by either (16.a) or (17.a)} \end{aligned} \quad (18)$$

$\Rightarrow$

for $y \in [B_{\tau-1}, B_\tau]$ $L_{\tau-1,\tau} \leq F(y | x, BR) \leq U_{\tau-1,\tau}$

where

$$L_{\tau-1,\tau} = \min\{L(B_\tau), L(B_{\tau-1})\} \quad \& \quad U_{\tau-1,\tau} = \max\{U(B_\tau), U(B_{\tau-1})\}$$

Compared to expression (7) the results in (18) show that allowing for anchoring effects implies a new interval to identify $F(y | x, BR)$. The interval in (18) implies a width equal to $(U_{\tau-1,\tau} - L_{\tau-1,\tau})$. For $B_\tau > B_{\tau-1}$ the width is always positive.⁹

A.3 Anchoring effects drawing from Herriges and Shogren (1996)

The Herriges and Shogren (1996) model offers an alternative explanation for the shift in the estimated distribution based upon unfolding bracket questions due to the order of the bids (the main empirical finding in Hurd et al., 1998). On the other hand, the Herriges and Shogren model cannot explain the main findings of Jacowitz and Kahneman (1995) since in the latter the findings relate to the first bid (i.e., subjects in the experiment receive only one bid), for which the Herriges and Shogren (1996) model impose the ‘no anchoring’ assumption. The set up in Herriges and Shogren (1996) makes use of a parametric interpretation to compare the effect of the first bid on all other follow up bids. Let B_1 be the first bid that serves as anchor for the second bid, say B_2 (which can be either higher or lower than B_1 according to what individuals answer to Q_1). Thus, in the second bracket question the respondent does

⁹ This can be easily shown: Write $F(y | x, BR) \in \{\min[L_{r-1}, L_r], \max[U_{r-1}, U_r]\}$. By definition $(U_r - L_r) > 0$ and $(U_{r-1} - L_{r-1}) > 0$. Thus ambiguity can only occur when $F(y | x, PR) \in \{\min = L_{r-1}, \max = U_r\}$ or when $F(y | x, BR) \in \{\min = L_r, \max = U_{r-1}\}$. If both B_{r-1} and B_r are perceived as high bids (according to A.2), using conditions (16) and (17) shows that in the case of $\{\min = L_{r-1}, \max = U_r\} = \{P(Q_{r-1} = 0 | \dots), 1\}$ and in the case of $\{\min = L_r, \max = U_{r-1}\} = \{P(Q_r = 0 | \dots), 1\}$. In both cases the width between upper and lower limits is positive and the bounds are informative when the involved probabilities are positive. If both B_{r-1} and B_r are perceived as low bids, $\{\min = L_{r-1}, \max = U_r\} = \{0, P(Q_r = 0 | \dots)\}$ and $\{\min = L_r, \max = U_{r-1}\} = \{0, P(Q_{r-1} = 0 | \dots)\}$, so that once more the distance between paired upper and lower limits is positive: in these two cases the bounds are unambiguously informative. Finally, if B_{r-1} and B_r are perceived as low and high, respectively, $\{\min = L_{r-1}, \max = U_r\} = \{0, 1\}$ and $\{\min = L_r, \max = U_{r-1}\} = \{0, 1\}$: in these two cases the bounds are not informative because the distance is identical to the natural distance in the presence of no information; $F(y | x, BR) \in [0, 1]$ would apply as in the case of NR .

not compare B_2 to Y , but to $Y^* = (1-g) + gB_1$. This reflects the intuition behind anchoring: the respondent is uncertain about the true value of Y . The bid B_1 is taken as informative about Y and the respondent's new estimate Y^* is drawn towards B_1 . Herriges and Shogren assume that g is a fixed parameter but also discuss an extension in which g can vary with B_1 . They apply their model to data on WTP for water quality improvement, and find an estimate for g of 0.36 with standard deviation of 0.14. Assuming an unfolding bracket design with J bids, it would be natural to consider that a preceding bound $B_{-\tau}$ always acts as informative about Y . Thus, the respondent's new valuation to Q_τ with respect to the bid B_τ is drawn towards $B_{-\tau} < B_\tau$.¹⁰ Instead of defining concurrent values of g_τ and allow for a parametric set up as in Herriges and Shogren (1996), we relax this parametric assumption to allow for a nonparametric interpretation of anchoring in Herriges and Shogren (1996) that applies to a design with potentially $J > 2$ number of bids, where $B_1(\tau=1)$ defines the first bid in the design according to (4). This leads to Assumption A3:

Assumption A.3

*"There is no anchoring at B_1 , the first bid in an unfolding bracket design.
For any bid $B_\tau > B_1$, anchoring implies that the answer to the question Q_τ
is pulled towards the preceding bid $B_{-\tau}$ "*

With Assumption A3, (19a), (19.b) and (19c) apply:

At B_1 , no anchoring :

$$P(Y \geq B_1 | x, BR) = P(Q_1 = 1 | x, BR) \Rightarrow P(Y < B_1 | x, BR) = P(Q_1 = 0 | x, BR) \quad (19.a)$$

¹⁰ An empirical application of the principle in Herriges and Shogren is found in O'Connor et al., (1999), who also find a similar significant positive value of g .

For all other B_τ following from B1, define $Q_\tau = \kappa$, $\kappa = (0,1)$

$$Q_\tau = 0 \Rightarrow B_{\tau+1} < B_\tau; \quad P(Y \geq B_{\tau+1} | x, BR, Q_\tau = 1) \leq P(Q_{\tau+1} = 1 | x, BR, Q_\tau = 0) \quad (19.b)$$

i.e., Individuals respond 'no' to B_τ , and when faced with a $B_{\tau+1}$ they respond so that on average there is a movement towards B_τ , and the unknown $P(Y \geq B_{\tau+1} | \dots)$ is 'overestimated' by the proportion $P(Q_{\tau+1} = 1 | \dots)$.

$$\begin{aligned} Q_\tau = 1 \Rightarrow B_{\tau+1} > B_\tau; \quad & P(Y \geq B_{\tau+1} | x, BR, Q_\tau = 1) \geq P(Q_{\tau+1} = 1 | x, BR, Q_\tau = 1) \\ \Rightarrow & P(Y < B_{\tau+1} | x, BR, Q_\tau = 1) \leq P(Q_{\tau+1} = 0 | x, BR, Q_\tau = 1) \end{aligned} \quad (19.c)$$

i.e., Individuals respond 'yes' to B_τ , and when faced with a $B_{\tau+1}$ they respond so that there is a movement towards B_τ , and the unknown $P(Y \geq B_{\tau+1} | \dots)$ is 'undermined' by the proportion $P(Q_{\tau+1} = 1 | \dots)$.

The results (19.a) to (19.c) implied by A3 are the only conditions used to find bounds on $F(y | x, BR)$. First we apply A3 to derive bounds on $F(y | x, BR)$ for $y < B_{\tau+1}$ where $B_\tau < B_{\tau+1}$ and B_τ stands for the 'preceding' bid:

$$\text{Assume } B_\tau \text{ is bounded such that } l(B_\tau) \leq (P(Y < B_\tau | x, BR)) \quad (*)$$

$\Rightarrow$

$$\text{If } Q_\tau = 1 \rightarrow B_{\tau+1} > B_\tau \text{ and (19.c) applies}$$

Upper Bound :

$$\begin{aligned} P(Y < B_{\tau+1} | x, BR) &= P(Y < B_{\tau+1} | x, BR, Q_\tau = 0)P(Q_\tau = 0 | x, BR) \\ &\quad + P(Y < B_{\tau+1} | x, BR, Q_\tau = 1)P(Q_\tau = 1 | x, BR) \\ &\leq P(Q_\tau = 0 | x, BR) + P(Q_{\tau+1} = 0 | x, BR, Q_\tau = 1)P(Q_\tau = 1 | x, BR) \end{aligned}$$

Lower Bound

$\Rightarrow$ (19.c) does not provide information to bound $P(Y < B_{\tau+1} | x, BR)$ from below (i.e., lower bound). But (*) implies,

$$l(B_\tau) \leq P(Y < B_\tau | x, BR)$$

$$\text{then, given that } B_{\tau+1} > B_\tau \Rightarrow l(B_\tau) \leq P(Y < B_{\tau+1} | x, BR)$$

Lower & Upper bound at the bid $B_{\tau+1} (> B_\tau)$

$$\begin{aligned} l(B_\tau) &\leq P(Y < B_{\tau+1} | x, BR) \leq \\ &P(Q_\tau = 0 | x, BR) + P(Q_{\tau+1} = 0 | x, BR, Q_\tau = 1)P(Q_\tau = 1 | x, BR) \end{aligned} \quad (20)$$

The outcome in (20) requires assuming that $P(Y < B_{\tau+1} | x, BR, Q_\tau = 0) = 1$. The reason is that (19.c) is derived because those answering $Q_\tau = 1$ and leading to $B_{\tau+1}$ answer to $Q_{\tau+1}$ with answers pulled towards B_τ so that $P(Y \geq B_{\tau+1} | x, BR, Q_\tau = 1)$ is undermined while the conditional distribution $P(Y < B_{\tau+1} | x, BR, Q_\tau = 0)$ is ‘over-valued’. Thus, assuming $P(Y < B_{\tau+1} | x, BR, Q_\tau = 0) = 1$ is a correct interpretation that takes this ‘over-valuation’ into account. Next, we apply Assumption A3 to derive bounds on $F(y | x, BR)$ for $y < B_{\tau-1}$ where $B_{\tau-1} < B_\tau$:

Assume B_τ is bounded such that $P(Y < B_\tau | x, BR) \leq u(B_\tau)$ (*)

$\Rightarrow$

If $Q_\tau = 0 \rightarrow B_{\tau+1} < B_\tau$ and (19.b) applies:

Lower Bound :

$$\begin{aligned} P(Y < B_{\tau+1} | x, BR) &= P(Y < B_{\tau+1} | x, BR, Q_\tau = 0)P(Q_\tau = 0 | x, BR) \\ &\quad + P(Y < B_{\tau+1} | x, BR, Q_\tau = 1)P(Q_\tau = 1 | x, BR) \\ &\geq P(Q_{\tau+1} = 0 | x, BR, Q_\tau = 0)P(Q_\tau = 0 | x, BR) \end{aligned}$$

Upper Bound

$\Rightarrow$ (19.b) does not provide information to bound $P(Y < B_{\tau+1} | x, BR)$ from above (i.e., upper bound). But (*) implies

$$P(Y < B_\tau | x, BR) \leq u(B_\tau)$$

$$\text{then, given that } B_{\tau+1} < B_\tau \Rightarrow P(Y < B_{\tau+1} | x, BR) \leq u(B_\tau)$$

Lower & Upper bound at the bid $B_{\tau+1}$ ($< B_\tau$)

$$P(Q_{\tau+1} = 0 | x, BR, Q_\tau = 0)P(Q_\tau = 0 | x, BR) \leq P(Y < B_{\tau+1} | x, BR) \leq u(B_\tau) \quad (21)$$

Similar arguments that justify $P(Y < B_{\tau+1} | x, BR, Q_\tau = 0) = 1$ in (20) justify the assumption $P(Y < B_{\tau+1} | x, BR, Q_\tau = 1) = 0$ in (21), since overestimation of $P(Y \geq B_{\tau+1} | x, BR, Q_\tau = 0)$ leads to underestimation of $P(Y < B_{\tau+1} | x, BR, Q_\tau = 1)$, thus, it is correct to let $P(Y < B_{\tau+1} | x, BR, Q_\tau = 1) = 0$ to bound $P(Y < B_{\tau+1} | x, BR)$ from below. The lower bound $l(B_\tau)$ in (20) is defined by (19.a) if $B_\tau = B_1$; otherwise, $l(B_\tau)$ equals the lower bound in (22) but substituting the subscript $(\tau + 1)$ for (τ) . The upper bound $u(B_\tau)$ in (21) is also defined by (19.a) if $B_\tau = B_1$; otherwise, $u(B_\tau)$ equals the upper bound in (20) also by substituting $(\tau + 1)$ for (τ) . To derive bounds on the cumulative distribution function $F(y | x, BR)$ with respect to intervals such as $y \in [B_{\tau-1}, B_\tau)$ where $B_{\tau-1} < B_\tau$ implies combining (20) and (21):

Let $y \in [B_{\tau-1}, B_\tau)$, $B_{\tau-1} < B_\tau$

For $y < B_{\tau-1}$: If $Q_{-(\tau-1)} = 0 \Rightarrow B_{\tau-1} > B_{-(\tau-1)}$
 $\Rightarrow$ (20) bounds $F(y|x, BR)$ for $y < B_{\tau-1}$
 If $Q_{-(\tau-1)} = 1 \Rightarrow B_{\tau-1} < B_{-(\tau-1)}$
 $\Rightarrow$ (21) bounds $F(y|x, BR)$ for $y < B_{\tau-1}$

Let either case be represented by:

$$y < B_{\tau-1}: \quad l(B_{\tau-1}) \leq F(y|x, BR) \leq u(B_{\tau-1}) \quad (22.a)$$

For $y < B_\tau$: If $Q_{-\tau} = 0 \Rightarrow B_\tau > B_{-\tau}$
 $\Rightarrow$ (20) bounds $F(y|x, BR)$ for $y < B_\tau$
 If $Q_{-\tau} = 1 \Rightarrow B_\tau < B_{-\tau}$
 $\Rightarrow$ (21) bounds $F(y|x, BR)$ for $y < B_\tau$

Let either case be represented by:

$$y < B_\tau: \quad l(B_\tau) \leq F(y|x, BR) \leq u(B_\tau) \quad (22.b)$$

$\Rightarrow$

$$\text{for } y \in [B_{\tau-1}, B_\tau), \quad L_{\tau-1, \tau} \leq F(y|x, BR) \leq U_{\tau-1, \tau}$$

where

$$L_{\tau-1, \tau} = \min\{l(B_{\tau-1}), l(B_\tau)\}, \quad U_{\tau-1, \tau} = \max\{u(B_{\tau-1}), u(B_\tau)\} \quad (22)$$

Compared to expression (7) the results in (22) show that allowing for anchoring effects implies a new interval to identify $F(y|x, BR)$ where the anchoring assumption is driven by a nonparametric interpretation of the parametric assumptions in Herriges and Shogren (1996). The interval in (22) implies a width equal to $U_{\tau-1, \tau} - L_{\tau-1, \tau}$. For $B_\tau > B_{\tau-1}$ the width is always positive¹¹

Bounds on $F(y|x)$ with full and partial respondents, and non-respondents

Sections 2.3 and 2.4 derive four bounding intervals $\{L_{\tau-1, \tau}, U_{\tau-1, \tau}\}$ on the conditional distribution function $F(y|x, BR)$ for any $y \in [B_{\tau-1}, B_\tau)$ category determined by CBR. These bounds are described in (7), (13), (18) and (22) with reference to A0, A1, A2 and A3, respectively, each of which implies a different assumption with respect to anchoring effects.¹² Let anyone of these four bounding intervals

¹¹ Showing this follows identical arguments as those employed to show a positive interval for $F(y|x, BR)$ in the case of Hurd et al. (1998).

¹² The models employed in Section 2 (i.e., Hurd et al., 1998, Jacowitz and Kahneman, 1995 and Herriges and Shogren, 1996) are not unique at explaining anchoring effects. Other alternatives could have been considered, for example, that of Cameron and Quiggin (1994) who interpret anchoring exclusively for a 2 bid model. Another alternative could have been to consider anchoring effects as described by the 'yea-saying' model in Green et al. (1998) which implies systematic asymmetric inequalities between reported and true fractions such that bounds should be derived with the condition $P(Q_\tau = 1|x, BR, Q_{-\tau}) \geq P(Y \geq B_\tau | x, BR, Q_{-\tau})$, or in its stronger form, $P(Q_\tau = 1|x, BR) \geq P(Y \geq B_\tau | x, BR)$.

together with $F(y|x, NR) \in [0, 1]$ and correctly assume that $F(y|x, FR)$ is fully identified by the sampling process. Bounds on (3) are defined as follows:

*Let an unfolding bracket design define regions $[B_{\tau-1}, B_\tau)$ in Y :
For any $y \in [B_{\tau-1}, B_\tau)$,*

$$\begin{aligned} & F(y|x, FR)P(FR|x) + (L_{\tau-1, \tau})P(BR|x) \\ & \leq F(y|x) \leq F(y|x, FR)P(FR|x) + (U_{\tau-1, \tau})P(BR|x) + P(NR|x) \end{aligned} \quad (23)$$

*where $L_{\tau-1, \tau}$ and $U_{\tau-1, \tau}$ are given by either (7), (13), (18) or (22)
and $L_{\tau-1, \tau} = 0$ if $B_{\tau-1} = 0$, $U_{\tau-1, \tau} = 1$ if $B_\tau = \infty$*

Bounds in (2) ignored partial response information thus resulting in a width between upper and lower bounds equal to $P(NR|x)$ at each y in the support of Y . Allowing for partial respondents and no anchoring effects leads to bounds as given in (23) with $\{L_{\tau-1, \tau}, U_{\tau-1, \tau}\}$ defined in (7); thus, in this case, the pointwise width between bounds is $\{P(Q_\tau = 0|\dots) - P(Q_{\tau-1} = 0|\dots)\}P(BR|x) + P(NR|x)$; the latter is smaller than $P(NR|x)$ if $\{P(Q_{\tau-1} = 0|\dots) - P(Q_\tau = 0|\dots)\}P(BR|x) > 0$ and this is always the case.¹³ Comparing the resulting widths in (23) with $\{L_{\tau-1, \tau}, U_{\tau-1, \tau}\}$ defined either by (13), (18) or (22) implies comparing different assumptions of anchoring. In the presence of a single bracket design we cannot test for anchoring or test which of the anchoring models best adjusts to the data. All models are valid and resulting widths can only be compared empirically. It can be said, however, that bounds in (23) based on (7) imply a stronger assumption than those based on (13), (18) and (22) where the assumption of no anchoring effects is relaxed. Thus, bounds in (23) based on (7) are (at most) as wide as those implied by bounds based on Assumptions A1 – A3. Likewise, bounds based on these latter anchoring assumptions are (at most) as wide as bounds implied by (2) where partial information is completely ignored. Section 4 makes use of the second wave of the Health and Retirement Study (HRS, 1996) to illustrate all bounds derived in Section 2 using the variable ‘earnings’.

As a final note, it is important to notice that although bounds in (23) identify pointwise bounds for each $y \in Y$, the specific contribution from partial information cannot identify $F(y|x, BR)$ for particular values $y = Y$, or for categories $[B_{\tau-1}, B_\tau)$ if one of the two bids ($B_{\tau-1}$ or B_τ) or both do not

¹³ $P(BR|x) > 0$ is assumed. So it remains to show that $P(Q_{\tau-1} = 0|x, BR) > P(Q_\tau = 0|x, BR)$. However, in the case of follow up bids $Q_{\tau-1} = 1 \Rightarrow Q_\tau$ so that within the subset of BR those that contribute to the event $P(Q_\tau = 0|x, BR)$ are a subset of those who contribute to the event $P(Q_{\tau-1} = 0|x, BR)$.

match the bids defined in the unfolding bracket design. This applies equally to all bounding intervals (7), (13), (18) and (22).

2.6 Allow for partial information with incomplete bracket respondents (IBR)

Initial non-respondents who decide to provide partial information may not necessarily complete the sequence of J conditional bids. Bounds (7), (13), (18) and (22) are derived assuming that bracket respondents are complete bracket respondents (CBR), that is, they answer with either ‘yes’ ($\kappa = 1$) or ‘no’ ($\kappa = 0$) to questions defined by (4) at each bid of the design and until this is completed. In practice, however, initial non-respondents that decide to initiate the unfolding bracket design (i.e., they answer such that $\kappa = 0,1$ at least at the first bid B_1) can always stop providing partial information (i.e., answer either ‘*don’t know*’ or ‘*refuse to answer*’) at some bid before the end of the unfolding design. These are incomplete bracket respondents (IBR) that may decide to stop providing information at the $(J - d)$ th bid where $d > 0$ is the number of remaining bids. Thus, the first difference is the number of categories resulting from each type of respondent: whereas CBR define as many as 2^J categories, IBR define combinations of $2^{(J-d)}$ categories for different values of $0 < d < J$. Empirically information from CBR and IBR cannot be mixed if the intend is to estimate $P(y \in [B_{\tau-1}, B_\tau] | \dots)$ because the categories $[B_{\tau-1}, B_\tau]$ between CBR and IBR do not match in range. On the other hand, it is possible to mix information from CBR and IBR to derive bounds on $F(y | x, BR)$ for $y \in [B_{\tau-1}, B_\tau]$ thus defining successive cumulative estimates on the conditional distribution at different B_τ bids. The result is a step function and the following applies:

$$F(y | x, BR) = F(y | x, BR, CBR) \times P(CBR | x, BR) + F(y | x, BR, IBR) \times P(IBR | x, BR) \quad (24)$$

The sampling process identifies $P(CBR | x, BR)$ and $P(IBR | x, BR)$. The conditional distribution functions $F(y | x, BR, CBR)$ and $F(y | x, BR, IBR)$ receive the same treatment as $F(y | x, BR)$ but for subgroups CBR and IBR . Thus, $F(y | x, BR, CBR)$ and $F(y | x, BR, IBR)$ are identified up to a bounding interval according to different anchoring assumptions defined by (7), (13), (18) and (22) but contributing differently according to how each subgroup CBR and IBR contributes to $y < B_\tau$. For example, let’s assume an unfolding bracket design with $J = 2$ bids; CBR complete the sequence and determine $2^2 = 4$ categories; $[\min(Y), B_{20}], [B_{20}, B_1], [B_1, B_{21}]$ and $[B_{21}, \max(Y)]$. With 2 bids it must be the case that IBR provide information only up to the first bid B_1 (either with $Q_1 = 0$ or $Q_1 = 1$) and answer ‘*don’t know*’ or ‘*refuse*’ at the second bid (i.e., at $B_{20} = B_2 | (Y < B_1)$ or $B_{21} = B_2 | (Y \geq B_1)$) to define $2^{J-2} = 2$ categories $[\min(Y), B_1]$ and $[B_1, \max(Y)]$. The result is that CBR contribute to

bound $F(y|x, BR, CBR)$ from below with lower limits at $\min(Y)$, $B20$, $B1$ and $B21$ defined in (7), (13), (18) and (22) whereas the upper limits are estimated at $B20$, $B1$, $B21$ and $\max(Y)$ also defined in (7), (13), (18) and (22). Instead, IBR contribute to bound $F(y|x, BR, IBR)$ with estimates of lower and upper limits at $\min(Y)$ & $B1$, and $B1$ & $\max(Y)$, respectively, with estimated using corresponding expressions also from (7), (13), (18) and (22).

2.7 Bounds on Quantiles of the distribution

Estimates on the conditional distribution of quantities (e.g., savings, income, earnings), are often described in terms of conditional quantiles. For $\alpha \in [0,1]$, the α -quantile associated with the conditional distribution of Y given $x \in X$ is defined as follows:

$$q(\alpha, x) = \inf_y \{y : F(y|x) \geq \alpha\} \quad (25)$$

For $\alpha > 1$, $q(\alpha, x) = \infty$, whereas for $\alpha < 0$, $q(\alpha, x) = -\infty$. Following Manski (1994), bounds on the quantiles can be derived by inverting the bounds on the conditional distribution function. Any of the bounds given by (7), (13), (18) and (22) can be written as follows:

$$lb(y, x) \leq F(y|x) \leq ub(y, x) \quad (26)$$

Both $lb(y, x)$ and $ub(y, x)$ are non-decreasing functions in y so that expression (26) can be inverted to give the following bounds on the conditional quantiles:

$$\inf \{y : ub(y, x) \geq \alpha\} \leq \inf \{y : F(y|x) \geq \alpha\} \leq \inf \{y : lb(y, x) \geq \alpha\} \quad (27)$$

The result in (27) relative to (26) is easily illustrated using a graph of the distribution function with y along the horizontal axis and $F(y|x)$ along the vertical axis: bounds in (26) on the distribution function squeeze $F(y|x)$ in between two curves and the vertical distance between these curves is the width between the bounds at each given value of $y \in Y$. Expression (27) results from reading the same graph horizontally so that for a given probability value $\alpha \in [0,1]$ the graph illustrates a lower and an upper bound on the α -quantile.

3 Estimation

All bounding intervals given by (23) in Section 2 are explained in terms of population characteristics. Each of these bounds can easily be estimated using the corresponding sample analogues for $F(y|x, R) = E[I(Y \leq y)|x, R]$, $P(R|x) = E[R|x]$ where R stands for any one of the conditional spaces in the sets (7),(13),(18) and (22). If the covariates in X contains a mixture of continuous and discrete variables bounds could be estimated using nonparametric weighting techniques, for example, the use of product Kernel estimation and an appropriate selection of bandwidths (for example, see Härdle, 1990, as reference for nonparametric regression). Our empirical example considers discrete covariates ($X = \text{gender}$) so that kernel weights are not required. Thus, the illustration will be based on estimating quantiles of the distribution, first by estimating distribution functions of a variable ‘earnings’ that suffers from non-negligent nonresponse with all the alternatives in (23), second by inverting these distribution to obtain quantiles as in (27). The use of cross-sectional weights is a requirement for the sample to become representative of the underlying population (see Section 4.1 below).

Estimates of the width between the upper and the lower bound describe uncertainty due to item nonresponse thus reflecting a population concept rather than a sampling concept. This means that for each of these bounding intervals we still need to estimate confidence bands to reflect finite sampling error. For any of the four bounding intervals in Section 2, a pointwise distance in the support of Y between the upper confidence band for the upper bound and lower confidence band for the lower bound reflects joint uncertainty due to both sampling error and error due to item nonresponse. This paper makes use of a naïve bootstrap procedure that consists on re-sampling 1000 times (with replacement) from the original data to estimate pairs of upper and lower 95% confidence bands for the upper bound and for the lower bound for each set of bounds. In all cases, the 95% confidence bands displayed in the empirical section are the lower confidence band for the lower bound and the upper confidence band for the upper bound.

4 Illustration

4.1 The Data

We use the 1996 wave of the Health and Retirement Study (HRS) to illustrate the empirical implications of bounds in Section 2. The HRS is a longitudinal study that has been conducted since 1992 by the University of Michigan on behalf of the US National Institute of Aging. The panel collects information on aspects of health, wealth, retirement and other socio-economic conditions for a representative cohort of US citizens born between 1931 and 1941. The 1996 wave applies an unfolding bracket design to elicit further earning information from initial non-respondents, while the number of

individuals who may still be earners (age 65 or below) is optimised in 1996 since the HRS weights only individuals born between 1931 and 1941. These reasons drive our selection of the HRS, 1996 wave.

In 1992 the panel surveyed 7,600 households. The 1996 has data from 6,736 household resulting in 8,436 eligible individuals.¹⁴ Each household has a representative member who answers all questions relating to own earnings and earnings from other surveyed household members (e.g., the spouse). The first sample selection criterion consists on selecting only those who declare to be household representatives (6,075) and who declare to have worked for wages and salary (i.e., no self-employed) during the last calendar year (i.e., 1995): this implies an initial sample selection of 2,862 individuals from the 6,075 representatives.¹⁵ We aim at bounding (27) with different bounds in (23), where $y \in Y$ stands for annual earnings and $x \in \{0,1\}$ indicates gender. This paper does not deal with misreporting or measurement error, thus, we assume that both earnings (when declared in full) and $X = \text{gender}$ are free from self-reporting error. All estimates that follow are based on weighted values using the corresponding cross section personal weights provided by the HRS (see HRS, tracker file and documentation).

The variable ‘earnings’ is not fully observed for all individuals in the selected sample. Initially all 2,862 respondents who declared a status of employee during 1995 face the following open-ended question:

*'about how much wage and /or salary income did you receive during the last
calendar year?'*

'.....any amount' (in USA dollars)

'.....don't know'

'.....refuse'

(28)

Out of the 2,862 individuals, 2,508 become full respondents (*FR*) because they answer with a specific amount to question (28), and the remaining 354 are the *initial* non-respondents (*NR*) who are then routed to an unfolding bracket design defined in Table 1;

¹⁴ The sample is representative using cross sectional weights, and those who are not age or sample eligible are given a weight of zero. In 1996 only 8,436 individuals (of 27,109 thus far interviewed since 1992) have a positive weight assigned to them.

¹⁵ We are dealing with nonresponse. Not responding to ‘own earnings’ may be due to reasons completely different to not responding to ‘other person’s earnings’. For example, reporting earnings from a partner can be affected by the fact that a partner is (or not) present in the household at the time of the interview, itself possibly as consequence of working, or even health reasons.

Table 1: Distribution of CBR and IBR with respect to an unfolding bracket design for the variable 'annual earnings', Health and Retirement Study, (HRS, 1996). Amounts represent nominal USA Dollars.

Group	1 st BID; B1	Answer to bid	2 nd BID: B20 or B21	Answer to BID	3 rd BID: only for Q21:B311	Anchor 2: B20 or B21	Resulting Bracketed categories
CBR	>\$25,000?	YES Q1=1	>\$50,000?	YES Q21=1	>\$100,000	YES (end) Q311=1	[100,000; INF)
				NO (end) Q21=0		NO (end) Q311=0	[50,000; 100,000)
				YES (end) Q20=1			[25,000; 50,000)
		NO Q1=0	>\$5,000?	NO (end) Q20=0			[5,000; 25,000)
							[0; 5,000)
IBR	>\$25,000?	YES Q1=1	>\$50,000?	YES Q21=1	>\$100,000	Don't know, or refuse to respond	[50,000; INF)
				Don't know, or refuse to respond			[25,000; INF)
		NO Q1=0	>\$5,000?	Don't know or refuse to respond			[0, 25,000)

Note: This table is the complete version of the unfolding bracket design. There is no natural upper bound to earnings other than to allow for it to be infinity (INF). However, restricting the example to cover employees, the natural lower bound for the lowest located category is earnings=0. Including self-employees would not had made possible this assumption since self-employed could in principle receive negative earnings in which case the lowest category would had been (-INF; 5,000) or (-INF, 25,000) in the case of IBR.

With 354 initial non-respondents facing the unfolding bracket design, the initial (weighted) nonresponse probability is 0.12. All 354 initial non-respondents face question (4) with the same bid of \$25,000 to start up the sequence. At this point 86 are classified as full non-respondents (thus, *NR* is now an event associated with these individuals) since they declare 'don't know' or 'Refuse to respond' to the initial bid and, consequently, are not driven through the design by the interviewer. The rest are such that 255 complete the sequence of brackets (CBR) and 13 leave the sequence (IBR) before this is completed (i.e., they report either 'don't know' or 'Refuse to respond' at some bid before the design ends). Thus, the new (weighted) distribution between sub-groups becomes such that $P(FR) = 0.88$ (2,508 units), $P(NR) = 0.031$ (86 units), $P(CBR) = 0.08$ (255 units) and $P(IBR) = 0.005$ (13 units). Table 1 shows the resulting categories from complete and incomplete bracket responses illustrating the dynamics implied by the unfolding bracket design, with the initial bid of $B1 = \$25,000$ followed by either one or two conditional bids depending on the route followed by the individual along the sequence. For example,

complete bracket respondents that answer $Q1=1$ to (4) when faced with $B1=\$25,000$ are routed to a second bid $B21=\$50,000$; where $B21$ indicates the 2nd bid conditional on having answered ‘yes’ to $B1$. If $Q21=1$ at $B21$ individuals are driven to a 3rd bid $B311=\$100,000$ and the sequence stops. However, the sequence stops at $B21$ if $Q21=0$. All other paths in Table 1 follow similar explanations. Table 2 complements Table 1 with the sample distribution of CBR, IBR and NR among the categories for the full initial non-response sample and by gender sub-groups.

Table 2: Distribution of initial Non-respondents between full non-respondents (NR), complete Bracket respondents (CBR) and incomplete bracket respondents (IBR) by categories on Annual Earnings (HRS, 1996)

Category	Full Sample n= 354	Males n= 94	Females n= 260
<u>Complete Bracket Respondents</u> (CBR)			
[100,000, INF)	n=4, 0.016 (0.007)	n=1, 0.014 (0.013)	n=3, 0.016 (0.008)
[50,000; 100,000)	n=19, 0.056 (0.013)	n=5, 0.043 (0.022)	n=14, 0.062 (0.015)
[25,000; 50,000)	n=69, 0.200 (0.022)	n=23, 0.240 (0.045)	n=46, 0.183 (0.025)
[5,000; 25,000)	n=134, 0.350 (0.026)	n=29, 0.290 (0.049)	n=105, 0.374 (0.031)
[0; 5,000)	n=29, 0.072 (0.014)	n=4, 0.032 (0.019)	n=25, 0.089 (0.018)
<u>Incomplete Bracket Respondents</u> (IBR)			
[50,000, INF)	n=2, 0.006 (0.004)	n=1, 0.014 (0.012)	n=1, 0.002 (0.003)
[25,000, INF)	n=6, 0.015 (0.006)	N=0	n=6, 0.022 (0.009)
[0, 25,000)	n=5, 0.020 (0.007)	n=3, 0.047 (0.022)	n=2, 0.01 (0.006)
Full Non-respondents (NR)	n=86, %= 0.266 (0.023)	n=28, %= 0.322 (0.044)	n=58, %= 0.242 (0.027)

Note: See notes in Table 2. All percentages are with respect to the 354 members (94 males and 260 females) in the sample of partial respondents including IBR. In each case the samples are weighted with personal level cross-sectional weights (HRS 1996, tracker file). Standard errors in brackets; ‘n’ indicates number of observed units per category.

In the full sample of 2,862 units is distributed between males and females such that 1,036 (62%) are males and 1,836 (38%) are females, where percentages are based on weighted values. In terms of partial respondents, units of females outnumber units of males for both CBR and IBR, but weighted probabilities show that relative to the sample of 354 there is an even representation of genders by categories. Full nonresponse probability is significantly higher for males than for females.

Specifically for Assumption A.2 (Jacowitz and Kahneman, 1995) it is important to examine the probability of being at particular ‘node’ in the path of the tree defined by Table 1, that is, we need to examine the sample properties for $P(Q1=1|PR)$, $P(Q21=1|Q1=1,PR)$, $P(Q311=1|Q21=1,PR)$ and $P(Q20=1|Q1=0,PR)$ so that Assumption A.2 in (15/16) becomes operational for each of the bids $\{\$5,000, \$25,000, \$50,000, \$100,000\}$, a necessary condition to develop and estimate bounds in (20). We show results for CBR (similar sample properties apply to IBR):

Table 3: Sample Probabilities for different nodes in Table 1 using CBR to Annual Earnings, (HRS 1996)

Probability at nodes defined by Table 3	Full CBR Sample n= 255	Males n= 62	Females n= 193	Conclusions with respect to HIGH or LOW ANCHOR
$P(Q1 = 1 PR)$	n=92/255 %= 0.283 (0.028)	n=29/62 %= 0.317 (0.059)	n=63/193 %= 0.217 (0.030)	Q1=1 → B1 (\$25,000) HIGH ANCHOR
$P(Q21 = 1 Q1 = 1, PR)$	n=19/92 %= 0.265 (0.048)	n=5/29 %= 0.192 (0.073)	n=14/63 %= 0.299 (0.058)	Q21=1 → B21 (\$50,000) HIGH ANCHOR
$P(Q311 = 1 Q21 = 1, PR)$	n=4/23 %= 0.217 (0.086)	n=1/6 %= 0.252 (0.178)	n=3/17 %= 0.206 (0.098)	Q311=1 → B311 (\$100,000) HIGH ANCHOR
$P(Q20 = 1 Q1 = 0, PR)$	n=134/163 %= 829 (0.029)	n=29/33 %= 0.906 (0.051)	n=105/130 %= 0.807 (0.035)	Q20=1 → B20 (\$5,000) LOW ANCHOR

Note: See notes in Table 1 and 2. All percentages are with respect to the corresponding sub-groups. The indicator 'n' shows the total units at each Q_i given the number of units in the corresponding space. The probabilities are based on weighted averages. Bracketed numbers show standard errors.

Thus, from Table 3 and recalling the Jacowitz and Kahneman (1995) implications, we conclude that the bids B1=\$25,000, B21=\$50,000 and B311=\$100,000 act as 'high anchors' and the bid B20=\$5,000 acts as a 'low anchor'. The information in Table 3 drives the characterization of (18) in Appendix 2 (see Table A2.3). Finally, Table 4 examines the variable 'earnings' with two distinct sets of information: weighted earnings using full respondents only (exogenous nonresponse assumption) and weighted earnings when both non-respondents and partial respondents have had their missing values imputed (constructed at sources, HRS, 1996 imputed variables files). The imputation method makes use of partial information for those in the *BR* group taking this information as correct, i.e., imputed earnings and bounds in (7) hold identical assumptions with respect to anchoring; in both cases the assumption is that of no anchoring. This is not the case for bounds based on (13), (18) or (22) where different models of anchoring relax the no anchoring assumption in (7).

Table 4: Sample statistics of gross annual earnings (standard errors in brackets)

	Full Sample	Males	Females
Earnings, FR only (n=2,508: male = 942, female = 1,566)			
Average (standard error)	\$31,342 (566)	\$36,202 (942)	\$28,148 (695)
Range	min=\$0 max=\$350,000	min=\$0 max=\$300,000	min=\$0 max=\$350,000
Imputed earnings (n=2,862: male = 1,036, female = 1,836)			
Average (standard error)	\$30,867 (525)	\$35,400 (876)	\$28,040 (645)
Range	min=\$0 max=\$350,000	min=\$0 max=\$300,000	min=\$0 max=\$350,000

Note 1: Estimates are based on the weighted sample. The variable 'earnings' refers to 'gross earnings' and is based on an initial open-ended question where individuals who declare to have worked over the last calendar year (i.e., 1995) are asked to declare the total gross amount of earnings received from such work. Amounts refer to nominal US dollars. The original variable draws from the publicly accessible HRS- 996 file named HR96J_H. The derived variable with the imputed amounts draws from the file named H96i_jh.

Table 4 suggests that males earn significantly more than females. Imputation of missing earnings implies a slight drop in mean values for the full sample and for the sub-sample of males, but has weaker effects on mean earnings for females. The standard errors do not differ much before and after imputation, thus reflecting small noise induced from imputing non-respondents and partial respondents. This feature is expected when imputation is performed in the support of the full response sample.

4.2 Specification

Section 2 provides the general expression for bounds that allow for partial information when initial respondents are routed to an unfolding design with J possible conditional bids resulting in 2^J categories for CBR and 2^{J-d} categories for IBR (see Section 2.3). Table 1 shows that for the specific case under consideration J is at most 3; in the case of CBR the number of resulting categories are $1 + 2^2 = 5$ where the ‘1’ accounts for the special case of an asymmetric design at $B3$. On the other hand, IBR define 3 distinct categories. The partition $Y = \{[0, B20), [B20, B1), [B1, B21), [B21, B311), [B311, \infty)\}$ defines the support for CBR, whereas IBR define an alternative partition with overlapping categories: $Y = \{[0, B21), [B1, \infty), [B21, \infty)\}$. The two sub-groups of partial respondents have bids in common and this helps to join the sample information from the two groups to estimate identification regions on $F(y | \text{gender})$ for different $y < B_r$. Appendix 2 characterizes (7), (13), (18) and (22) for the particular case when the unfolding design follows the example as in Table 1.

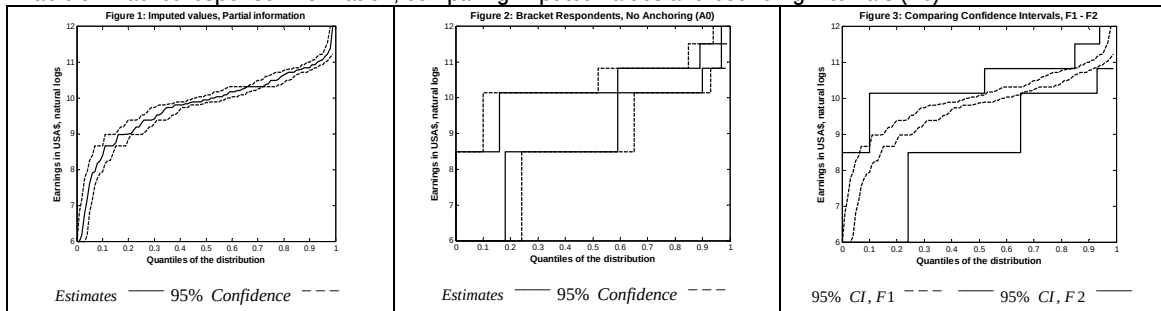
4.3 Bounds, Bracket Respondents and anchoring

First we use bracket respondents to validate the properties of imputed earning. The HRS imputation procedure assumes no anchoring effects. This is in fact the same assumption characterizing bounds in (7). Comparing estimates based on imputed values to estimate of bounds in (7) is a test of the assumptions (other than no anchoring) underlying the imputation process. Figure 1 in Table 5 plots the distribution using the 268 bracket respondent’s imputed earnings. Figure 2 plots the interval identified using the original information provided by the 268 bracket respondents.¹⁶ Figure 3 compares Figure 1 to Figure 2 with the corresponding 95% confidence intervals (dashed lines in each original figure). Imputing values results in a continuous earning’s variable and therefore in a continuous distribution function (Figure 1); in this case the median is located between \$19,200 -- \$23,900 with 95% confidence. Using the original unfolding bracket’s information respondents identify an earning’s category so that information is discrete in nature and the results are bounds shaped as step functions (Figure 2): in this case the median is

¹⁶ Bounds in Figure 2 are based on estimates from (27) thus inverting the estimate of the distribution function from (7). The upper bound in (7) becomes the lower bound for quantiles in (27) and the lower bound in (7) becomes the upper bound for quantiles in (27).

located between \$4,800 – \$25,000; this region of identification is also based on a 95% confidence interval, but it includes uncertainty from both sampling error and partial (as opposed to exact) information. Thus, in the case of the median, the result is a width that is 77% wider than that in Figure 1. The estimated upper and lower bounds in Figure 2 meet exactly at \$25,000 and \$50,000; the gap at \$5,000 and \$100,000 reflects the effect of asymmetric overlapping categories from IBR. In the absence of this subgroup of partial respondents upper and lower bounds would meet exactly at all bids in the design. Finally, and ‘holding the no anchoring assumption’, Figure 3 demonstrates that the imputed variable cannot be rejected because the 95% confidence region identified using imputation is nested by that identified with the step function interval (also with 95% confidence).

Table 5: Bracket response information, comparing imputed values and bounding intervals (A0)



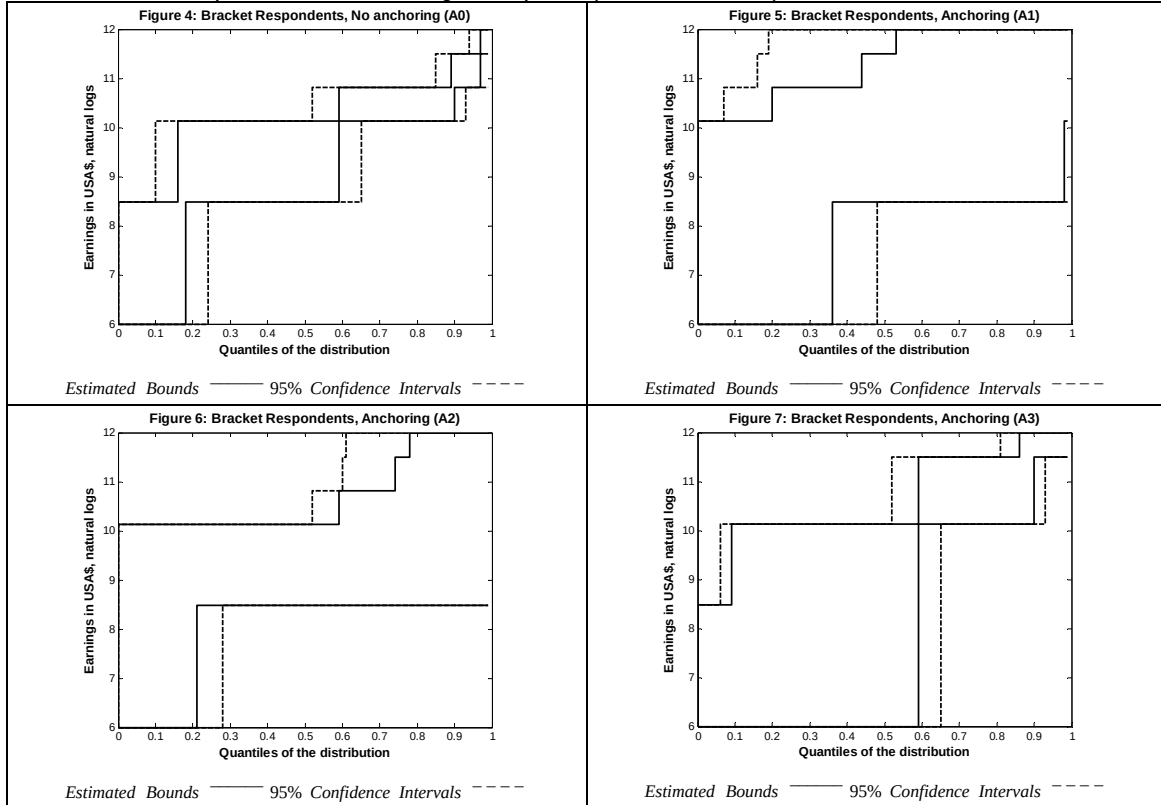
Note: The horizontal range has been artificially trimmed between \$500 and \$160,000 for a clearer illustration. Below and above these values the plots do not provide further information.

Figures 1 to 3 ignore possible anchoring effects. Table 6 compares the effect of three different anchoring assumptions (A1 to A3) against the no anchoring Assumption A0.¹⁷ Compared to Figure 4 (A0), all other figures clearly show that relaxing the assumption of no anchoring leads to a loss in the power of the identification region. The loss differs between anchoring assumptions and according to quantiles. With 95% confidence, Assumption A1 identifies the median between \$4,800 and the maximum value in the range of earnings (\$350,000). Also with 95% confidence Assumption A2 identifies the same quantile between \$4,800 and \$25,000 whereas Assumption A3 identifies the median between \$0 and \$25,000. Figure 5 show anchoring effects based on Hurd et al. (1998) where the median assumption widens the identification region for all the quantiles in the distribution. Assumption A2 implies the findings in Jacowitz and Kahneman where ‘high’ and ‘low’ perceived anchors affect the tails of the distribution: accounting for these effects leads to bounding intervals with thicker tails as reflected in Figure 6. Assumption 3 defines no anchoring at the first bid and this is reflected with a point estimate at \$25,000 that is identical to that in Figure 4 (A0): since anchoring based on A3 implies answers that are

¹⁷ Figure 2 is identical to Figure 4. It is included in Table 6 for better visual comparison.

pulled towards $B1$ (i.e., \$25,000) from either side, the regions around \$25,000 from below and above became wider when compared to the identification regions in Figure 4 (A0).

Table 6: Bracket respondents and anchoring assumptions (A0, A1, A2 and A3)



4.4 Distribution of earnings by gender

In this exercise we illustrate the use of bounding intervals as a tool to detect difference (in earnings) between distinct subgroups. Studies on labour market outcome (e.g., ‘earnings’) often disaggregate the sample between males and females thus assuming systematic differences between genders. It is therefore interesting to compare cross-sectional difference on earnings between genders when the sample represents individual’s close to completing their labour market cycle (i.e., nearing retirement age) and earnings reflect near human capital development.¹⁸ Our analysis is only an illustration

¹⁸ We are comparing ‘annual earnings’ between genders where the variable ‘earnings’ is assumed to measure similar intensity at work for all the members in the sample. This would require that the average number of months worked per calendar year is similar among those with similar earnings. We looked at months worked over 1995 and found that on average individuals work 10.8 months per year. The extreme values of months worked were between 2 and 12 but we also saw similarity number of months worked by earnings categories (e.g., those working below 4 months are associated with the lowest category of earnings). It is not possible for us to compare individuals by ‘monthly averages’ because this would distort the lower and upper limits for the categories and information from partial respondents would need to be disregarded.

of Section 2 with no policy implication since we are not conditioning on variables that might imply policy analysis.

Our concern is to perform the comparison between genders taking into account that 12% in the representative sample do not declare annual earnings, although they have declared to have worked for pay during the corresponding calendar year. Ignoring these 12% may bias our inference if nonresponse is not random with respect to earnings. Allowing for imputed variables implies the use of the assumptions underlying the imputation procedure (e.g., ignoring possible anchoring effects from bracket respondents) that may also bias our inference: Figure 8 shows the result of comparing 95% confidence bands on the quantiles of earnings between genders and using the complete sample with missing values imputed.¹⁹ Alternatively, we estimate identification regions on the quantiles of earnings that allow for any type of non-random nonresponse at the expense of this nonresponse contributing towards total uncertainty. Figures 9 to 13 do this with various assumptions on how to deal with bracket response information; in these figures the vertical distance between bounds is the sum of nonresponse error and sampling error also with 95% confidence. All figures in Table 7 help to test for earnings difference between the two genders by comparing the location of identification regions between the two subgroups. Table 8 complements Figures 8 – 13 with a formal test on a selection of quantiles. All figures in Table 7 are based on estimating all versions of (23) and inverting these estimates to obtained (27).

Figure 8 shows that once the missing values are imputed the identification region for male is always above that identified for females, with 95% confidence. Thus, imputation would imply rejecting the null of earnings equality between genders in favour of males receiving significantly higher earnings than females at any level of earning. Figure 9 relaxes all assumptions implied by imputation to allow for any type of nonresponse error thus increasing total uncertainty when estimating the identification regions. At the same time, Figure 9 is based on expression (2) so that information from partial respondents is ignored. Thus, Figure 9 implies a loss in the power of the test when compared to Figure 8. This is clear since overlapping regions in Figure 9 imply that we cannot reject the null of earning equality between genders at any level of earnings. Figure 10 improves the power of the test by including information provided by bracket respondents but ignoring possible anchoring effects (A0). In Figure 10 overlapping of the 95% confidence intervals is only visible a α -quantiles below 0.3 and above 0.85, but the null of earning's equality between genders is rejected for middle earning's quantiles. Figures 11 to 13 apply (13), (18) and (22) to relax the assumption of no anchoring implied by Figure 10. The results vary between the

¹⁹ Pointwise estimate with imputation, or estimates of bounds are always nested by their 95% confidence intervals. Figures in Table 7 reflect these 95% confidence intervals for ease of illustration. For example, Figure 9 shows the median for males bounded between \$26,200 and \$36,900 with 95% confidence, but the actual pointwise estimated upper and lower bounds locate the median between \$29,900 and \$34,600; this latter interval reflects *only* uncertainty as result of item nonresponse. We choose to illustrate upper and lower bounds using 95% confidence intervals because the use of 4 (rather than 8 curves) in one plot clarifies the comparison between genders. Table 8 complements these figures showing point estimates for a selection of quantiles.

three figures but all have in common a loss in the power of the test as total uncertainty in estimation increases. Table 8 complements Figures 8 – 13 with a formal test that covers a selection of quantiles.

Table 7: Bounding intervals with and without bracket respondents and no anchoring effects

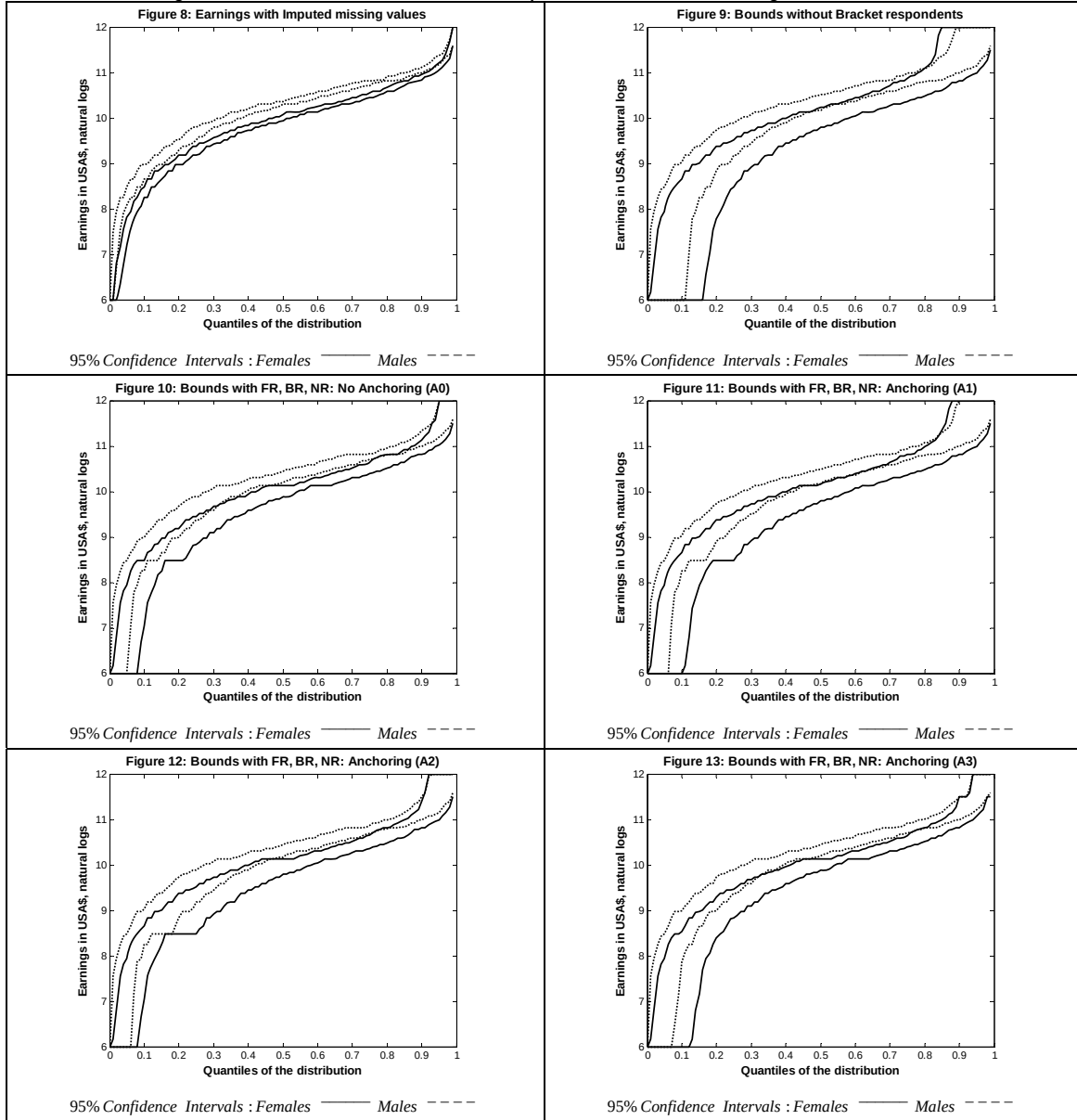


Table 8: Testing for Quantile difference between genders. Amounts refer to annual earnings in USA Dollars (HRS, 1996)

	Quantile Estimates Male Upper bound (standard error)	Quantile Estimate Female Lower bound (standard error)	Quantile Difference: Male (Lower bond) – Female (Upper bound)	T-statistic (bootstrap critical values)
IMPUTATION (FIG. 8) ¹				
25 th Percentile	14,700 (1,285)	10,800 (800)	3,900	2.6 (2.8)
40 th Percentile	25,000 (786)	17,900 (522)	7,100	7.4 (2.2)
50 th Percentile	29,900 (764)	21,900 (765)	8,000	7.5 (1.8)
60 th Percentile	35,600 (1,219)	25,900 (879)	9,700	6.5 (1.9)
75 th Percentile	47,500 (1,146)	35,600 (1,027)	11,900	7.7 (2.1)
90 th Percentile	61,000 (2,199)	53,900 (1,307)	7,100	2.8 (2.7)
BOUNDS, NO BRACKETS (FIG. 9)				
25 th Percentile	11,900 (1,035)	12,750 (503)	-850	-0.8 (-2.7)
40 th Percentile	22,800 (1,322)	20,500 (946)	2,300	1.4 (2.0)
50 th Percentile	29,900 (1,104)	25,900 (815)	4,000	2.9 (0.7)
60 th Percentile	34,600 (1,250)	32,000 (1,177)	2,600	1.5 (1.8)
75 th Percentile	48,000 (1,528)	49,700 (1,386)	-1,760	-0.9 (-2.3)
90 th Percentile	65,500 (2,350)	350,000 (0,00)	-285,500	-121.5 (-2.3)
BOUNDS, and BRACKETS with A0 (FIG. 10)				
25 th Percentile	13,700 (1,180)	11,900 (537)	1,800	1.4 (1.6)
40 th Percentile	25,000 (751)	19,500 (581)	5,500	5.8 (1.1)
50 th Percentile	29,900 (985)	25,000 (178)	5,000	4.9 (1.1)
60 th Percentile	35,600 (1,235)	27,900 (946)	7,800	5.0 (1.5)
75 th Percentile	48,000 (1,414)	39,400 (1,080)	8,600	4.8 (1.5)
90 th Percentile	64,500 (2,337)	62,000 (2,784)	2,500	0.7 (1.5)
BOUNDS, and BRACKETS with A1 (FIG. 11)				
25 th Percentile	11,900 (1,032)	12,800 (503)	-850	-0.7 (-2.1)
40 th Percentile	22,800 (1,292)	20,500 (946)	2,300	1.4 (2.0)
50 th Percentile	29,900 (1,104)	25,000 (255)	5,000	4.3 (0.0)
60 th Percentile	34,600 (1,250)	29,900 (872)	4,700	3.1 (2.0)
75 th Percentile	47,900 (1,528)	44,500 (2,535)	3,500	1.2 (1.3)
90 th Percentile	64,500 (2,350)	98,000 (89,461)	-33,500	-0.4 (-2.9)
BOUNDS, and BRACKETS with A2 (FIG. 12)				
25 th Percentile	11,900 (1,015)	12,800 (503)	-850	-0.8 (-2.7)
40 th Percentile	22,800 (1,322)	20,500 (946)	2,300	1.4 (2.0)
50 th Percentile	29,900 (1,104)	25,000 (240)	5,000	4.5 (2.0)
60 th Percentile	34,600 (1,250)	27,900 (947)	6,700	4.3 (1.9)
75 th Percentile	48,000 (1,530)	39,400 (1,107)	8,600	4.5 (0.9)
90 th Percentile	64,500 (2,350)	69,000 (6,985)	-4,500	-0.6 (-3.2)
BOUNDS, and BRACKETS with A3 (FIG. 13)				
25 th Percentile	13,700 (1,180)	12,800 (482)	950	0.7 (2.3)
40 th Percentile	25,000 (751)	19,500 (680)	5,500	5.4 (0.8)
50 th Percentile	29,900 (894)	25,000 (83)	5,000	5.0 (0.1)
60 th Percentile	35,600 (1,235)	27,900 (946)	7,700	5.0 (1.5)
75 th Percentile	48,000 (1,414)	39,400 (1,100)	8,600	4.8 (1.5)
90 th Percentile	64,500 (2,317)	74,200 (6,800)	-9,700	-1.4 (-3.3)

1 In the case of imputed values the point estimates are exactly identified. Thus, the test is based on the quantiles difference (males - females) normalized by the corresponding standard error. For all other rows in Table 8 the test is based on (male's lower bound – female's upper bound) also normalized by the corresponding standard error.

2 The 1000 bootstrap estimates of the difference are used to find the empirical distribution for the t-statistic centered on the original estimate of the difference. This provides the bootstrap critical values based on the upper 2.5 percentile of the empirical t-distribution using an equal tail test. Therefore, the null of equality is rejected at a 0.05 level when the estimate of the t-statistic is greater or equal than the bootstrap critical value (in brackets).

All tests in Table 8 are based on an equal tail test for the null of earnings equality between genders. We test for pointwise zero difference between the estimated quantiles (Table 8, row 2), and pointwise zero difference between the estimated lower bound and estimated upper bound on the quantiles of the distribution for males and females, respectively (Table 8, rows 3 to 7); in the latter case the test is justified because the corresponding Figures 9 to 13 show that if an overlap exists, this is due the upper 95% confidence band for females being above the lower 95% confidence band from males. Whereas Figures 8 to 13 are based on plotting 95% confidence bands for easiness of illustration, Table 8 illustrates point estimates and corresponding standard errors. Bootstrap critical values based on the centered t-statistic are used to draw inference from the estimate t-statistic for each of the quantiles. Estimates at the 90th percentiles (especially for the subgroup of females) are sometimes unreliable due to low density at the upper tail of the distribution.²⁰

As illustrated with Figure 8, Table 8 shows that once the missing earnings values are imputed, inference imply that males earn significantly more than females at all levels of earnings. Dropping the imputation procedure and allowing for any type of nonresponse while ignoring partial information (Table 8, row 3) shows that the null of equality is only rejected in the neighbourhood of the median. Allowing for bracket respondents and no anchoring effect (Table 8, row 4, assumption A0) differs from conclusions based on imputed values in that the estimates with A0 cannot detect significant difference at the tails of the distribution, thus, the formal test in Table 8 suggest that Figure 10 rejects the null of equality in favour of higher earnings for males at α -quantiles in the (0.4; 0.75) interval. Once the assumption of no anchoring is relaxed (Table 8, rows 5 to 7), the power of the test is reduced and the earning's equality rejection region shrinks for estimates based on Assumptions A1 and A2: in both cases the test suggest that males earn significantly more than females but only for α -quantiles in the (0.5; 0.75) and not outside this range, a result that contrast with inference based on the imputed variable. Assumption A3 provides similar inference to those based on Assumption A0. Recall that A3 implied only 'partial anchoring effects', i.e., there is no anchoring in the first bid: since in this unfolding bracket design most bracket respondents face 2 bids (only a few face B3), it is not surprising that results using A3 are similar to those using A0. Finally, bounding intervals using the subgroup of females leads to wider intervals at the extreme tails of the distribution despite an overall lower sampling variance for this subgroup (see the standard errors in Table 8 for females who constitute a larger sample size). The reason is not a higher nonresponse error for females relative to males (see Table 2) but a very low density at the extreme of the distribution for the subgroup of females. This, in turn, explains the negative estimates at the extremes in Table 8, rows 2 to 7.

²⁰ We do not consider high values (e.g., \$300,000, \$350,000) as outliers because 'outliers' are given a special code in the HRS, and this was not the case for such values.

5 Conclusions

Manski's approach to deal with item nonresponse avoids the assumptions usually associated with parametric and semi-parametric methods. On the other hand, it identifies the unknown parameters up to an upper and lower bound only. In this paper, these bounds are extended to take account of the information contained in follow up categorical questions to initial non-respondents. Such questions are included in many current household surveys (e.g., the Health and Retirement Study and related panels, or the SHARE, the counterpart of the HRS in Europe). Several studies have shown that responses to such questions can be subject to response errors due to anchoring effects. Some existing studies model this response error with a parametric set up. We have extended the Manski's bounds to allow for anchoring in a nonparametric way, starting from various nonparametric anchoring assumptions inspired by the existing parametric models. Thus, our bounds simultaneously allow for any type of selective nonresponse and various forms of anchoring.

Using the variables earnings (wages and salary) of the household representative taken from the 1996 wave of the Household and Retirement Study, the empirical section shows estimates of Manski's basic worst case bounds that do not use the bracket respondents information and compares these with estimates of the new bounds. For the variable considered, the initial (weighted) nonresponse rate is 12%. Most of the initial non-respondents answer unfolding bracket questions, and the percentage of (weighted) full nonresponse rate drops to 3%. Incorporating information provided by bracket respondents tightens the bounds. Allowing for anchoring effects reduces the gains in information but still leads to estimated bounds that are substantially more informative than the bounds that do not use bracket information. This is illustrated by using the bounds to test for equality of quantiles between genders. Adding the information provided by bracket respondents improves the power of the tests, and leads to rejecting the null more often. How much the power of the tests increases depends on whether and how anchoring is allowed for.

Manski's bounds are an elegant, intuitively plausible and extremely flexible way to allow for selective nonresponse. Their flexibility is at the same time their main weakness: the bounds are often so wide that they do not provide enough information for the economic issue under consideration. This paper shows that additional information on bracket responses by initial non-respondents can be useful to make the bounds more informative. This is still true if anchoring is allowed for, though to a lesser extent. Furthermore, we have used the bounds to validate the imputation procedure used to impute missing earnings values for the HRS publicly released data. Missing values are imputed using information from bracket respondents and assuming no anchoring effects. We show that using imputed values to make inference on earnings equality between genders leads to different conclusions than those implied by the bounding interval: with this latter earning's differentials between genders is often rejected at the upper and lower tail of the distribution. In the light of the findings in Hurd et al. (1998) where anchoring effects

are detected, the use of bounds might be called for instead of an imputed variable that ignores possible anchoring effects.

The bounds are estimated allowing for different types of anchoring each generalizing a different parametric model in the existing literature. The paper does not analyse which model of anchoring is most appropriate: this is not a relevant question for this framework. With the current data, however, selective nonresponse and anchoring interact, and it is hard to say something about anchoring without making strong assumptions about the nature of nonresponse. For an analysis of anchoring itself, therefore, experimental data where all respondents get bids that vary randomly across the sample, such as in the experimental HRS module used by Hurd et al. (1998), is more appropriate. With more knowledge about the nature of the anchoring process, the analysis here could be refined.

References

- Borgers, N and J. Hox, (2001), Item nonresponse in questionnaire research with children, *Journal of Official Statistics*, 17,2, 321-335
- Cameron, T. and J. Quiggin, (1994), Estimation using contingent valuation data from a dichotomous choice with follow-up questionnaire, *Journal of Environmental Economics and Management*, 27, 218-234
- Ginther, D., (2000), Alternative estimates of the effect of schooling on earnings, *Review of Economics and Statistics*, 82,1, 102 - 116
- Green, D., K. Jacowitz, D. Kahneman and D. McFadden (1998), Referendum contingent valuation, anchoring, and willingness to pay for public goods, *Resource and Energy Economics*, 20, 85-116
- Härdler, W., (1990), Applied nonparametric regression, *Econometric Society Monographs*, 19, Cambridge University Press, Cambridge
- Herriges, J. and C. Shogren, (1996), Starting point bias in dichotomous choice valuation with follow-up questioning, *Journal of Environmental Economics and Management*, 30, 112-131
- Horowitz, J. and C. Manski, (1995), Identification and robustness with contaminated and corrupted data, *Econometrica*, 63,2, 281-302
- Horowitz, J. and C. Manski, (1998), Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations, *Journal of Econometrics*, 84, 37-58
- Hurd, M., D. McFadden, H. Chand, L. Gan, A. Merrill and M. Roberts, (1998), Consumption and savings balances of the elderly: Experimental evidence on survey response bias, in D. Wise (ed.), *Frontiers in the Economics of Aging*, University of Chicago Press, Chicago, 353-387
- Jacowitz, K. and D. Kahneman (1995), Measures of anchoring in estimation tasks, *Personality and Social Psychology Bulletin*, 21, 1161- 1166
- Juster, T. and J. Smith (1997), Improving the quality of economic data: Lessons from the HRS and AHEAD, *Journal of the American Statistical Association*, 92, 1268-1278
- Manski, C., (1989), Anatomy of the selection problem, *Journal of Human Resources*, 24, 343-360
- Manski, C., (1990), Non-parametric bounds on treatment effects, *American Economic Review, Papers & Proceedings*, 80, 319-323
- Manski, C., (1994), The selection problem in 'Advances in Econometrics', C. Sims (ed.), Cambridge University Press, 143-170
- Manski, C., (1995), Identification problems in the social sciences, Harvard University Press
- Manski, C., (1997), Monotone treatment response, *Econometrica*, 68, 1311-1334
- Manski, C., and J. Pepper, (2000), Monotone instrumental variables with an application to the returns to schooling, *Econometrica*, 68, 997-1010

- Mason, R., V. Lesser, and M. Traugott, Effect of item nonresponse on nonresponse error and inference, In Groves, R., D. Dillman, J. Eltinge and R. Little (eds.), *Survey Nonresponse*, Wiley, New York, 2002.
- O'Connor, R. M. Johansson and P. Johansson (1999), Stated preferences, real behaviour and anchoring: some empirical evidence, *Environmental and Resource Economics*, 13, 235-248.
- Pepper, J. (2000), The inter-generational transmission of welfare receipts: a non-parametric bounds analysis, *The Review of Economics and Statistics*, 82, 472-488
- Pepper, J. (2003), Using experiments to evaluate performance standards, what do welfare to work demonstrations reveal to welfare reformers?, *Journal of Human Resources*, 38,4, 860-880
- Rabin, M. (1998), Psychology and Economics, *Journal of Economic Literature*, 11-46.
- Switzer, F. and J. Snizek, (1991), Judgement processes in motivation: anchoring and judgement effects on judgement and behaviour, *Organizational Behaviour and Human Decision Process*, Vol. 49, 208-229
- Tourangeau, R., L. Rips and K. Rasinski (2000), *The psychology of survey response*, New York and Cambridge University Press
- Tversky, A. and D. Kahneman, (1974), Judgement under uncertainty: Heuristics and biases, *Science*, 185, 1124-1131
- Vazquez-Alvarez, R., (2006), Anchoring effects and covariates nonresponse, Department of Economics DP series, University of St.Gallen
- Wilson, T. D., C. Houston, K. Etling, and N. Brekke (1996), A new look at anchoring effects: basic anchoring and its antecedents, *Journal of Experimental Psychology*,. 135, 387-402

Appendix 1: Deriving bounds (13) on $F(y|x, BR)$ with Assumption A.1

Bounds in (13) result from applying conditions in (12) to derive upper and lower bounds on $P(Y < B_\tau | x, BR)$. Conditions in (12) imply the following:

Let B_τ be the bid that follows $B_{-\tau}$ from $Q_{-\tau}$ and Q_τ :

$$(i) \quad Q_\tau = 1 \ \& \ Q_{-\tau} = 0 \quad \Rightarrow \quad B_\tau > B_{-\tau} \\ \Rightarrow \quad (12.a), (12.e) \text{ apply to bound } P(Y < B_\tau | x, BR)$$

$$(ii) \quad Q_\tau = 1 \ \& \ Q_{-\tau} = 1 \quad \Rightarrow \quad B_\tau > B_{-\tau} \\ \Rightarrow \quad (12.c), (12.e) \text{ apply to bound } P(Y \geq B_\tau | x, BR)$$

$\Rightarrow$ (i) and (ii) define upper and lower bounds on $P(Y < B_\tau | x, BR)$ for a bids B_τ above some bid $B_{-\tau}$ implied by $Q_{-\tau}$, where $Q_{-\tau}$ leads to $B_{-\tau}$

$$(iii) \quad Q_\tau = 0 \ \& \ Q_{-\tau} = 0 \quad \Rightarrow \quad B_\tau < B_{-\tau} \\ \Rightarrow \quad (12.b), (12.e) \text{ apply to bound } P(Y < B_\tau | x, BR)$$

$$(vi) \quad Q_\tau = 1 \ \& \ Q_{-\tau} = 0 \quad \Rightarrow \quad B_\tau < B_{-\tau} \\ \Rightarrow \quad (12.d), (12.e) \text{ apply to bound } P(Y \geq B_\tau | x, BR)$$

$\Rightarrow$ (iii) and (vi) define upper and lower bounds on $P(Y < B_\tau | x, BR)$ for a bids B_τ below some bid $B_{-\tau}$ implied by $Q_{-\tau}$, where $Q_{-\tau}$ leads to $B_{-\tau}$ (A1.1)

Apply the rubric in (A1.1) to both $P(Y < B_\tau | x, BR)$ and $P(Y \geq B_\tau | x, BR)$, where the following interpretation applies to link reported fractions and unknown conditional distributions:

$$\begin{aligned}
(a) \quad P(Y < B_{\tau} | x, BR) &= P(Y < B_{\tau} | x, BR, Q_{\tau} = 1)P(Q_{\tau} = 1 | x, BR) \\
&\quad + P(Y < B_{\tau} | x, BR, Q_{\tau} = 0)P(Q_{\tau} = 0 | x, BR) \\
\text{and} & \\
\end{aligned} \tag{A1.2}$$

$$\begin{aligned}
(b) \quad P(Y \geq B_{\tau} | x, BR) &= P(Y \geq B_{\tau} | x, BR, Q_{\tau} = 1)P(Q_{\tau} = 1 | x, BR) \\
&\quad + P(Y \geq B_{\tau} | x, BR, Q_{\tau} = 0)P(Q_{\tau} = 0 | x, BR)
\end{aligned}$$

Once the position of B_{τ} is decided relative to $B_{-\tau}$, i.e., $B_{\tau} < B_{-\tau}$ or $B_{\tau} > B_{-\tau}$,²¹ deriving the two bounds in (13) is rooted in applying the two sets of results in (A1.1) to (A1.2). We proceed to show the application of (i) and (ii) to (A1.2):

For $B_{\tau} > B_{-\tau}$, (12.a) and (12.e) bounds $P(Y < B_{\tau} | x, BR)$ from above:

$$\begin{aligned}
(a) \quad P(Y < B_{\tau} | x, BR) &= P(Y < B_{\tau} | x, BR, Q_{\tau} = 0)P(Q_{\tau} = 0 | x, BR) \\
&\quad + P(Y < B_{\tau} | x, BR, Q_{\tau} = 1)P(Q_{\tau} = 1 | x, BR) \\
&\leq P(Y < B_{\tau} | x, BR)P(Q_{\tau} = 0 | x, BR) && \left(\text{applying (12.e)} \right) \\
&\quad + 2P(Q_{\tau} = 0 | x, BR, Q_{-\tau} = 1)P(Q_{\tau} = 1 | x, BR) && \left(\text{applying (12.a)} \right) \\
\text{i.e.,} &\leq 2P(Q_{\tau} = 0 | x, BR, Q_{-\tau} = 1)
\end{aligned}$$

$$\text{Upper Bound} \Rightarrow P(Y < B_{\tau} | x, BR) \leq \min[1, 2P(Q_{\tau} = 0 | x, BR, Q_{-\tau} = 1)]$$

For $B_{\tau} > B_{-\tau}$, (12.c) and (12.e) bounds $P(Y \geq B_{\tau} | x, BR)$ from below:

$$\begin{aligned}
(b) \quad P(Y \geq B_{\tau} | x, BR) &= P(Y \geq B_{\tau} | x, BR, Q_{\tau} = 0)P(Q_{\tau} = 0 | x, BR) \\
&\quad + P(Y \geq B_{\tau} | x, BR, Q_{\tau} = 1)P(Q_{\tau} = 1 | x, BR) \\
&\leq P(Y \geq B_{\tau} | x, BR)P(Q_{\tau} = 0 | x, BR) && \left(\text{applying (12.e)} \right) \\
&\quad + 2P(Q_{\tau} = 1 | x, BR, Q_{-\tau} = 1)P(Q_{\tau} = 1 | x, BR) && \left(\text{applying (12.c)} \right) \\
\text{i.e.,} &\leq 2P(Q_{\tau} = 1 | x, BR, Q_{-\tau} = 1) \\
\Rightarrow P(Y \geq B_{\tau} | x, BR) &\leq \min[1, 2P(Q_{\tau} = 1 | x, BR, Q_{-\tau} = 1)] \\
\Rightarrow 1 - P(Y \geq B_{\tau} | x, BR) &\geq 1 - \min[1, 2P(Q_{\tau} = 1 | x, BR, Q_{-\tau} = 1)] && (A1.3)
\end{aligned}$$

$$\text{Lower Bound} \Rightarrow P(Y < B_{\tau} | x, BR) \geq \max[0, 1 - 2P(Q_{\tau} = 1 | x, BR, Q_{-\tau} = 1)]$$

The result in (A1.3) is shown in Section 2 with (13.b). Deriving bounds in (13.a) follow in the same manner, but using (iii) and (iv) in (A1.1) in (A1.2) instead of (i) and (ii). Notice that the result in

²¹ For example, $B_{\tau} = \$5,000$ is such that $B_{\tau} < B_{-\tau}$ if compared to $B_{-\tau} = \$25,000$, whereas $B_{\tau} = \$50,000$ is such that $B_{\tau} > B_{-\tau}$ if compared to $B_{-\tau} = \$25,000$.

(A1.3) is not trivial when compared to (12.a), i.e., Assumption A1 leads to a bound on the conditional distribution $P(Y < B_{\tau} | x, BR, Q_{-\tau} = 1)$, and we use this together with (12.e) to derive an upper bound on $P(Y < B_{\tau} | x, BR)$ given in (A1.3).

Appendix 2: Characterizing bounds (7), (13), (18) and (22)

Section 4.1, Table 1 defines an unfolding bracket design such that respondents can face as many as 3 bids. The unfolding design is asymmetric in the sense that not all respondents face the same number of bids. Only those follow the design towards the higher conditional bids face at most 3 bids. Otherwise, individuals face at most 2 bids. With this, those who become CBR are able to classify the otherwise undisclosed amount in one out of five categories: $[0, B_{20})$, $[B_{20}, B_1)$, $[B_1, B_{21})$, $[B_{21}, B_{311})$ and $[B_{311}, \infty)$; empirically, all these categories are nonempty. The IBR subgroup defines three categories, $[0, B_1)$, $[B_1, \infty)$, and $[B_{21}, \infty)$, and all three are empirically nonempty. Bounds on $F(y | x, BR)$ are characterized according to these categories.

Characterising Bounds in (7)

Table A2.1: Characterizing $F(y | x, BR)$ for bounds in (7): Assumption A0

	Contribution to Lower Bound on $F(y x, BR)$ weighted by $P(R x, BR)$	Contribution to Upper Bound on $F(y x, BR)$ weighted by $P(R x, BR)$
Complete Bracket Respondents where $P(R x, BR) = P(CBR x, BR)$		
$[0, B_{20})$	0	$P(Q_{20} = 0 x, CBR)$
$[B_{20}, B_1)$	$P(Q_{20} = 0 x, CBR)$	$P(Q_1 \& Q_{20} = 0 x, CBR)$
$[B_1, B_{21})$	$P(Q_1 \& Q_{20} = 0 x, CBR)$	$P(Q_{21} \& Q_1 \& Q_{20} = 0 x, CBR)$
$[B_{21}, B_{311})$	$P(Q_{21} \& Q_1 \& Q_{20} = 0 x, CBR)$	$P(Q_{311} \& \dots \& Q_{20} = 0 x, CBR)$
$[B_{311}, \infty)$	$P(Q_{311} \& \dots \& Q_{20} = 0 x, CBR)$	1
Complete Bracket Respondents where $P(R x, BR) = P(IBR x, BR)$		
$[0, B_1)$	0	$P(Q_1 = 0 x, IBR)$
$[B_1, \infty)$	$P(Q_1 = 0 x, IBR)$	1
$[B_{21}, \infty)$	$P(Q_1 \& Q_{21} = 0 x, IBR)$	1

Characterising Bounds in (13)

A pre-requisite is to define the relation between B_{τ} & $B_{-\tau}$ for each bid in $[B_{\tau-1}, B_{\tau})$. For example, Table 1 shows that to be in $[0, B_{20})$, B_{20} is determined because $Q1 = 0$ to $B1$ and $Q20 = 0$. Thus, relative to $B1 = (B_{-\tau})$, $B_{20} < B1$ and (13.a) applies to B_{20} for bounds on $[B_{\tau-1}, B_{\tau}) = [0, B_{20})$; it implies that $\{l(B_{\tau}), u(B_{\tau})\} = \{\max[0, 1 - 2P(Q20 = 1 | x, BR, Q1 = 0)], \min[1, 2P(Q20 = 0 | x, BR, Q1 = 0)]\}$ whereas $\{l(B_{\tau-1}), u(B_{\tau-1})\} = \{0, 0\}$. From (13), $L_{\tau-1, \tau} = \min\{0, \max[0, 1 - 2P(Q20 = 1 | \dots)]\}$ so that $L_{\tau-1, \tau} = 0$, while $U_{\tau-1, \tau} = \max\{0, \min[1, 2P(Q20 = 0 | x, BR, Q_{-\tau} = 0)]\}$ so that $U_{\tau-1, \tau} = \min[1, 2P(Q20 = 0 | x, BR, Q_{-\tau} = 0)]$. Similar arguments determine all entries in Table A2.2 below:

Table A2.2: Characterizing $F(y | x, BR)$ for bounds in (13): Assumption A1

	Contribution to Lower Bound on $F(y x, BR)$ weighted by $P(R x, BR)$	Contribution to Upper Bound on $F(y x, BR)$ weighted by $P(R x, BR)$
Complete Bracket Respondents where $P(R x, BR) = P(CBR x, BR)$		
$[0, B_{20})$	0	$\max\left\{\begin{matrix} 0 \\ \min[1, 2P(Q20 = 0 x, CBR, Q1 = 0)] \end{matrix}\right\}$
$[B_{20}, B1)$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q20 = 1 x, CBR, Q1 = 0)] \\ \max[0, 1 - 2P(Q1 = 1 x, CBR)] \end{matrix}\right\}$	$\max\left\{\begin{matrix} \min[1, 2P(Q20 = 0 x, CBR, Q1 = 0)] \\ \min[1, 2P(Q1 = 0 x, CBR)] \end{matrix}\right\}$
$[B1, B_{21})$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q1 = 1 x, CBR)] \\ \max[0, 1 - 2P(Q21 = 1 x, CBR, Q1 = 1)] \end{matrix}\right\}$	$\max\left\{\begin{matrix} \min[1, 2P(Q1 = 0 x, CBR)] \\ \min[1, 2P(Q21 = 0 x, CBR, Q1 = 1)] \end{matrix}\right\}$
$[B_{21}, B_{311})$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q21 = 1 x, CBR, Q1 = 1)] \\ \max[0, 1 - 2P(Q311 = 1 x, CBR, Q21 = 1)] \end{matrix}\right\}$	$\max\left\{\begin{matrix} \min[1, 2P(Q21 = 0 x, CBR, Q1 = 1)] \\ \min[1, 2P(Q311 = 0 x, CBR, Q21 = 1)] \end{matrix}\right\}$
$[B_{311}, \infty)$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q311 = 1 x, CBR, Q21 = 1)] \\ 1 \end{matrix}\right\}$	1
Complete Bracket Respondents where $P(R x, BR) = P(IBR x, BR)$		
$[0, B1)$	0	$\max\left\{\begin{matrix} 0 \\ \min[1, 2P(Q1 = 0 x, IBR)] \end{matrix}\right\}$
$[B1, \infty)$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q1 = 1 x, IBR)] \\ 1 \end{matrix}\right\}$	1
$[B_{21}, \infty)$	$\min\left\{\begin{matrix} \max[0, 1 - 2P(Q21 = 1 x, IBR, Q1 = 1)] \\ 1 \end{matrix}\right\}$	1

Characterising Bounds in (18)

Table 3, Section 4.1 determines that in our empirical application $B20$ is perceived and should act as a low anchor, whereas $B1$, $B21$ and $B311$ are perceived and act as high anchors. These determine the application of (18.a) to $B20$, and the application of (17a) to $B1$, $B21$ and $B311$. Entries in Table A2.3 below follow:

Table A2.3: Characterizing $F(y | x, BR)$ for bounds in (18): Assumption A2

	Contribution to Lower Bound on $F(y x, BR)$ weighted by $P(R x, BR)$	Contribution to Upper Bound on $F(y x, BR)$ weighted by $P(R x, BR)$
Complete Bracket Respondents where $P(R x, BR) = P(CBR x, BR)$		
$[0, B20)$	$\min \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = 0$	$\max \begin{Bmatrix} 0 \\ P(Q20 = 0 x, CBR, Q1 = 0) \end{Bmatrix}$
$[B20, B1)$	$\min \begin{Bmatrix} 0 \\ P(Q1 = 0 x, CBR) \end{Bmatrix}$	$\max \begin{Bmatrix} P(Q20 = 0 x, CBR, Q1 = 0) \\ 1 \end{Bmatrix}$
$[B1, B21)$	$\min \begin{Bmatrix} P(Q1 = 0 x, CBR) \\ P(Q21 = 0 x, CBR, Q1 = 1) \end{Bmatrix}$	$\max \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = 1$
$[B21, B311)$	$\min \begin{Bmatrix} P(Q21 = 0 x, CBR, Q1 = 1) \\ P(Q311 = 0 x, CBR, Q21 = 1) \end{Bmatrix}$	$\max \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = 1$
$[B311, \infty)$	$\min \begin{Bmatrix} P(Q311 = 0 x, CBR, Q21 = 1) \\ 1 \end{Bmatrix}$	$\max \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = 1$
Complete Bracket Respondents where $P(R x, BR) = P(IBR x, BR)$		
$[0, B1)$	$\min \begin{Bmatrix} 0 \\ P(Q1 = 0 x, IBR) \end{Bmatrix}$	$\max \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = 1$
$[B1, \infty)$	$\min \begin{Bmatrix} P(Q1 = 0 x, IBR) \\ 1 \end{Bmatrix}$	$\max \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = 1$
$[B21, \infty)$	$\min \begin{Bmatrix} P(Q21 = 0 x, IBR, Q1 = 1) \\ 1 \end{Bmatrix}$	$\max \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = 1$

Notice that in the case of CBR, when a ‘perceived’ low bounds and a ‘perceived’ high bound mix to create a category, in this case $[B20, B1)$, the identification region according to the Jacowitz and Kahneman assumption are not informative, that is, the upper and lower bounds imply a $[0, 1]$ interval. The same applies to $[0, B1)$ for IBR, where the argument is the same if we define the natural lower bound of ‘zero’ as a ‘perceived’ lower bound.

Characterising Bounds in (22)

Table A2.4: Characterizing $F(y | x, BR)$ for bounds in (22): Assumption A3

	Contribution to Lower Bound on $F(y x, BR)$ weighted by $P(R x, BR)$	Contribution to Upper Bound on $F(y x, BR)$ weighted by $P(R x, BR)$
Complete Bracket Respondents where $P(R x, BR)$ $= P(CBR x, BR)$		
$[0, B20)$	$\min \left\{ \begin{array}{c} 0 \\ P(Q20 = 0 x, CBR, Q1)P(Q1 = 0 x, CBR) \end{array} \right\}$	$\max \left\{ \begin{array}{c} 0 \\ P(Q1 = 0 x, CBR) \end{array} \right\}$
$[B20, B1)$	$\min \left\{ \begin{array}{c} P(Q20 = 0 x, CBR, Q1)P(Q1 = 0 x, CBR) \\ P(Q1 = 0 x, CBR) \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) \\ P(Q1 = 0 x, CBR) \end{array} \right\}$
$[B1, B21)$	$\min \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) \\ P(Q1 = 0 x, CBR) \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) \\ P(Q1 = 0 x, CBR) + \\ + P(Q21 = 0 x, CBR, Q1 = 1)P(Q1 = 1 x, CBR) \end{array} \right\}$
$[B21, B311)$	$\min \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) \\ P(Q1 = 0 x, CBR) + \\ + P(Q21 = 0 x, CBR, Q1 = 1)P(Q1 = 1 x, CBR) \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) + \\ + P(Q21 = 0 x, CBR, Q1 = 1)P(Q1 = 1 x, CBR) \\ P(Q21 = 0 x, CBR) + \\ + P(Q311 = 0 x, CBR, Q21 = 1)P(Q21 = 1 x, CBR) \end{array} \right\}$
$[B311, \infty)$	$\min \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) + \\ + P(Q21 = 0 x, CBR, Q1 = 1)P(Q1 = 1 x, CBR) \\ 1 \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q21 = 0 x, CBR) + \\ + P(Q311 = 0 x, CBR, Q21 = 1)P(Q21 = 1 x, CBR) \\ 1 \end{array} \right\}$
Complete Bracket Respondents where $P(R x, BR)$ $= P(IBR x, BR)$		
$[0, B1)$	$\min \left\{ \begin{array}{c} 0 \\ P(Q1 = 0 x, IBR) \end{array} \right\}$	$\max \left\{ \begin{array}{c} 0 \\ P(Q1 = 0 x, IBR) \end{array} \right\}$
$[B1, \infty)$	$\min \left\{ \begin{array}{c} P(Q1 = 0 x, IBR) \\ 1 \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q1 = 0 x, IBR) \\ 1 \end{array} \right\}$
$[B21, \infty)$	$\min \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) \\ 1 \end{array} \right\}$	$\max \left\{ \begin{array}{c} P(Q1 = 0 x, CBR) + \\ + P(Q21 = 0 x, CBR, Q1 = 1)P(Q1 = 1 x, CBR) \\ 1 \end{array} \right\}$