CentER

Discussion Paper

No. 2004–27

A TEST FOR ANCHORING AND YEA-SAYING IN EXPERIMENTAL CONSUMPTION DATA

By A.H.O. van Soest, M. Hurd

March 2004

ISSN 0924-7815

TILBURG UNIVERSITY

A Test for Anchoring and Yea-Saying in Experimental Consumption Data[1]

Arthur van Soest and Michael Hurd

RAND Corporation

December 2003

Abstract

In the experimental module of the AHEAD 1995 data, the sample is randomly split into respondents who get an open-ended question on the amount of total family consumption - with follow-up unfolding brackets (of the form: is consumption $X or more?) for those who answer "don't know" or "refuse" - and respondents who are immediately directed to unfolding brackets. In both cases, the entry point of the unfolding bracket sequence is randomized. These data are used to develop a nonparametric test for whether people make mistakes in answering the first bracket question, allowing for any type of selection into answering the open-ended question or not. Two well-known types of mistakes are considered: anchoring and yea-saying (or acquiescence). While the literature provides ample evidence that the entry point in the first bracket question serves as an anchor for follow-up bracket questions, it is less clear whether the answers to the first bracket question are already affected by anchoring. We reject the joint hypothesis of no anchoring and no yea-saying at the entry point. Once yea-saying is taken into account, there is no evidence of anchoring.

Correspondence: mhurd@rand.org, vansoest@rand.org; RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138, U.S.A.

JEL classification: C81, D12
Key words: Unfolding brackets, Non-response, Anchoring, Yea-saying

---

[1] Financial support was provided by the National Institute of Aging. The comments of Arie Kapteyn and seminar participants at Bonn University and the University of Southern California Los Angeles are gratefully acknowledged.

1. Introduction

Item non-response in survey data is a well-known problem, particularly if questions relate to sensitive information that respondents are not willing to provide or to information that respondents do not know exactly and find hard to estimate. Examples are the amounts invested in saving accounts or the value of assets such as stocks and bonds, or the value of total family income or consumption. Item non-response becomes particularly problematic if the information is not missing at random, i.e., if the probability of not responding correlates with the amount in question conditional on a set of always observed covariates. See, for example, Manski (1989, 1995) and Little and Rubin (2002, Chapter 15).

Follow-up bracketing questions are often used to reduce the loss of information due to item non-response and to extract at least partial information from initial non-respondents. See, for example, Juster and Smith (1997). Particularly in telephone interviews where it is difficult to show range cards on which respondents can choose in one step from a larger number of categories, unfolding brackets can be used to collect banded information. For example, a respondent who answers "don't know" or "refuse" to a question on total family consumption in the past month then gets the question "Can you say whether it was $2000 or more?" If the answer is affirmative, the next question is "Was it $5000 or more?" etc.

The majority of initial non-respondents typically appear to be willing to answer one or more unfolding bracket questions, making unfolding brackets a useful tool to collect at least some information from the initial non-respondents (Juster and Smith, 1997). Several experiments, however, have shown that the distribution of the categorical answers obtained in an unfolding bracket design depends on the order in which the bracket points are presented. For example, if the first question is "Was it $ 2000 or more?" i.e., the entry point is $2000, the distribution gives more weight to the high consumption outcomes than if the entry point were $500. A psychological explanation is that if people are unsure about the exact amount, the entry point will serve as an anchor that provides some information about their own amount. This phenomenon is known as anchoring (see Tversky and Kahneman, 1974).

The literature shows that anchoring exists and becomes more prominent the more uncertain the respondent is about the exact answer (see, for example, Jacowitz and Kahneman, 1995). In order to use the answers to unfolding bracket questions for meaningful analysis, the possibility of an anchoring bias needs to be taken into account. Several models for what the answers to bracket questions look like if they suffer from anchoring have been introduced and estimated, using different sources of experimental data with random entry points. Examples are Hurd et al. (1998), Herriges and Shogren (1996), and Cameron and Quiggin (1994). A comparison between competing models for anchoring, however, has not yet been performed.

In this paper, some existing models for anchoring are considered and their implications are analyzed for differences between the true distribution of the variable of interest and the distribution obtained if anchoring according to a given model is present. Experimental data on household consumption from the AHEAD 1995 survey will then be used to test some of these implications. The specific feature of these experimental data on total family consumption is that the sample is randomly split in respondents that start with an open question - and get follow-up brackets if they do not answer that - and respondents who are immediately directed to the unfolding bracket questions without being asked an open question (see Hurd et al., 1998). Although in the first sub-sample initial non-response is substantial and respondents to the open-ended question are probably not a random sub-sample, it will be shown that comparing the combined distribution of open responses and bracket responses in the first sub-population with the distribution in the sub-population of immediate bracket responses provides a test for certain types of anchoring.

Another problem that leads to incorrect answers in an unfolding bracket design is acquiescence or “yea- saying” This is the problem that people have a tendency to answer yes rather than no. See Schuman and Presser (1981) or Hurd (1999). It implies that answers may depend on the wording of the question. For example, the number of “yes” answers to the question “Was it 2000 or more?” will be higher than the number of “no” answers to the question “Was it less than 2000?” The design of the experimental data used here makes the answers vulnerable to yea-saying.

The remainder of this paper is organized as follows. In the next section, some existing models for anchoring are reviewed. In Section 3, we describe the experimental data. Section 4 explains how these data can be used to distinguish between anchoring models in a simple, nonparametric framework. This is applied in Section 5. In sections 4 and 5, full non-respondents who provide neither open-ended nor bracket information are discarded. Section 6 takes them into account, using Manski's worst-case bounds (Manski 1989). Section 7 incorporates a simple way of "fixed probability" anchoring and acquiescence in the model and leads to the conclusion that acquiescence alone can explain the rejections in Section 5. Moreover, it shows that ignoring the acquiescence problem leads to misleading conclusions about selectivity of open-ended answers and the distribution of consumption that we are ultimately interested in. It also looks at estimates of the acquiescence problem for different groups, defined by, for example, education level or memory skills. Section 8 concludes.

2. Anchoring Models and Acquiescence Bias

An intuitively appealing anchoring model is the fixed-point model used by Herriges and Shogren (1996), Hurd (1997), O'Connor et al. (1999), and Hurd et al. (2001). In this model, the entry point E in the first bracket question serves as an anchor for follow-up questions. In the second bracket question the respondent does not compare E to the true amount Y, but to the weighted mean $Y^* = (1-\alpha)Y + \alpha E$ for some α between 0 and 1. The intuition is that the respondent is uncertain about the true amount. The entry point E is seen as an indication of what Y could be, and the respondent's updated estimate Y^* is drawn towards E. Herriges and Shogren (1996) apply their model to data on willingness to pay for water quality improvement and find an estimate for α of 0.36, with standard error 0.14. They also discuss the possibility that α varies with Y and E.

An important feature of the fixed-point model is that answers to the first bracket question are not affected. Although the respondent's estimate of the true amount will be drawn towards the entry point, it will remain on the same side. Thus if anchoring according to the fixed-point model is the only source of incorrect answers, answers to the first bracket question can be taken at face value. The fixed-point model can be interpreted as a Bayesian model, where respondents update their beliefs about an unknown amount if new information (in the form of an entry point or other bracket point) arrives.

Jacowitz and Kahneman (1995) and Green et al. (1998) find evidence suggesting that respondents already give biased answers to the first bracket question. For estimates of objective quantities as well as willingness to pay for public goods, they find that, for high entry point values, the proportion of bracket respondents who report that the amount is larger than the entry point value exceeds the fraction of open-ended answers larger than the entry point value.

Hurd et al. (1998) specify a parametric model that captures this phenomenon in a symmetric way, biasing probabilities smaller than 0.5 upwards and probabilities larger than 0.5 downward. The idea is that respondents do not compare the true value Y to the entry point E, but instead compare Y to $E+\varepsilon$, where ε is a mean zero error term, assumed to be normal and independent of all other components of the model. Hurd et al. (1998) use the same device also at follow-up bracket questions, with independent errors that can have a different variance. The model is called a gating model, since respondents have to pass a number of gates to reach their final bracket answer.

Somewhat similar, though not yet applied in this context, are the models for binary regression with contaminated data (e.g., Copas, 1988) or misclassification (e.g. Hausman et al., 1998). Assume that with some fixed “gating” probability P(Gat), people give the wrong answer. It implies $P[\text{“no”}] = P[Y<E](1-P(Gat)) + P[Y \geq E]P(Gat)$ and $P[\text{“yes”}] = P[Y \geq E](1-P(Gat)) + P[Y<E]P(Gat)$. If this probability does not depend on the true value and is thus the same whether the true amount is larger or smaller than the entry point, the reported probabilities will be too large if they are larger than 0.5 and too small if they are smaller than 0.5.

A direct test of yea-saying is preformed by Hurd (1999). He uses experimental data from the 1996 wave of the Health and Retirement Study, a representative sample of the US population aged 54-64 with their spouses. Respondents were asked “About how much could you sell your home for in today’s housing market?” They were forced into brackets and were randomly assigned to an entry point (E=$50,000, E=$100,000 or E=$150,000) and to a question format: “Would it be more than E,” “Would it be E or more?” or “Would it be less than E or more than E?” He found that the first two (unbalanced) formats led to many more “yes” answers than the third (balanced) format for non-financial respondents, while there was hardly any difference for financial respondents (i.e., the person in the household most knowledgeable in financial matters). He interpreted this as evidence of acquiescence related to uncertainty. Acquiescence bias is also a well-known problem in the willingness-to-pay (WTP) literature. For example, Boyle et al. (1998) find evidence of yea-saying for bid levels in the upper tail by comparing open-ended answers and bracket answers on the WTP for a moose hunting site. On the other hand, Frykblom and Shogren (2000) used experimental data to compare open-ended and discrete choice answers on students’ WTP for a book (“The Environment”) and found no differences, leading them to conclude that problems with discrete choice answers can be due to how the survey is framed rather than to the questions themselves.

3. The AHEAD Wave 2 Consumption Experiment

We use basically the same data as Hurd et al. (1998), who describe the data in detail. The AHEAD panel (Assets and Health Dynamics Among the Oldest Old) is roughly representative for the US population over 70 and their spouses. The second wave was mainly done by telephone interviews in 1995. At the end of the regular survey, regular AHEAD participants were asked to complete an additional experimental module. About 75% of all respondents were willing to do this. This is the sample used by Hurd et al. (1998) and for this study.

The participants in the experimental module were randomly separated into 7 groups of approximately equal size. All these groups got some questions on household consumption in the last month, but the question format differed across the seven groups. Three groups (sample I) started with an open question: “How much did your household spend on consumption in the past month?”[2] Those who answered “don’t know” or “refuse” then got unfolding bracket questions, with different entry points for the three groups. The first unfolding bracket question was formulated as “Would the amount be $E

---

[2] The exact wording of the question was: “About how much did you and your household spend on everything in the past month? Please think about all bills such as rent, mortgage loan payments, utility and other bills, as well as all expenses such as food, clothing, transportation, entertainment and any other expenses you and your household may have.”

or more in the past month?” with E=500, E=1000 or E=2000. If a respondent answered “don’t know” or “refuse” to a bracket question, the sequence was stopped. The other four groups (sample II) immediately were given bracket questions, with different entry points for the four groups (500, 1000, 2000 and 5000).

This sample consists of 4928 observations. About 2.6% of them did not give an answer to either an open-ended question or a follow-up or direct bracket question; these observations are discarded for the main part of the analysis but will be incorporated in Section 6. This leaves 4759 observations. About 42% of them (sample I) started with an open question. Almost two thirds of these gave an open-ended answer (1416 observations). The remaining 681 gave at least one bracket answer. The first panel of Table 1 presents the fractions with consumption less than each of the entry point values in the open-ended answers and according to the follow-up bracket answers, separately for the three entry point groups. There are several explanations for differences between open-ended answers and the bracket answers and between bracket answers with different entry points. High consumption families could be more concerned about their privacy and less willing to give a precise number, i.e., open-ended answers could under-represent high amounts (selection into open answers is correlated with consumption level, i.e., missing values in open-ended answers are not completely random). Alternatively, respondents could make mistakes in either the open-ended answers or in their bracket answers or both. In particular, the difference between the distributions of bracket responses for the three entry points suggest that some bracket answers may depend on the (randomly assigned) entry point. In other words, there is an anchoring problem. Significance levels are rather low, however, due to the relatively small numbers of observations.

Table1. Distribution of Consumption in Experimental Data AHEAD 1995

Sample I: open-ended question first

	Sub-sample: Entry point 500	Entry point 1000	Entry point 2000	All entry points
observations	674	698	725	2097
obs. open answer	435	485	496	1416
(percentage)	(64.5)	(69.5)	(68.4)	(67.5)
% of open answers				
< 500	16.6	16.9	17.3	16.9
< 1000	46.4	49.7	50.8	49.1
< 2000	80.0	81.0	83.3	81.5
obs. bracket answer	239	213	229	681
(percentage)	(35.5)	(30.5)	(31.6)	(32.5)
% of bracket answers				
< 500	19.7	15.2	11.1	
< 1000	47.4	49.3	40.4	
< 2000	91.9	85.0	75.1	

Sample II: forced brackets

	Sub-sample: Entry point 500	Entry point 1000	Entry point 2000	Entry point 5000
observations	690	612	582	778
% of bracket answers				
< 500	17.0	11.8	7.6	6.4
< 1000	57.0	41.7	31.1	32.2
< 2000	86.4	83.5	70.1	65.2
< 5000	98.9	98.8	97.4	94.1

Note: respondents who answer neither an open question nor the first bracket question are not included.

The bottom panel of Table 1 shows the distributions for the immediate bracket respondents by entry point. Again, non-random response behavior might in principle explain why the open-ended answers tend to be small compared to the bracket responses. A salient feature of the figure is the fact that the distribution shifts with the entry point. In this case the numbers of observations are much larger and most of the differences are significant. The anchoring models discussed in the previous section provide several explanations for these differences.

There seems to be common agreement in the literature that fixed-point anchoring affects the answers to bracket questions from the second bracket question onwards. The entry point in the first bracket question acts as an anchor and the respondent’s estimate is pushed towards it. This phenomenon makes intuitive sense and is studied relatively often in the literature and is not the issue of the current paper. Fixed-point anchoring, however, does not affect the answer to the first bracket question. In this paper we focus on errors in answers to the first bracket questions. We do not analyze the answers to the later bracket question, thus avoiding the need to cope with fixed-point anchoring.

Our nonparametric testing strategy is illustrated in Figure 1, which is largely derived from the numbers in Table 1. The solid line is the empirical distribution of open-ended answers. The other curves use only the answers to the first bracket question to estimate the distributions for the samples of immediate bracket respondents and follow-up bracket respondents. For example, for the immediate bracket respondents, the fraction of households with consumption less than $2000 is based upon the group of immediate respondents who got $2000 as their entry point. This is the point (20,0.701) on the short-dashed dash-dotted curve in Figure 1. Similarly, the fraction of households with consumption less than $1000 is based upon the sub-sample of immediate bracket respondents with entry point $1000, etc.

The three distributions in Figure 1 will differ if non-response to the open-ended question is non-random. If all answers are correct, 81.0% in the sub-sample of sample I that gave an open-ended answer have consumption less than $2000, compared to 75.1% in the sub-sample who gave a follow-up bracket answer. There is no reason why these two numbers should be estimates of the same thing, since whether or not a respondent gives an open answer may be correlated with the level of household consumption (missing open answers are nonrandom; selective non-response). An estimate for the fraction of people with household consumption less than $2000 in the total population of interest on the basis of sample I is 0.675*0.81+0.325*0.751=0.791. Under the assumption that all answers are correct, this estimate is consistent under any form of selective non-response between open answers and follow-up bracket answers.[3] The open-ended answers (solid curve) might be negatively selected, implying that follow-up bracket answers (long-dashed curve) are positively selected.

[3] It is assumed that selection does not depend on whether the follow-up bracket questions start with $500, $1000 or $2000. This seems quite reasonable since entry points are assigned randomly and people do not know their entry point at the time they have to decide whether to give an open answer or not. Moreover, this can easily be tested for and is not rejected.

The immediate bracket responses are, because of the random assignment, drawn from the same population as the combined sample of open-ended and follow-up bracket respondents. This implies that, if all answers to open-ended questions and first bracket questions are correct, the distribution of immediate bracket respondents should be a weighted mean of the other two distributions. Figure 1 suggests this is not the case; the point estimate for sample II is only 0.701. In the next section, we develop a formal test based upon this intuition that looks at these differences at several entry points.

4. Testing the Fixed-Point Model of Anchoring

A major distinction between the models discussed in section 2 is their implication for the first bracket point. The fixed-point model assumes that the entry point acts as an anchor for later bracket questions but does not affect whether the amount is larger or smaller than or equal to the first entry point itself, and therefore does not lead to an incorrect answer to the first bracket question. In other words, the fixed-point model implies that anchoring does not bias the answers to the first bracket question. The same applies to any Bayesian model: the first entry point leads to an update of prior information, but will not move the mean or any other location measure of the respondent's subjective distribution past the entry point. Thus a Bayesian model will not lead to biased answers to the first bracket question.

The gating model, on the other hand, assumes that errors can be made at every bracket question, including the first one. Explaining the stylized fact that the distribution shifts in the direction of the first entry point even requires that the errors made in the first question are typically larger than those in the later stages of the unfolding bracket design (Hurd et al., 1998). Under plausible distributional assumptions, the gating model implies that bracket answers give too high probabilities of both very low outcomes and very high outcomes. If Y is consumption, then, according to the gating model, the fraction of "no" answers in bracket responses will overestimate $P[Y<E]$ if the entry point E is, for example, the first decile of the consumption distribution, and the fraction of "yes" bracket responses will overestimate $P[Y \geq E]$ if E is, for example, the ninth decile. The gating bias on the estimate of $P[Y<E]$ is unclear (zero in case of complete symmetry) if E is the median.

Since all the questions in the experiment are of the form “Is it E or more?” acquiescence bias always goes in one direction: $P[Y<E]$ will be underestimated, $P[Y\geq E]$ will be overestimated, no matter whether E is larger or smaller than (or equal to) the median.

If samples of open-ended answers and bracket answers drawn randomly from the same population would be available, the arguments given above would make it possible to test the joint null hypothesis of no acquiescence bias and no anchoring in the first bracket question, i.e., anchoring according to the fixed-point model or a Bayesian model, but not according to a gating model. Moreover, if this null hypothesis would be rejected, looking at what happens at a low entry point would make it possible to say whether yea-saying or anchoring according to the gating model is the main reason for rejection. If there is yea-saying, the bracket answers will lead to a larger estimate of $P[Y\geq E]$ than the open-ended answers whether E is large or small, but if there is gating, the bracket answers based estimate of $P[Y\geq E]$ will be smaller when E is small.

The reality of the experiment, however, is slightly more complicated. First, there are respondents who do not reveal any information. In the sample of immediate bracket respondents (sample II), these are the respondents that do not answer the first bracket question. In the sample of those who start with an open-ended question (sample I), some people answer neither the open-ended nor the first bracket question in the follow-up unfolding brackets design (i.e., they always answer “don’t know” or “refuse”). We will assume that the groups of complete non-respondents are the same in the two sub-samples. That is, we assume that people who do not answer an initial bracket question would also not have answered an open-ended question, and people who do not answer a follow-up bracket question would not have answered the same bracket question either if it had been preceded by an open-ended question. In other words: the population actually analyzed excludes the people who do not reveal any information at all, and we assume that both samples (I and II) are random samples from this population. Appendix A presents detailed information on complete non-response for each group. It suggests that no systematic differences between samples I and II can be detected, backing up the assumption that is made. Moreover, in section 6 we show how the assumption can be relaxed using Manski bounds (Manski 1989).

Second, there is a substantial number of respondents in sample I who do not answer the open-ended question but do answer one or more follow-up bracket questions. Whether someone in sample I gives an open-ended answer or a bracket answer may very well be non-random, in which case ignoring the bracket respondents in sample I leads to selection bias. The bracket responses in sample I may suffer from the same anchoring or acquiescence biases as those in sample II. We will combine estimates of $P[Y<E]$ for various (entry point) values of E in the sub-samples of open-ended and bracket respondents in sample I and compare them with estimates on sample II. We will only use the open answers and the first bracket questions, and not the information in other bracket questions: the latter would suffer from anchoring in any anchoring model, including the fixed-point model. Under our null hypothesis, answers to the first bracket question will not suffer from anchoring in either group, and the difference between the sample I and II based estimates should be insignificant. Under the alternative, both the sample I and the sample II based estimates will be biased. Since the majority of the sample I answers are open answers, however – which are assumed to be always correct under the null as well as the alternative – the bias is expected to be larger in sample II than in sample I. This will drive the power of the test and will tell us how to interpret deviations from the null.

To make this more precise, some notation is needed. First consider sample I (excluding complete non-respondents). Let $P[O]$ denote the probability that an open answer is given. For $E=500$, 1000 or 2000, let $P[Y<E|O]$ be the (population) fraction of people with consumption less than E among those who give an open answer.

Let $P[Y <_r E|B]$ be the fraction among initial non-respondents who, if their entry point is E, *report* that their consumption is less than E. Under the null hypothesis, the fraction $P[Y <_r E|B]$ is equal to the fraction $P[Y < E|B]$ of initial non-respondents whose true consumption level is less than E. Due to selective initial non-response, $P[Y < E|B]$ and $P[Y<E|O]$ can be different, so that $P[Y <_r E|B]$ and $[Y<E|O]$ can also be different, even under the null.

With the survey design of sample I, the fraction of people who report consumption less than E in the whole population (again, excluding full non-respondents, as explained above) is equal to

$$P[Y<_r E|I] = P[O]\,P[Y<E|O] + (1-P[O])\,P[Y <_r E|B].$$

Under the null hypothesis, this is equal to the fraction of people with actual consumption less than E, given by

$$P[Y< E|I] = P[O]\,P[Y<E|O] + (1-P[O])\,P[Y < E|B].$$

Sample II (excluding full non-respondents) is drawn from the same population. It gives a direct estimate of $P[Y <_r E|II]$, the probability that someone who gets an initial bracket question with entry point E *reports* that consumption is lower than E. Under the null, $P[Y<E|II] = P[Y<_r E|II]$. Since I and II are from the same population, $P[Y<E|II] = P[Y<E|I]$ (= $P[Y<E]$, the fraction in the population with consumption below E). Under the alternative, both $P[Y <_r E|I]$ and $P[Y <_r E|II]$ will be different from $P[Y<E]$, and the sign of the difference depends on the nature of the alternative and the position of E in the distribution of Y. It seems likely that $P[Y<_r E|I]$ will then be closer to $P[Y<E]$ than $P[Y<_r E|II]$, since the bias is attenuated by the open-ended responses (which are assumed to be always correct). Thus comparing estimates of $P[Y<_r E|I]$ and $P[Y<_r E|II]$ will give us an idea of the sign of the bias.

The easiest way to implement the joint test for the three entry points E=500, 1000 and 2000 is to formulate the above as a model that can be estimated by maximum likelihood and perform a likelihood ratio test. To account for bunching at the entry points in the open answers, which is evident in Figure 1, we explicitly estimate mass point probabilities for the outcomes 500, 1000 and 2000. The complete unrestricted model has thirteen parameters: $P[Y<500|O]$, $P[Y=500|O]$, $P[500<Y<1000|O]$, …. $P[Y=2000|O]$, $P[Y<_r 500|B]$, $P[Y<_r 1000|B]$, $P[Y<_r 2000|B]$, $P[Y<_r 500|II]$, $P[Y<_r 1000|II]$ and $P[Y<_r 2000|II]$. The three restrictions to be tested are $P[Y<_r E|II]=P[O]P[Y<E|O]+(1-P[O])P[Y<_r E|B]$, E=500,1000,2000, where $P[Y<1000|O] = P[Y<500|O] + P[Y=500|O] + P[500<Y<1000|O]$, and $P[Y<2000|O]$ is defined similarly.[4] The unrestricted ML

[4] We cannot use $P[Y<5000]$ since there is no group of follow-up bracket respondents with entry point \$5000, making it impossible to estimate $P[Y<_r 5000|B]$.

estimates are given by the corresponding sub-sample fractions. The restricted estimates have to be determined numerically.

5. Results

Table 2 presents the unrestricted and restricted estimates using the complete samples of continuous and bracket respondents. The null hypothesis is clearly rejected. Comparing the implied estimates $P[Y<_r E|I]$ with the estimates of $P[Y<_r E|II]$ in the unrestricted model gives some idea why this is the case. The estimates of $P[Y<_r 500|I]$ and $P[Y<_r 500|II]$ are not significantly different (t-value of the difference: 0.4). The estimate of $P[Y<_r 1000|I]$ is significantly larger than the estimate of $P[Y<_r 1000|II]$ (t-value: 2.3). Since this entry point is close to the median, it seems implausible that this is due to gating. On the other hand, it could very well be due to acquiescence bias, implying that too many people give an affirmative answer to the question whether their consumption is at least $1000. Similarly, the finding that $P[Y<_r 2000|I]$ is significantly larger than the estimate of $P[Y<_r 2000|II]$ (t-value 3.9) could also be due to acquiescence bias. This finding, however, would also be in line with gating, since it implies a higher probability of an “extreme” outcome (>2000) for bracket answers than for continuous answers.

Qualitatively, one interpretation of these results seems to be a combination of acquiescence bias and gating. The acquiescence bias makes $P[Y<_r E|II]$ smaller than $P[Y<_r E|I]$ for all E. Gating at 500 increases $P[Y<_r 500|II]$ more than $P[Y<_r 500|I]$. Thus for E=500, the two effects are opposite and almost cancel, apparently. For E=1000 there is no strong gating effect and the acquiescence bias raises $P[Y<_r E|II]$ more than $P[Y<_r E|I]$. For E=2000, gating and acquiescence bias are in the same direction, explaining why in this case, the difference between $P[Y<_r E|II]$ and $P[Y<_r E|I]$ is particularly large.

As we will show in Section 7, however, this explanation is likely to be incorrect. The effect of acquiescence will automatically be increasing with the entry point, because only those whose correct answer is “no” can make an acquiescence error. Few respondents have monthly household consumption less than $500 and should answer “no” if the entry point is $500, but many more have consumption level less than $2000 and should answer “no” if the entry point is $2000. If the acquiescence probability is the same for everyone, we can therefore expect more incorrect answers at $2000 than at $500.

Table 2: Nonparametric Models: ML estimates

	Unrestricted model		Restricted model	
	Estimate	St. error	Estimate	St. error
P[O]	0.6752	0.0102	0.6721	0.0103
$P[Y<500 \mid O]$	0.1695	0.0100	0.1621	0.0088
$P[Y=500 \mid O]$	0.0848	0.0074	0.0812	0.0071
$P[500<Y<1000 \mid O]$	0.2366	0.0113	0.2276	0.0107
$P[Y=1000 \mid O]$	0.1448	0.0094	0.1468	0.0093
$P[1000<Y<2000 \mid O]$	0.1794	0.0102	0.1821	0.0101
$P[Y=2000 \mid O]$	0.0726	0.0069	0.0813	0.0074
$P[Y<_r 500 \mid B]$	0.1967	0.0257	0.1936	0.0225
$P[Y<_r 1000 \mid B]$	0.4883	0.0342	0.4451	0.0302
$P[Y<_r 2000 \mid B]$	0.7511	0.0286	0.6949	0.0276
$P[Y<_r 500 \mid II]$	0.1696	0.0143		
$P[Y<_r 1000 \mid II]$	0.4167	0.0199		
$P[Y<_r 2000 \mid II]$	0.7010	0.0190		
Implied estimates:				
$P[Y<_r 500 \mid I]$	0.1783	0.0186	0.1724	0.0168
$P[Y<_r 1000 \mid I]$	0.4901	0.0253	0.4624	0.0231
$P[Y<_r 2000 \mid I]$	0.7943	0.0142	0.7654	0.0135
Log likelihood	-5456.57		-5469.49	

A potential caveat of our analysis might be the way in which we treat the focal points in the open-ended answers, particularly since the entry points are also focal points. In the approach discussed above, we have taken the answers literally and have used a mixed distribution for the open-ended answers. For the bracket answers, we have taken “E or more” literally also, including those with consumption exactly equal to E. An alternative would be to interpret the open-ended focal point reports as rounded values, assuming that the underlying distribution of true consumption values is continuous. This would mean that we would have to add about 0.5P[Y=E|O] to P[Y<E|O] before carrying out the test. It would lead to even higher values of P[Y$<_r$ E|I] and would increase the deviations between P[Y$<_r$ E|I] and P[Y$<_r$ E|II], thus further increasing the value of the test statistic. This would not change the qualitative conclusion of the test, but it would make the case for an acquiescence bias even stronger.

According to Hurd (1999), acquiescence bias and anchoring become more important if respondents are more uncertain about the actual amounts. Hurd (1999) analyzes experimental data on asset holdings, and finds that the answers of the financial respondent – the person in the household who is more knowledgeable in financial matters and answers most of the financial questions – are hardly subject to acquiescence bias, while for others, acquiescence bias is quite prominent. We can check whether this also applies to consumption by considering financial respondents (almost 75% of the sample) and other respondents separately.

Table 3. Unrestricted estimates for Financial respondents and Others

	Financial respondents		Others	
	Estimate	St. Error	Estimate	St. Error
P[O]	0.6941	0.0116	0.6176	0.0214
P[Y<500|O]	0.1934	0.0119	0.0875	0.0158
P[Y=500|O]	0.0830	0.0083	0.0906	0.0160
P[500<Y<1000|O]	0.2500	0.0131	0.1906	0.0220
P[Y=1000|O]	0.1369	0.0104	0.1719	0.0211
P[1000<Y<2000|O]	0.1779	0.0116	0.1844	0.0217
P[Y=2000|O]	0.0602	0.0072	0.1156	0.0179
P[Y$<_r$500|B]	0.2485	0.0336	0.0811	0.0317
P[Y$<_r$1000|B]	0.5267	0.0408	0.3968	0.0616
P[Y$<_r$2000|B]	0.7857	0.0312	0.6557	0.0608
P[Y$<_r$500|II]	0.2070	0.0179	0.0618	0.0180
P[Y$<_r$1000|II]	0.4773	0.0232	0.2282	0.0344
P[Y$<_r$20000|II]	0.7346	0.0211	0.6000	0.0407
Implied estimates:				
P[Y$<_r$500|I]	0.2103	0.0248	0.0850	0.0219
P[Y$<_r$1000|I]	0.5265	0.0308	0.3795	0.0422
P[Y$<_r$2000|I]	0.8243	0.0166	0.6985	0.0269
Log likelihood				
Unrestricted	-4115.58		-1274.89	
Restricted	-4123.86		-1282.15	

The estimates of the unrestricted model and the restricted likelihood for both groups are presented in Table 3. The underlying distributions are obviously different: all nonfinancial respondents are married and therefore are in high consumption households, on average, whereas some financial respondents are single and therefore more often in low consumption households. The probability of a focal point answer is smaller for financial respondents than for others, already suggesting that financial respondents are less uncertain about their consumption levels than others. However, the null hypothesis of neither gating nor acquiescence bias is rejected for both financial respondents and others. The deviations between estimates of $P[Y <_r E|I]$ and $P[Y <_r E|II]$ are similar for the smallest and largest entry points E=\$500 and E=\$2000. The t-statistics on these differences are 0.1 and 0.8 at E=\$500, and 3.3 and 2.0 at E=\$2000, for financial respondents and others, respectively. Only for the intermediate entry point E=\$1000, it seems that the (acquiescence) bias is smaller for financial respondents than for others. Here the t-test does not reject equality for financial respondents (t-statistic 1.3), while it does reject equality for others (t-statistic 2.8).

6. Relaxing the Assumptions about Full Non-response

Until now we have assumed that there is no selective complete non-response, i.e., the respondents in sample I who neither give an open-ended nor a bracket answer have the same consumption distribution as the respondents in sample II who do not give any bracket answers. In this section we will investigate whether the result obtained above could be due to that assumption instead of to acquiescence and/or gating.

There are several types of non-response. First, the experimental module is given to the respondents at the end of the interview, and before it starts, respondents are explicitly asked whether they are willing to cooperate. About 19.2% of all respondents are not willing to cooperate and thus do not answer any of the questions in the experimental module. Since this selection takes place before respondents know which experimental module they are assigned to (and do not know whether they will be in sample I or sample II), it seems quite reasonable that this selection is not systematically different for the two samples. It may mean that both samples are no longer representative of the population we would ultimately be interested in but will not invalidate the test discussed in the previous sections.

Of the remaining respondents, 0.87% stop somewhere during the experimental module, before coming to the consumption questions (0.83% in sample I, 0.94% in sample II). The samples are split into seven sub-samples (three in sample I, four in sample II) that do not get exactly the same questions, but there seems to be no reason for any systematic difference between samples I and II. It thus seems reasonable to treat these non-respondents to the consumption questions in the same way as those who did not want to participate in the modules. Since they stop before they know which consumption questions they will get, we will assume that their consumption distributions are not systematically different for the two samples.

We can therefore condition on not opting out before the consumption questions start and focus on the third type of full non-response. In sample I, 2.37% of all respondents answer “don’t know” or “refuse” to the open-ended consumption question and subsequently also answer “don’t know” or “refuse” to the first follow-up bracket question. In sample II, 2.74% answer “don’t know” or “refuse” to the first bracket question. Since the survey designs are systematically different here – an open-ended question first versus forced brackets – selection into full non-response might be different for the two samples. We will now check whether such differences could explain the test results in the previous section.

Without further assumptions, even under the null hypothesis of correct answers only, the probability $P[Y<E]$ is no longer identified if full non-respondents are included in the population. The reason is that nothing is known about the values of Y for the full non-respondents. We can, however, apply Manski (1989, 1995) to obtain upper and lower bounds for the probabilities. In that way we can, under the null hypothesis, construct intervals for $P[Y<E]$ on the basis of I and II. Under the null hypothesis these intervals should both contain the common value $P[Y<E]$. To take account of sampling uncertainty, we will construct confidence bands for the lower and upper bounds of the intervals and thus construct intervals that contain the Manski bounds with at least 95% confidence. The test will then be based upon comparing the intervals based upon samples I and II.

To make this more precise some notation needs to be introduced. For sample I, there are three types of response behavior, open-ended (O), bracket (B) or none at all (N), with probabilities $P[O|I]$, $P[B|I]$ and $P[N|I]=1-P[O|I]-P[B|I]$. The probabilities that were used in the previous section were conditional on giving some response, i.e., on not N. Under the null hypothesis that all bracket answers are correct, Table 2 thus gives the estimates (with standard errors in parentheses) 0.1783 (0.0186), 0.4901 (0.0253) and 0.7943 (0.0142) of $P[Y<E|\text{not } N, I]$, for E=500, 1000 and 2000, respectively. The probabilities that are not conditional upon some response are $P[Y<E|I] = (1-P[N|I])P[Y<E|\text{not } N,I] + P[N|I]P[Y<E|N,I]$. Without further assumptions, nothing is known about $P[Y<E|N,I]$, except of course that it is between 0 and 1. Using this gives the Manski bounds on $P[Y<E|I]$:

$$(1-P[N|I])P[Y<E|\text{not } N,I] \leq P[Y<E|I] \leq (1-P[N|I])P[Y<E|\text{not } N,I] +P[N|I]$$

Similarly, the following Manski bounds can be derived for sample II:

$$(1-P[N|II])P[Y<E|\text{not } N,II] \leq P[Y<E|II] \leq (1-P[N|II])P[Y<E|\text{not } N,II] +P[N|II]$$

Applying this to the entry points E=500, 1000 and 2000 gives the following point estimates of the Manski bounds:

	Sample I	sample II
E=500	[0.1741; 0.1978]	[0.1650; 0.1924]
E=1000	[0.4785; 0.5022]	[0.4053; 0.4327]
E=2000	[0.7755; 0.7992]	[0.6818; 0.7092]

First, consider E=500. The two intervals overlap. If all full non-respondents in sample I have consumption below \$500 and all full non-respondents in sample II have consumption above \$500, the sample I estimate is 0.1978 and the sample II estimate is 0.1650. If both assumptions are reversed, the estimates are 0.1741 and 0.1924. Many intermediate cases can be thought of for which the two estimates are identical. We already concluded in Section 5 that the difference at $500 can also be explained from sampling uncertainty – under the assumption that selection into full non-response was the same, the point estimates are not significantly different. Thus at E=$500, the current exercise does not add much.

Now consider E=$1000. Here the two intervals do not overlap. Even if all full non-respondents in sample II have low consumption and all full non-respondents in sample I have high consumption, the estimate for sample II (0.4327) remains below that for sample II (0.4785). If there were no sampling uncertainty, this would mean that the null hypothesis should be rejected (at the zero significance level). Of course there is sampling uncertainty, and the question thus becomes whether the upper bound for sample II is significantly lower than the lower bound for sample I. The standard error on the difference between these two can be derived along the same lines as in the previous section and appears to be about 0.033, giving a t-value of about 1.4. In other words, if we are willing to assume that full non-respondents in the two samples are completely opposite groups (low consumption versus high consumption), and only consider entry point E=$1000, then we cannot reject the null hypothesis that all reports are correct at the 5% (or even 10%) two-sided significance level.

Similarly, considering the entry point E=$2000, the question becomes whether the upper bound for sample II (0.7092) is significantly different from the lower bound for sample I (0.7755). In this case, the standard deviation on the difference is about 0.025 and the t-value on the difference becomes 2.6. That is, the null hypothesis of correct reports only can be rejected at the 5% level, even if we allow for opposite selection into full non-response in samples I and II. Maintaining the assumption that the open-ended answers are correct, this provides evidence of anchoring or acquiescence bias in the bracket answers. The evidence is quite strong in the sense that any type of selection into full non-response is allowed for, even the most extreme – and obviously not very plausible – case that all non-respondents in one group are from low consumption (i.e., lower than $2000 per month) families while all non-respondents in the other group are from high consumption families. Imposing more assumptions on response behavior (such as monotonicity, see Manski 1989) will make the test more powerful and reinforce the result.

We conclude that the non-respondents cannot explain the variation by entry point, and that a role remains for gating and acquiescence bias.

7. Fixed Probability Gating and Acquiescence

To investigate more formally whether gating or acquiescence bias (or both) can indeed explain why the null hypothesis is rejected, the model can be extended with a simple form of gating and acquiescence bias, using fixed probabilities of acquiescence and gating. We go back to the assumptions in sections 4 and 5 about full non-response and consider the subpopulation of respondents who provide either open-ended or bracket information. First, we assume that there is a fixed fraction of the population that will automatically answer a bracket question with "yes," irrespective of the true amount and the entry point. Second, we assume that in the rest of the population there is a fixed probability P(Gat) that people give the wrong answer ("yes" if $Y<E$ or "no" if $Y\geq E$).[5]

With these two additions to the nonparametric model, the probability that a random person in the population answers "yes" if the true consumption amount is less than E, is given by P(Acq) + (1-P(Acq))P(Gat) where P(Acq) is the probability the person is the acquiescent type. The probability that someone answers "no" if the true amount is at least E, is given by (1-P(Acq))P(Gat). For the symmetric case with P(Acq)=0, this model is essentially the same as what Copas (1988, p.234) calls "a simple model for resistant fitting." The general case is similar to the misclassification models used by, for example, Ekholm and Palmgren (1982), Lee and Porter (1984) or Hausman et al. (1998).[6]

Adding these features to the unrestricted model gives a model with one parameter less than the unrestricted model, thus imposing one restriction in the general model. This restriction is not rejected by a likelihood ratio test: the log likelihood of the new model is –5456.91, close to the log likelihood of the unrestricted model in Table 2. The results of

[5] With just one observation we cannot distinguish this model of heterogeneity from one in which everyone in the population has a combination of gating and acquiescence errors such that P("yes"|$Y<E$)>P("no"|$Y>E$)>0. For convenience of exposition we adopt the heterogeneity model.

[6] The current model is less general than the models in the literature in the sense that yea-saying works in only one direction and makes the probability or reporting a "yes" where "no" would be correct larger than the probability of reporting "no" where "yes" would be correct, but this inequality restriction appears not to be binding.

this intermediate model are presented in Table 4. The estimates of P(Gat) and P(Acq) suggest that the latter is much more important than the former: the estimate of the gating probability is 0.018 with standard error 0.039, while the acquiescence probability is 0.170 with standard error 0.037.

To explain why the fixed acquiescence probability works well in spite of the seemingly different biases at the three entry points (revealed by Table 1), consider some of the results of the model with fixed acquiescence probability but without gating. In this model, the estimate of P(Acq) is 0.176 (with standard error 0.035) . The estimated fraction of people with (true) consumption less than $500 is 0.196. The acquiescence bias in the sample of forced bracket respondents is therefore estimated to be 0.196*0.176=0.034. In sample I, it is about one third as large (since one third goes to brackets). This would therefore lead to a difference of about 2.3 % points between samples I and II. At entry point $1000, the estimated acquiescence bias in sample II would be 0.519*0.176=0.091, leading to a 6.0%-points difference between samples I and II. At entry point $2000, the difference would be about two thirds of 0.847*0.176=0.149, i.e., about 10.0%-points. These differences are so close to the observed differences in the data that the null hypothesis that this simple model is correctly specified cannot be rejected.

The other panels of Table 4 present the models with either fixed probability gating or fixed probability acquiescence. In the model without fixed probability acquiescence, the gating probability is much larger than if acquiescence is also allowed for. This model is misspecified, however. In the model without gating, the acquiescence probability is close to its estimate if gating is also allowed for, as are all the other parameters. All this suggests that acquiescence alone might explain the anomalies in the data. In any case, it can explain the result of the nonparametric tests.

Figure 2 compares the implications for the distribution of consumption of the three models in Table 4 and the restricted “benchmark” model without gating or acquiescence in the right hand panel of Table 2. Compared to the benchmark model, the model allowing for gating only implies a smaller dispersion in the distribution of consumption. This is because part of the “extreme” bracket answers (“no” at E=$500; “yes” at E=$2000) are explained by gating rather than true low or high consumption values. While the benchmark model estimates just reflect the observed data, the gating model recognizes that part of the large dispersion in the observed data is due to gating, and produces a less dispersed picture of the true consumption values. The differences are small, however.

Larger differences arise when acquiescence is allowed for. In this case, allowing for gating in addition makes no difference – the two curves are virtually the same. But both imply lower true values of consumption than the models without acquiescence. The explanation is clear: in the models allowing for yea-saying, some of the “yes” answers to the bracket questions are explained as yea-saying instead of necessarily reflecting high consumption values. Thus the model with yea-saying can explain the large number of “yes” answers without high underlying consumption values.

The difference can also be explained in terms of selection into open-ended answers versus bracket response (full non-response is ignored). Although we have not made the relation between consumption level and willingness to respond to the open-ended answer explicit yet, the model certainly allows us to do so. Using Bayes rule, the probability of giving an open-ended answer (in the survey design of sample I) given that consumption is lower than the entry point level E can be written as:

$$P[O|Y<E] = P[O]P[Y<E|O]/(P[O]P[Y<E|O] + \{1-P[O]\}P[Y<E|B])$$

All the probabilities on the right hand side can be computed directly from the estimates in Table 2 or Table 4. For the benchmark (restricted) model in Table 2 we get, for example, an estimate of 0.684 for $P[O|Y<1000]$. This is somewhat higher than the estimate of P[O] which is 0.672. Thus the benchmark model would imply a small (and insignificant) selection effect, which is positive in the sense that the higher consumption values have higher probability of self-selection in the bracket
questions.

According to the most general model in Table 4, however, we get an estimate of 0.638 for $P[O|Y<1000]$, which is much lower than the estimate of 0.675 for P[O]. Thus the model allowing for acquiescence implies negative selection: respondents with low consumption more often become bracket respondents. The many “yes” responses in the (follow-up) bracket answers do not always reflect high consumption but can partly be explained by yea-saying. The distinction between the two explanations is identified because of the group of forced bracket respondents, which is not subject to selection effects.

Table 4. Models with Fixed Acquiescence and/or Gating Probabilities

	Gating and Acquiescence		Gating only		Acquiescence only	
	Est.	S.e.	Est.	S.e.	Est.	S.e.
P[O]	0.6753*	0.0102	0.6739*	0.0103	0.6752*	0.0102
P[Y<500|O]	0.1695*	0.0100	0.1578*	0.0091	0.1709*	0.0096
P[Y=500|O]	0.0843*	0.0073	0.0823*	0.0072	0.0841*	0.0073
P[500<Y<1000|O]	0.2352*	0.0111	0.2298*	0.0109	0.2348*	0.0111
P[Y=1000|O]	0.1456*	0.0093	0.1475*	0.0094	0.1452*	0.0093
P[1000<Y<2000|O]	0.1804*	0.0102	0.1828*	0.0103	0.1800*	0.0101
P[Y=2000|O]	0.0727*	0.0069	0.0785*	0.0073	0.0727*	0.0069
P[Y<500|B]	0.2308*	0.0511	0.1262*	0.0435	0.2490*	0.0310
P[Y<1000|B]	0.5771*	0.0511	0.4356*	0.0351	0.5798*	0.0494
P[Y<2000|B]	0.9253*	0.0667	0.7441*	0.0424	0.9141*	0.0599
P[Gating]	0.0178	0.0391	0.0659*	0.0313	0	
P[Acquiescence]	0.1703*	0.0371	0		0.1760*	0.0349
log likelihood	-5456.91		-5467.29		-5457.01	
Implied estimates:						
P[Y<500]	0.1894	0.0344	0.1475	0.0295	0.1962	0.0222
P[Y<1000]	0.5176	0.0356	0.4587	0.0261	0.5190	0.0346
P[Y<2000]	0.8508	0.0142	0.7819	0.0138	0.8472	0.0141

Prior findings in the literature suggest that acquiescence bias should vary with the amount of uncertainty a subject has about the quantities being queried. Table 5, shows the results of estimating the model of Table 4 but with acquiescence probabilities only. The estimations are over a number of sub-samples. For example, we find an acquiescence probability of 0.147 among financial respondents only. In almost all cases the estimated acquiescence probability is substantial and its 95% confidence interval does not contain the value zero. There is some variation across subgroups, but this variation is not very large and usually insignificant, because of the smaller numbers of observations (and perhaps also the non-parametric nature of the approach). Still, most of the findings are in line with the notion that people who are more uncertain have a larger probability of acquiescence.

Table 5. Estimated Probabilities of Acquiescence
(Model with fixed acquiescence probability, without gating)

subsample	# obs.	P[Acquiescence]	
		Estimate	St. error
Financial respondent	3548	0.147	0.036
Not financial respondent	1211	0.311	0.095
Borne after 1919	2088	0.190	0.059
Borne before 1920	2671	0.171	0.042
Woman	3111	0.187	0.041
Man	1648	0.182	0.064
Years education >12	1445	0.138	0.081
Years education ≤12	3314	0.170	0.036
High household income[a]	2380	0.173	0.061
Low household income	2379	0.134	0.035
High household wealth[a]	2380	0.222	0.061
Low household wealth	2379	0.133	0.037
Self-rated memory high[b]	1494	0.072	0.069
Self-rated memory low	3265	0.226	0.042
Memory test high[c]	1943	0.120	0.059
Memory test low	3016	0.193	0.042
All	4759	0.176	0.035

Notes:
[a]High and low household income and wealth: above or below median; missing income and wealth values imputed (RAND version C of HRS/AHEAD); wealth is total wealth (including housing and IRAs).
[b]Self-rated memory: Answer to "How would you rate your memory at the present time?" Possible answers: excellent, very good, good, fair or poor. High is defined as excellent or very good.
[c]Memory test: number of words immediately recalled from a list of 40 words that are read to the respondent. High is defined as at least 6 (the maximum was 10).

For example, financial respondents and respondents with high education level are less subject to yea-saying than others. Using either a self-reported indicator or a test-based indicator of memory quality, we find that people with better memory are less subject to yea-saying. There is hardly any relation between age or gender and the tendency of yea-saying. The income and wealth patterns seem somewhat surprising, with the higher income and higher wealth respondents more subject to yea-saying. As said before, however, significance levels are low. Only the difference between those with low and high self-rated memory skills is close to significant at the conventional two-sided 5% level (t-value 1.91).[7]

7. Conclusions

In this study, we have investigated the importance of gating and acquiescence bias at the entry point of an unfolding bracket design. Experimental data on consumption where respondents are randomly either first given an open-ended question on household consumption or immediately directed to bracket questions were used to test whether these phenomena are present without making any model assumptions. The main finding is that some bias is present even at the entry point. Further analysis – making additional assumptions on the nature of acquiescence and gating – suggests that acquiescence bias is the main problem, while gating is less important. For the consumption question at hand, ignoring acquiescence leads to misleading conclusions on the selective nature of item non-response and on the inference on the distribution of consumption when selective item non-response is taken into account. It would lead to overestimation of consumption levels and underestimation of poverty rates.

Reducing acquiescence bias in data collection is easy: the bracket questions can be formulated in a neutral way, asking people to choose a category rather than answering "yes" or "no". This has already been implemented in the most recent waves of, for example, the Health and Retirement Study (HRS) and AHEAD. The results of this study suggest that, with these improved questions, it seems at least safe to use the first bracket question in an unfolding bracket design at face value. Given the evidence in the literature that anchoring is a problem also in absence of acquiescence bias, this at the same time means that anchoring at follow up bracket questions must be taken seriously. The fixed point model for anchoring – or another model with Bayesian learning – in which answers to the later bracket questions are affected by the entry point may then be a useful tool.

[7] Since the estimates are based upon two independent samples, the standard deviation of the difference is the square root of the sum of the squared standard deviations of the two estimates.

References

Boyle, K.J., H.F. MacDonald, H. Cheng and D.W. McCollum (1998), Bid design and yea saying in single-bounded, dichotomous-choice questions, *Land Economics*, 74(1), 49-61.

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.

Copas, J.B. (1988), Binary regression models for contaminated data, *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(2), 225-253.

Ekholm, A. and J. Palmgren (1982), A model for a binary response with misclassification, *GLIM-82: Proceedings of the International Conference on Generalized Linear Models*, R. Gilchrist (ed.), Springer, Heidelberg, 128-143.

Frykblom, P. and J. Shogren (2000), An experimental testing of anchoring effects in discrete choice questions, *Environmental and Resource Economics*, 16, 329-341.

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.

Hausman, J.A., J. Abrevaya and F.M. Scott-Morton (1998), Misclassification of the dependent variable in a discrete response setting, *Journal of Econometrics*, 87(2), 239-269.

Herriges, J. and J. Shogren (1996), Starting point bias in dichotomous choice valuation with follow-up questioning, *Journal of Environmental Economics and Management*, 30, 112-131.

Hurd, M. (1997), A model of anchoring, typescript.

Hurd, M. (1999), Anchoring and acquiescence bias in measuring assets in household surveys, *Journal of Risk and Uncertainty*, 19(1-3), 111-136.

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.

Hurd, M., A. Kapteyn and J. Zissimopoulos (2001), Recovering the distribution from bracketed responses, paper presented at TMR Savings and Pensions workshop in Evian, France, August 21-23.

Jacowitz, K. and D. Kahneman (1995), Measures of anchoring in estimation tasks, *Personality and Social Psychology Bulletin*, 21, 1161-1166.

Juster, T., and J.P. Smith (1997), Improving the quality of Economic data: Lessons from the HRS and AHEAD, *Journal of the American Statistical Association*, 92, 1268-1278.

Lee, L.F. and R.H. Porter (1984), Switching regression models with imperfect sample information with an application on cartel stability," *Econometrica*, 52, 391-418.

Little, R.J.A., and D.B. Rubin (2002), *Statistical analysis with missing data*, Wiley, New York.

Manski, C.F., (1989), Anatomy of the selection problem, *Journal of Human resources*, 24, 343-360.

Manski, C.F., (1995), *Identification problems in the social sciences*, Harvard University Press.

O'Connor, R., M. Johansson and P.-O. Johansson (1999), Stated preferences, real behaviour and anchoring: some empirical evidence, *Environmental and Resource Economics*, 13, 235-248.

Schuman, H. and S. Presser (1981), *Questions and answers in attitude surveys*, Academic Press, New York.

Tversky, A. and D. Kahneman (1974), Judgment under uncertainty: heuristics and biases, *Science*, 185(4157), 1124-1131.

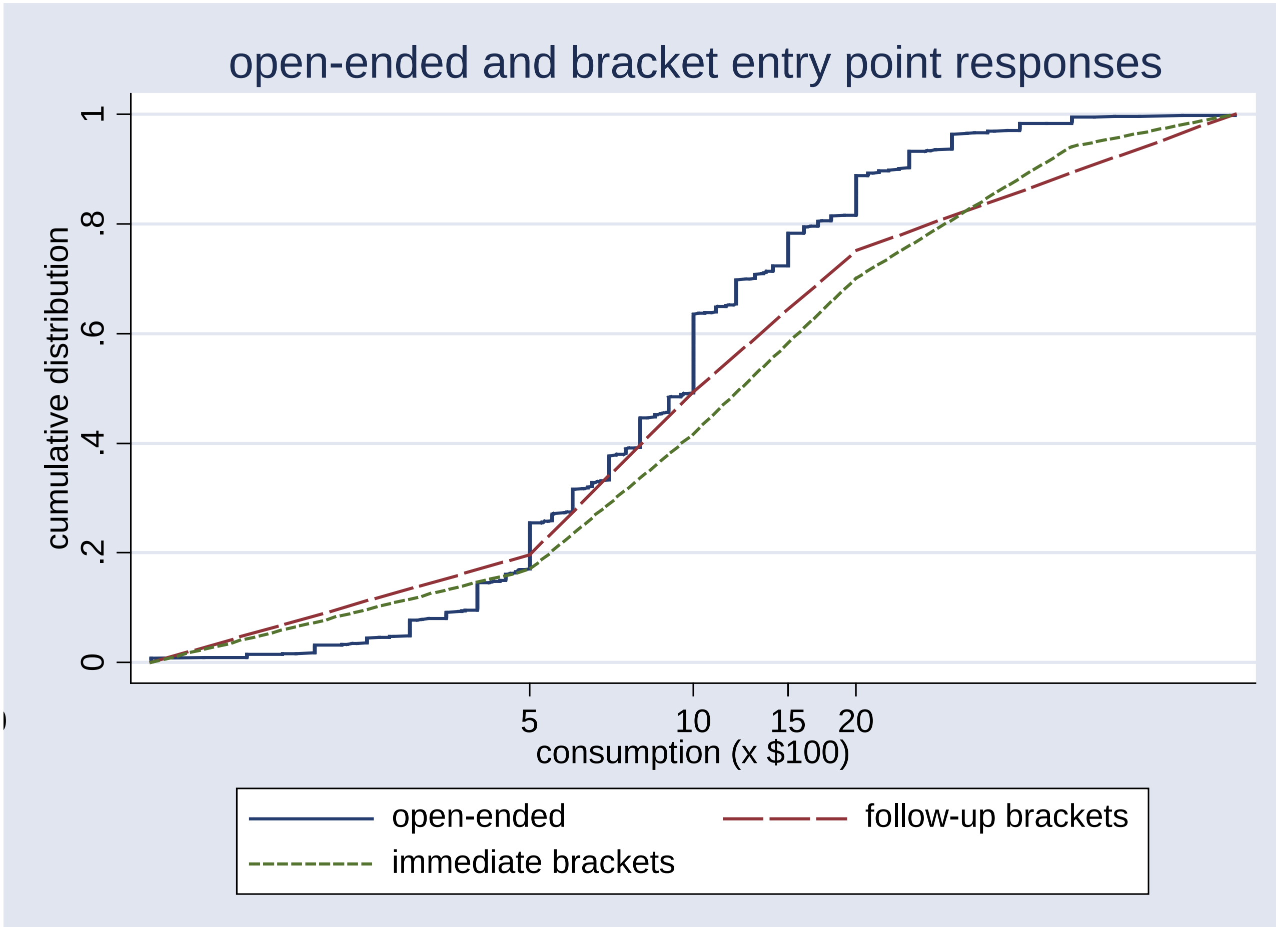

Figure 1. Cumulative Distributions of Monthly Consumption (on logarithmic scale)
Open-ended responses (sample I);
Follow-up bracket responses using entry point question only (sample I);
Immediate bracket responses using entry point question only (sample II)

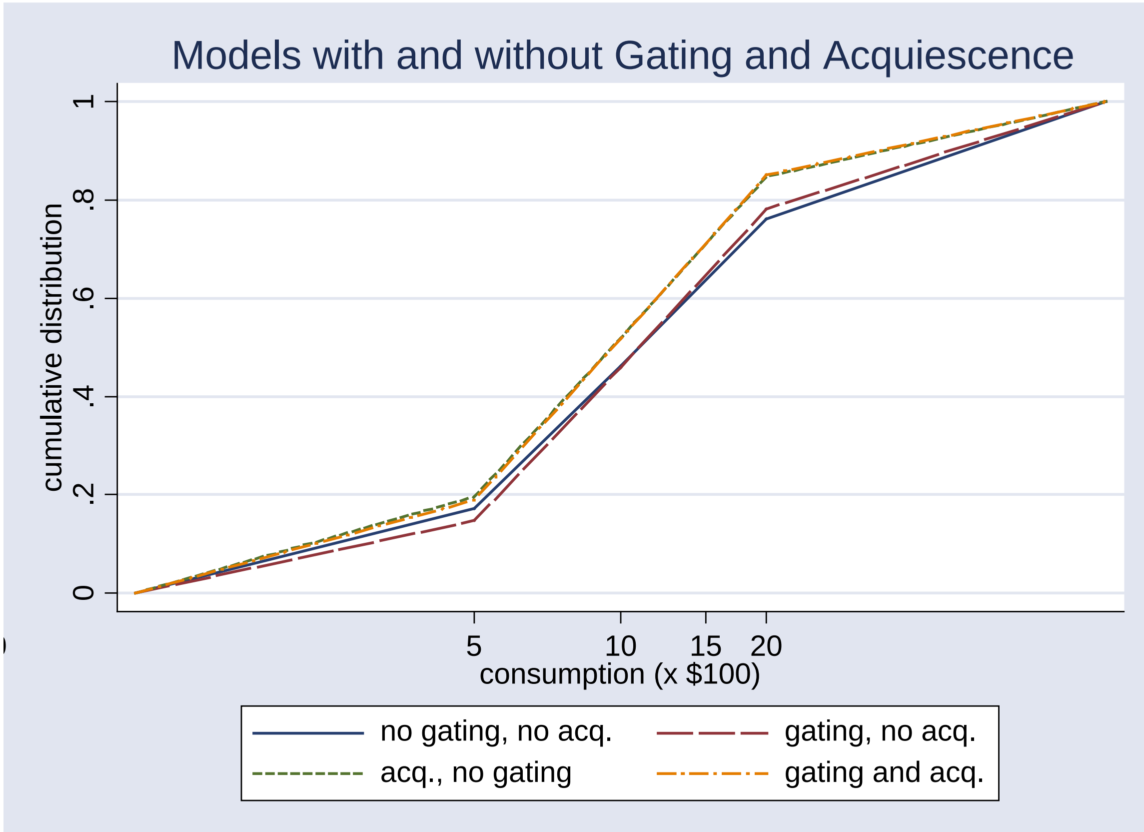

Figure 2. Distribution of Consumption According to Models with and without Fixed Probability Gating and Acquiescence

Appendix A: Details on the Consumption Data and on Complete Non-Response

Table A1 explains how many respondents were willing and not willing to answer the experimental modules. About 19.2% of all respondents did not participate.

Table A1: Willingness to cooperate in experimental modules

Module and entry point	Willing	Not willing
1.Open-ended; E=2000	746	194
2.Open-ended; E=500	695	199
3.Forced bracket; E=1000	636	174
4.Forced bracket; E=5000	801	141
5.Forced bracket; E=500	720	171
6.Forced bracket; E=2000	605	139
7.Open-ended; E=1000	725	155

Table A2 gives an overview of complete item non-response in the consumption question. It concerns respondents who were willing to answer the questions in the experimental module. Those who were not willing are excluded; there cannot be any difference in selection between the groups there, since these respondents refused to do the experimental module before they knew the questions.

Table A2: Response and non-response counts and percent distribution; All respondents

Module and entry point	Type of Response				Total
	Given	DK	RF	Missing	
1. Open-ended E=2000	725	11	5	5	746
	97.18	1.47	0.67	0.67	100.00
2. Open-ended E=500	674	13	6	2	695
	96.98	1.87	0.86	0.29	100.00
3. Forced Bracket E=1000	612	19		5	636
	96.23	2.99		0.79	100.00
4. Forced Bracket E=5000	778	17		6	801
	97.13	2.12		0.75	100.00
5. Forced Bracket E=500	690	19		11	720
	95.83	2.64		1.53	100.00
6. Forced Bracket E=2000	582	20		3	605
	96.20	3.31		0.50	100.00
7. Open-ended E=1000	698	12	4	11	725
	96.28	1.66	0.55	1.52	100.00
	4759	111	15	43	4928
Total	96.57	2.25	0.30	0.87	100.00

Groups 3,4,5 and 6 immediately go to the bracket questions. Those who do not answer the first bracket question are coded as 98 (DK/RF); no distinction between DK (don't know) and RF (refuse) is made. There are also some people who have not given any answer to the first (or other) bracket questions; their answer to the bracket question is coded as missing, coded as –1 in Table A2. They dropped out during the module, before coming to the consumption questions.

Groups 1,2 and 7 first get an open-ended question. Those who answer DK and then answer DK/RF at the first bracket question are given misc_b1=8; those who answer RF to the continuous question and then answer DK/RF to the first bracket question get misc_b1=9. Some people have a missing value for this question (as well as the first bracket question); they get misc_b1=-1. (This might apply to people in nursing homes for whom the consumption question is skipped.) All those for whom either continuous or bracket information is available get misc_b1=0.

The results show that complete item non-response varies between 2.8% in group 1 and 3.8% in group 6. The differences between the groups are small, and there seem to be no systematic differences between the continuous groups (1,2,7) and the immediate bracket respondents (3,4,5,6).

In Tables A3 and A4, the same thing is done for financial respondents and non-financial respondents separately. Complete non-response is somewhat more common for non-financial respondents, but the differences are small. Neither for financial respondents nor for non-financial respondents, it seems possible to detect any systematic pattern or a relation between complete non-response and whether a continuous question is asked first.

Table A3: Response and non-response counts and percent distribution; Financial respondents

Module and entry point	Type of Response				Total
	Given	DK	RF	Missing	
1. Open-ended E=2000	543 96.96	8 1.43	5 0.89	4 0.71	560 100.00
2. Open-ended E=500	499 97.46	6 1.17	5 0.98	2 0.39	512 100.00
3. Forced Bracket E=1000	463 96.66	19 2.99		5 1.04	479 100.00
4. Forced Bracket E=5000	567 97.26	17 2.12		6 1.03	583 100.00
5. Forced Bracket E=500	512 96.06	19 2.64		11 2.06	533 100.00
6. Forced Bracket E=2000	437 97.76	20 3.31		2 0.45	447 100.00
7. Open-ended E=1000	537 96.24	7 1.25	4 0.72	10 1.79	558 100.00
Total	3558 96.90	60 1.63	14 0.38	40 1.09	3672 100.00

Table A4: Response and non-response counts and percent distribution; Non-financial respondents

Module and entry point	Type of Response: Given	DK	RF	Missing	Total
1. Open-ended E=2000	182 97.85	3 1.61	0 0.00	1 0.54	186 100.00
2. Open-ended E=500	175 95.63	7 3.83	1 0.55	0 0.00	183 100.00
3. Forced Bracket E=1000	149 94.90	19 2.99		0 0.00	157 100.00
4. Forced Bracket E=5000	211 96.79	17 2.12		0 0.00	218 100.00
5. Forced Bracket E=500	178 95.19	19 2.64		0 0.00	187 100.00
6. Forced Bracket E=2000	145 91.77	20 3.31		1 0.63	158 100.00
7. Open-ended E=1000	161 96.41	5 2.99	0.00	1 0.60	167 100.00
Total	1201 95.62	51 4.06	1 0.08	3 0.24	1256 100.00