CONSUMPTION AND SAVINGS BALANCES OF THE ELDERLY: EXPERIMENTAL EVIDENCE ON SURVEY RESPONSE BIAS

Michael Hurd, Daniel McFadden, Harish Chand, Li Gan, Angela Merrill, Michael Roberts[1]

N.B.E.R. Conference of the Economics of Aging

April 24-27, 1997

ABSTRACT: The unfolding bracket method for eliciting quantitative information in economic surveys is effective in reducing item non-response and outliers, but is vulnerable to bias induced by anchoring. To test for anchoring effects, this study introduced an experimental module in the AHEAD (Asset and Health Dynamics of the Oldest Old) panel in which the unfolding bracket (or gate) sequence varied by treatment. We find that there are strong anchoring effects in the two household variables studied, savings account balances and monthly consumption. We construct a model of anchoring that is successful in capturing much of the effects of anchoring. Thus, it appears that experimental variation in the unfolding bracket design combined with a suitable model of anchoring can be used to predict and undo the biases introduced by anchoring. These conclusions have important implications for the design of quantitative questions in current economic panels such as AHEAD and the Health and Retirement Survey (HRS), and for other applications that use gate designs, such as elicitation of willingness-to-pay by contingent valuation.

---

[1] The first author is at S.U.N.Y. Stony Brook and the RAND Corporation, and the remaining authors are at the University of California, Berkeley. The first two authors are associates of the National Bureau of Economic Research. Correspondence should be directed to Daniel McFadden, Department of Economics, University of California, Berkeley CA 94707-3880, or mcfadden@econ.berkeley.edu. A postscript version of this paper is posted at http://elsa.berkeley.edu/~mcfadden. This research was supported by a grant from the Institute on Aging of the National Institute of Health to the National Bureau of Economic Research. We are indebted to Danny Kahneman, whose remarks on anchoring motivated this research.

CONSUMPTION AND SAVINGS BALANCES OF THE ELDERLY: EXPERIMENTAL EVIDENCE ON SURVEY RESPONSE BIAS

Michael Hurd, Daniel McFadden, Harish Chand, Li Gan, Angela Merrill, Michael Roberts

"Collecting bracket responses without varying the anchors is criminally negligent."
Danny Kahneman, 1993

I. Psychometric Biases in Economic Survey Data

A. The Need for Accurate Data. A prerequisite for understanding the economic behavior of the elderly, and the impacts of public policy on their health and well-being, is accurate data on key economic variables such as income, consumption, and assets, as well as expectations regarding future economic and demographic events such as major health costs, disabilities, and death. Standard practice is to elicit such information in economic surveys, relying on respondents' statements regarding the variables in question.

Economists are generally aware that stated responses are noisy. Item non-response is a common problem, and carefully done surveys are designed to minimize it. Well-designed analyses of economic survey data are careful about detecting implausible outliers, imputing missing values, and correcting for selection caused by dropping missing observations. Circumstances are recognized that tend to produce systematic biases in response, such as telescoping in recall of past events that arises from the psychophysical perception of time intervals, or overstatement of charitable contributions that arises from the incentive to project a positive self-image. Nevertheless, economic studies are often too sanguine about the reliability of subject's statements regarding objective economic data.

B. Stated Versus Revealed Economic Data. For many economic variables, it is possible in principle to obtain the accounting or administrative records necessary to verify stated responses. For example, subjects may be asked to consult or provide copies of utility bills, bank statements, or income tax records, or to give permission for linking to medicare or social security records. In practice, this is rarely done, due to cost, the difficulty of obtaining compliance from the subjects, and privacy and disclosure issues surrounding government administrative records.

In cases where direct comparisons of stated and revealed economic data are available, the results are sobering. For example, Poterba and Summers (1986) find that misstatements regarding employment status lead to underestimates by a factor of two of the duration of unemployment. Cowing *et al* (1982) found in an EIA panel of houses that five percent of the basements reported in one wave of the survey disappeared in the next wave. There may be parallels between the "disappearing basement" problem and the "disappearing asset" problem following sales of homes in economic panels such as the PSID and HRS.

Other techniques for investigating the accuracy of economic survey responses are to compare survey aggregates with national administrative aggregates, and to vary the survey by experimental design to obtain internal evidence on consistency of responses. An example of the first type is comparison of stated days of hospitalization with aggregate hospital statistics. This paper focuses on an example of the second type, in which the elicitation format for several economic questions in the Asset and Health Dynamics of the Oldest Old (AHEAD) survey was varied by design.

C. Response Errors. There are at least five reasons that a subject may fail to give accurate information on an economic variable: question ambiguity, subject concerns about confidentiality of sensitive information, incentives for strategic misrepresentation, imperfect knowledge of the facts, and psychometric context effects. In addition, errors may be introduced in the coding of responses by interviewers, and in processing the survey data.

Question Ambiguity. Questions about economic quantities may confuse subjects even if they are prepared to give accurate responses. Consider annual savings as an illustration. The question "How much did you save last year?" is not straightforward. Should accumulation of equity in durables such as real property be included? What about vehicles? Accumulated earnings and capital gains in asset accounts? Depreciation? Changes in checking account balances? Should changes in asset values be in nominal terms, or in inflation-corrected dollars? Is the year in the question from the date of the interview, or the last calendar year? In the absence of detailed instructions, which would themselves be vulnerable to misunderstanding, even subjects with precise knowledge of their economic position may find such questions difficult to answer. The result may be non-response, or a dispersion in responses resulting from different implicit assumptions on the definition of the economic variable. This paper will not address question ambiguity, but we note that it may be a significant source of error in economic surveys, and the elicitation and analysis methods used to moderate the impact of other sources of error do nothing to control these errors.

Confidentiality. Concerns of subjects about privacy and disclosure are likely to contribute to item non-response. In addition, survey organizations may be reluctant to ask sensitive questions because of the possibility of upsetting respondents and endangering the rest of the survey responses. However, the experience with AHEAD, LSOA, and other contemporary panel studies is that subjects are remarkably willing to discuss areas that have traditionally been considered difficult areas for questioning, such as health events. While it is clearly essential to establish that the survey has a useful social purpose and that confidentiality will be maintained, it seems clear that establishing rapport between the interviewer and the subject, and structuring the questionnaire so that sensitive questions are not surprising or boring, will be sufficient to eliminate confidentiality concerns as a major source of error for most topics.

Strategic Misrepresentation. Economic theory suggests that when subjects anticipate a possible connection between their response and some economic outcome in which they have an interest, they may have strategic incentives to misrepresent information. To illustrate, subjects asked about their interest in nursing home insurance may overstate their willingness-to-pay (WTP) if they believe a large response will increase the probability they will have this service as an option, without committing them to this cost. On the other hand, they may understate WTP if they believe that their actual cost would be tied to their response. In practice, most standard economic surveys have no linkage from response to subsequent economic events that would create incentives for misrepresentation. Further, there is at least fragmentary evidence that subjects are usually truthful when there are no positive incentives for misrepresentation, and even in some circumstances where there are such incentives; see Bohm (1972), Smith (1979). There are, however, some areas where there may be strong *non-pecuniary incentives* for misrepresentation. For example, subjects asked questions like "How often do you go to church?", or "How much did you contribute to charity last year?", may give biased responses in order to project a more favorable image to the interviewer and to themselves. In contingent valuation surveys, this phenomenon is sometimes called the "warm glow" motivation for overstating WTP for public goods. There are some elementary precautions in economic survey design that decouple responses from economic consequences, and eliminate obvious sources of *economic* incentives for misrepresentation. One way to control misrepresentation arising from *non-pecuniary* incentives is to present subjects with tasks that are "ethically neutral". For example, subjects may have no incentive to misrepresent trade-offs between different public goods, even when "warm glow" distorts their stated trade-off between public goods and personal private goods or money.

Imperfect Knowledge. Most economic surveys assume that subjects can readily and reliably recall household economic facts. This may be valid for regularly monitored quantities, such as checking account balance or monthly social security check, and subjects may find that being truthful minimizes response effort. However, subjects are likely to be uncertain about quantities that are irregularly monitored or require processing multiple numbers, such as net wealth or monthly consumption. In such circumstances, subjects may refuse to answer, or may construct estimates. Svenson (1996) describes a decision process in which simple heuristics are used to produce a preliminary estimate, using *markers* and *editing* to simplify and group information; see Kahneman and Tversky (1979) and Coupey (1994). Next, the decision-maker engages in a process of differentiating the test estimate from alternatives, through an internal dialogue in which ambiguities are resolved so that consistent aspects of the test estimate are emphasized, through sharpening of perceptions of the plausibility of the test estimate and implausibility of alternatives, and through restructuring of the task by adding or resurrecting alternatives. There may also be consolidation of perceptions following selection of a final estimate, to reduce dissonance and promote development of rules and principles for future decisions. It is a particular problem for analysis if constructed estimates are systematically biased.

Psychometric Bias. There is extensive evidence from psychological experiments that humans are vulnerable to systematic cognitive illusions when dealing with uncertainty; see Rabin (1996). Table 1 lists some of the cognitive errors that appear regularly in experimental settings, and may be factors in economic survey responses. This paper will focus on biases induced by anchoring to prompts presented by questions on economic variables, and show that anchoring bias is a significant issue in consumption and savings variables of key interest for study of the elderly.

Anchoring describes a family of effects observed in many psychological studies of beliefs about uncertain quantities, such as the length of the Amazon or the height of the tallest redwood (Tversky and Kahneman, 1974). Subjects in these studies are asked to judge whether a particular value (the bid) is higher or lower than the uncertain quantity, before stating their own estimate. A robust result is that subjects start from the bid, and fail to adjust fully to their base beliefs, so that their estimates are pulled toward the bid. Open-ended responses that follow up a bid are pulled toward the bid, and in "Yes/No" responses to the bid, minority responses are more prevalent than they would be if subjects were not influenced by the bid. Even an explicitly uninformative prompt, such as the output of a random device, can operate as an anchor. Large anchoring effects have been reported in diverse contexts and populations of respondents, including experts; for example, see Northcraft and Neale (1987).

A psychological explanation for the phenomenon of anchoring is that a prompt creates in the subject's mind, at least temporarily, the possibility that the uncertain quantity could be either above or below the prompt. This could result from classical psychophysical discrimination errors, or from a cognitive process in which the subject treats the question as a problem-solving task and seeks an appropriate framework for "constructing" a correct solution, utilizing the prompt as a cue. Both formal and informal education train individuals to use problem-solving protocols in which responses to questions are based not only on substantive knowledge, but also on contextual cues as to what a correct response might be. Consequently, it should be no surprise if subjects apply these protocols in forming survey responses.

In psychological experiments, anchoring is found even when the gate amount is explicitly random, suggesting that there is more to anchoring than "rational" problem solving. This could happen because subjects are sub-rational, making cognitive errors and processing information inconsistently, or because they are "super-rational", going beyond the substantive question to "model" the mind of the questioner and form superstitious beliefs about the behavior of Nature.[2]

Several other effects in Table 1 may be significant in economic survey responses, and are topics for further research. *Focal* effects occur when the

[2] It may seem a contradiction in terms to label superstition "super-rational". However, systems of superstitious beliefs may well be consistent with a probability model for Nature that contains elaborate patterns of causality and correlation too complex to effectively reject empirically. A complete Bayesian facing a complex world admits the possibility that apparently random events have a hidden structure of causation. There are powerful psychological forces, related to limited memory and recall, and the cogency of coincidences, that reinforce complex, superstitious world views; see McFadden (1974).

cognitive organization of qualitative information is categorical rather than extensive, and/or when categorical approximations are used to minimize reporting effort. Open-ended responses on many economic variables exhibit the focal phenomenon, with responses piled up at rounded off numbers. For example, travel times are usually reported in five minute intervals, monthly income in multiples of $500, etc. One explanation of focal effects is that quantitative information is stored in a series of successively refined partitions, with increasing effort and uncertainty associated with retrieval from deeper levels. Survey responses may correspond to the finest partition the individual maintains, or may correspond to a coarser level that can be accessed with less effort. We have found in AHEAD data that focal responses are more common among the cognitively impaired, and that the probabilities of giving focal responses are correlated across questions. The focal response phenomenon can have significant impacts on analysis of economic data. Since focal responses concentrate at rounded-off dollar amounts, growth or inflation are captured mostly through switches between focal points, rather than marginal adjustments. Because focal responses are "sticky", questions stated in nominal dollars are likely to lead to an overestimate of money illusion. Similarly, "no change" may be a focal point in expectations questions. Focal effects interact with framing, as changing the reporting periods or units change the focal points.

Several cognitive illusions are related to the effort required to retrieve various pieces of information; these might all be referred to as *availability* effects. Examples are *primacy/recency* effects, in which initial or most recent experiences are more readily recalled than ones in between, *saliency* effects in which the information that seems most important or relevant is emphasized to the exclusion of other information, and status quo effects in which historical experience is more easily retrieved than hypothetical alternatives. *Framing* and *anchoring* phenomena may be related to availability as well, with the question itself providing immediately accessible information. The possible impacts on economic survey responses are obvious: information on social security income is more accessible than asset income, so the former may provide an internal anchor for the latter; beliefs about mortality may be unduly influenced by the ages attained by relatives and friends, to the exclusion of baseline information from life tables; recent changes in health status may be weighed too heavily in predicting future health status, with insufficient allowance for regression to the mean.

Finally, there are several cognitive distortions related to motivation, self-control, and projection of self-image that may affect the recall and filtering of information. Subjects may overestimate quantities associated with socially desirable behavior and positive self-image. For example, the elderly may overstate their ability to do the routine tasks asked for in IADL batteries, and may underestimate their consumption of tobacco and alcohol. Individuals appear to establish rules that precommit themselves to strategically desirable behavior, and then to shape perceptions so they are consonant with these rules.

II. Elicitation Protocols and the Unfolding Bracket Method

A. Elicitation Formats. The most direct way to ask about a quantitative economic variable is to request an open-ended quantitative response. A problem with this method is that it often results in relatively high item non-response, as well as implausible extreme responses. A popular alternative that is effective in reducing these problems is to use an *unfolding bracket* elicitation format, which converts the quantitative question into an unfolding series of "Yes/No" questions. Subjects are presented with a series of gates or bids, and at each bid are asked whether the quantity of interest is at least as large as the presented bid. The bids are determined sequentially; i.e., a "Yes" response is followed by a larger bid, a "No" response by a smaller bid. The sequence of bids and responses establish a bracket for the quantity of interest.

There is considerable evidence that the unfolding bracket method is effective in reducing non-response rates. The method also avoids implausible extreme responses, although the confusion or inattention on the part of subjects that produces these problems in open-ended questions may also distort bracket responses. One reason that it works well may be that memory for quantitative data is organized in hierarchical categories, with recall at one level aiding (or anchoring) recall at the next. Then, qualitative or range questions require relatively little effort, whereas increasing detail requires an increasingly complex web of interconnected facts and prompts.

A disadvantage of the unfolding brackets method is that the presentation of bids may influence beliefs or response protocols, so that the distribution of responses over brackets is sensitive to the unfolding design. A common psychological phenomenon is that subjects asked to estimate a quantity following a prompt are <u>anchored</u> to the prompt: in "Yes/No" responses to the bid, minority responses are more prevalent than they would be if subjects were not influenced by the bid. For example, McFadden (1994) finds that unfolding bracket questions on willingness-to-pay for natural resources contain anchoring distortions; Green *et al* (1996) find similar results for both objective estimation and willingness to pay tasks.[3]

B. A Model for Unfolding Bracket Responses. We shall be interested in an economic variable q, such as log monthly consumption or log savings balances, that can be related to a vector of covariates **x** via a linear model

(1) $$q = \mathbf{x}\beta - \upsilon \ ,$$

where β is a vector of parameters and υ is a disturbance, independent of **x**, that has mean zero and a cumulative distribution function $G(\upsilon)$.[4] Then, $G(\mathbf{x}\beta - q)$ is the complementary cumulative distribution function (CCDF) of q, given **x**, in the population.

In case of an unfolding bracket response, an observation will be denoted $t,\mathbf{x},(b_1,y_1),(b_2,y_2),...,(b_K,y_K)$, where t is a *treatment*, b_k is the k-th bid; y_k is a response indicator for this bid, 1 if "Yes" and 0 if "No"; and K is the number of bid questions presented *and* answered. The treatment t determines the bids b_k, conditional on previous responses $y_1,...,y_{k-1}$. Assume a "Yes" response leads to a larger gate amount for the next question, and vice versa. Let q^{bot} and q^{top} denote bounds on beliefs, $-\infty \le q^{bot} \le q < q^{top} \le +\infty$, and augment the response pairs (b_k,y_k) with the pairs $(q^{bot},1)$ and $(q^{top},0)$. The notation $B = (b',b'')$ will be used for the interval determined by an unfolding bracket response; i.e., $(b',1)$ is the largest bid in an unfolding bracket sequence that elicits a "Yes" response, and $(b'',0)$ is the smallest bid that elicits a "No" response.

---

[3] In a variety of tasks involving estimation of an unknown physical quantity, such as the height of the tallest redwood, the sample distribution of unprompted open-ended responses is presumably an unbiased estimator of the population distribution of beliefs about the quantity, and the standard against which the accuracy of unfolding bracket responses should be judged. For economic questions such as willingness to pay or asset balances, unprompted open-ended responses may themselves be biased for various reasons, and these tests should simply be interpreted as tests of whether open-ended and unfolding bracket questions yield the same distributions of responses.
[4] For a variable such as savings balances that has a significant probability of being zero and is highly skewed, a bivariate selection model is a convenient setup. Assume $p = x\alpha - \eta$, $S = 0$ if $p \le 0$, and $q \equiv \log S = x\beta - \upsilon$ if $p > 0$, with (η,υ) having a bivariate distribution.

C. Outcomes When Responses are Error-Free. If there are no response errors, then a subject asked an open-ended question will give the true value q, and if asked an unfolding bracket question will indicate correctly the bracket in which his latent q falls. The probability that q exceeds a *bid* b is

$$(2) \qquad P(Y|\mathbf{x}) = G(\mathbf{x}\beta - b) \ ,$$

and the probability of observing a bracket $B = (b',b'')$ as a result of unfolding bracket responses under treatment t is

$$(3) \qquad P\{B|\mathbf{x},t\} = G(\mathbf{x}\beta - b') - G(\mathbf{x}\beta - b'') \ .$$

A completed sequence of gate responses will pick out a single final bracket; however, incomplete responses will span several final brackets.

When G can be placed in a parametric family, root-N consistent asymptotically normal (RCAN) estimates of the parameter vector β and parameters of G can be obtained by maximum likelihood, subject to identification and regularity conditions. When G is normal, this reduces to least squares in the case of open-ended responses, and ordered probit (with thresholds specified by the bracket boundaries) in the case of unfolding bracket responses; both β and the variance σ^2 of υ can be identified as long as there are at least three brackets or two treatments. Another case that permits parametric analysis occurs when there are no covariates; then $P\{b',b''|\mathbf{x}\}$ is a sum of multinomial probabilities, and one can estimate a *saturated* multinomial model.

Suppose G cannot be placed in a parametric family. Least squares remains a RCAN estimation method for β from open-ended data. Sample moments of q observations are RCAN estimates of the corresponding unconditional population moments. For a continuous function $r(q,\mathbf{x})$ for which the population conditional moment $\mathbf{E}_{q|\mathbf{x}}r(q,\mathbf{x})$ exists, an estimator that is RCAN under mild regularity conditions is

$$(4) \qquad \hat{\mathbf{E}}_{q|\mathbf{x}}q^k = \frac{1}{N}\sum_{i=1}^{N} r(q_i + (\mathbf{x}-\mathbf{x}_i)\hat{\beta}) \ ,$$

where $i = 1,...,N$ indexes observations and $\hat{\beta}$ is the least squares estimator of β. Leading examples are $r(q,\mathbf{x}) = q^k$ and $r(q,\mathbf{x}) = e^{kq}$, the k-th moments of q and e^q, respectively.

The case of unfolding bracket responses and nonparametric G presents a semiparametric estimation problem, with (3) specifying the probability of an observation, conditioned on $\mathbf{x}$. This falls within the general class of single-index models for which Horowitz and Neumann (1989), Ichimura (1993), Lee (1995), and others have provided RCAN estimators for β. When the distribution of bid levels specified by the experimental design has a positive density, as can be the case with a randomized design in which the bid levels are drawn from continuous distributions, Lewbel (1997) and Lewbel and McFadden (1997) give a simple weighted least squares estimator that is RCAN for β. However, this is not applicable to the AHEAD experiments, which followed a fixed design. An alternative when there is open-ended data available is to estimate β by least squares using this external data. This is the route we will follow in analyzing the AHEAD experiments. It should be noted that differences in question formatting and context between surveys could confound this analysis; we will introduce rescaling factors that will absorb some, although not necessarily all, of the effects of mismatches across surveys.

Given an external RCAN estimator $\hat{\beta}$ for β, it is possible to construct a simple RCAN estimator of the population moment $\mathbf{E}_{q|\mathbf{x}}r(q,\mathbf{x})$. For each subject in a sample i = 1,...,N, define residuals $\upsilon_i' = \mathbf{x}_i\hat{\beta} - b_i'$ and $\upsilon_i'' = \mathbf{x}_i\hat{\beta} - b_i''$, where b_i' is the highest bid at which the subject says "Yes" and b_i'' is the lowest bid at which the subject says "No". (By convention, we assume that q^{bot} would elicit a "Yes" response and q^{top} would elicit a "No" response.) Consider the 2N pairs $(\upsilon_i',1)$ and $(\upsilon_i'',0)$, sort them so the first arguments are in non-decreasing order, and let (υ_m,y_m) denote the m-th pair in this order, for m = 1,...,2N. Integration by parts gives

$$(5)\qquad \mathbf{E}_{q|\mathbf{x}_o}r(q,\mathbf{x}_o) = r(\mathbf{x}_o\beta,\mathbf{x}_o) + \int_{-\infty}^{+\infty} r'(\mathbf{x}_o\beta - \upsilon,\mathbf{x}_o)\cdot[G(\upsilon) - \mathbf{1}(\upsilon>0)]d\upsilon$$

$$\approx r(\mathbf{x}_o\beta,\mathbf{x}_o) + \sum_{m=1}^{2N} r'(\mathbf{x}_o\beta - \upsilon_m,\mathbf{x}_o)\cdot[G(\upsilon_m) - \mathbf{1}(\upsilon_m>0)]\cdot\frac{\upsilon_{m+1} - \upsilon_{m-1}}{2} + O(N^{-2}) ,$$

with the last approximation obtained by application of the trapezoid rule for numerical integration, subject to some regularity conditions, including smoothness conditions on r, and tail conditions on G; tail values of υ are defined by convention to take care of tail terms in the summation. Equation (5) suggests the estimator

$$(6)\qquad \hat{\mathbf{E}}_{q|\mathbf{x}_o} r(q,\mathbf{x}_o) = r(\mathbf{x}_o\hat{\beta},\mathbf{x}_o) + \sum_{m=1}^{2N} r'(\mathbf{x}_o\hat{\beta} - \upsilon_m,\mathbf{x}_o)\cdot[\bar{y}_m - \mathbf{1}(\upsilon_m>0)]\cdot\frac{\upsilon_{m+1} - \upsilon_{m-1}}{2} ,$$

where $\bar{y}_m$ is the average (in case of ties) of the y's corresponding to the value υ_m. Lewbel and McFadden (1997) show that this estimator is RCAN under fairly mild regularity conditions. However, a critical requirement is that either the bids levels be drawn randomly from a positive density, or that the covariates have a continuous component with sufficient variation so that G has a positive density. This excludes the case of no covariates and fixed number of treatments without randomized bids. One cannot achieve a consistent estimator of the expectation of r for this case, but the estimator

$$(7)\qquad \hat{\mathbf{E}}_{q|\mathbf{x}_o} r(q) = r(0) + \sum_{m=1}^{2N} r'(q_m)\cdot[\mathbf{1}(q_m<0) - \bar{y}_m]\cdot\frac{b_{m+1} - b_{m-1}}{2} ,$$

will have relatively satisfactory finite-sample properties because of the approximation properties of $\bar{y}_m$ and the trapezoid rule.

The models and estimators above developed for the case of no response error can be applied to data pooled across treatments under the null hypothesis of no treatment effects. Treatment effect parameters can then be used to test this hypothesis. Surveys in which the same bracket can be reached under alternative gating designs provide a simple but powerful nonparametric test of the hypothesis of no anchoring effects. McFadden (1994) uses this test to show that a two-gate protocol for eliciting willingness to pay, also called a double-referendum elicitation, produces anchoring distortions. An empirical test for the hypothesis of no anchoring distortion in response can also be carried out by eliciting unprompted open-ended responses from one sample, using the empirical distribution of responses from this sample to estimate the bracket probabilities for a second sample where an unfolding bracket protocol is used, and calculating a goodness of fit test of the estimated probabilities to the second sample frequencies. An asymptotically efficient version of this test carries out maximum likelihood estimation of the probabilities of responses in each final bracket, separately for each sample and for the pooled samples, and calculates a likelihood ratio statistic.

D. A Model for Anchoring to Unfolding Brackets. We develop a simple model for anchoring that combines features of a model proposed in Green *et al* (1996) for anchoring to a starting point prompt, and a model proposed by Hurd for unfolding brackets. The premise of this model is that beliefs are stationary: gate choices create a temporary *discrimination* problem, but past history has no effect on current discrimination tasks. When a subject with a belief q is presented with a sequence of gate amounts b_k, responses are based on a comparison of the latent q with $b_k + \eta_k$, where η_k is a *perception error.* We assume the errors η_k are distributed independently across successive gates, and have CDF's $T_k(\cdot)$ that are symmetric about zero. Let $s_k = 2\cdot y_k - 1$ be a response indicator that is +1 for "Yes" and -1 for "No". Then, $s_k = 1$ if $b_k + \eta_k < \mathbf{x}\beta - \upsilon$, so that the probability of this event given υ is $T_k(s_k(\mathbf{x}\beta - b_k - \upsilon))$. We use the notation (b_k,y_k) and (b_k,s_k) interchangeably. The probability of an observation is then

$$\text{(8)} \qquad P((b_1,s_1),(b_2,s_2),\ldots,(b_K,s_K)\,|\,\mathbf{x},t) = \int_{q^{bot}}^{q^{top}} \prod_{k=1}^{K} T_k(s_k(\mathbf{x}\beta - b_k - \upsilon))\cdot G'(\upsilon)d\upsilon \ .$$

We will term this the *imperfect discrimination model of anchoring to unfolding brackets*. Imperfect discrimination is usually associated with physical stimuli, such as pitch or loudness of sounds, or brightness of lights, and it is not obvious that it is relevant when the stimuli are precisely stated numbers. However, ambiguity about whether q and the gate amounts refer to precisely the same quantity, or whether there are differences in scope or scale, can induce perception errors. For example, in unfolding bracket questions about savings balances, the subject may be unsure whether balances in Certificates of Deposit or IRA accounts should be included.

This model is able to capture several of the stylized features of anchoring. First, $P((b_1,1)\,|\,\mathbf{x},t) = \int_{q^{bot}}^{q^{top}} T_1(\mathbf{x}\beta - b_1 - \upsilon)\cdot G'(\upsilon)d\upsilon \equiv R(\mathbf{x}\beta - b_1)$, where R is the CDF of the random variable $\upsilon + \eta_1$. This a mean-preserving spread of υ, so that the probability of minority responses to bids will be increased. Second, suppose two gate designs lead to the same bracket; e.g., $((b',1),(b'',-1))$ and $((b'',-1),(b',1))$. The respective probabilities

$$P((b',1),(b'',-1) \mid \mathbf{x},t) = \int_{q^{bot}}^{q^{top}} T_1(\mathbf{x}\beta - b' - \upsilon) \cdot T_2(-\mathbf{x}\beta + b'' + \upsilon) \cdot G'(\upsilon) d\upsilon$$

$$P((b'',-1),(b',1) \mid \mathbf{x},t) = \int_{q^{bot}}^{q^{top}} T_1(-\mathbf{x}\beta + b'' + \upsilon) \cdot T_2(\mathbf{x}\beta - b' - \upsilon) \cdot G'(\upsilon) d\upsilon$$

can differ if $T_1 \neq T_2$. For example, if T_1 is more disperse than T_2 and $b' < \mathbf{x}\beta$, then $P((b',1),(b'',-1) \mid \mathbf{x},t) > P((b'',-1),(b',1) \mid \mathbf{x},t)$.

A parametric version of the model assumes that υ and the η_k are all normally distributed, with standard deviations σ and λ_k, respectively. Then discrimination follows a classical psychophysical model of Thurstone (1927). The probability of an observation in this *normal imperfect discrimination model* is then

$$(9) \qquad P((b_1,s_1),(b_2,s_2),\ldots,(s_K,y_K) \mid \mathbf{x},t) = \int_{-\infty}^{+\infty} \prod_{k=1}^{K} \Phi\left(\frac{s_k(\mathbf{x}\beta - b_k - \upsilon)}{\lambda_k}\right) \cdot \phi\left(\frac{\upsilon}{\sigma}\right) \cdot \frac{d\upsilon}{\sigma},$$

The parameters of this model can be estimated by maximum likelihood, with numerical integration used to evaluate the integral. An overall assessment of the goodness of fit of the normal model can be performed by a likelihood ratio test against the saturated model. Rejection of the normal model could occur either because the discrimination process above is not adequate to describe anchoring behavior, or because q is not normally distributed.

A relaxation of this parametric model retains the Thurstonian discrimination, but takes G to be an empirical distribution obtained from external data. To assure that the model is well-behaved numerically for small λ's, these empirical distributions are interpolated. This corresponds to assuming that the external distribution is a linear spline with knots at the observed points, and can be interpreted probabilistically as sampling from piecewise uniform densities with breaks at the observations.

Consider an external open-ended sample of size J in which q is observed, along with covariates $\mathbf{x}$. Assume that least squares applied to the regression equation $q = \mathbf{x}\beta - \upsilon$ yields a RCAN estimate $\hat{\beta}$ of β, and define the least squares residuals $u_j = \mathbf{x}_j\hat{\beta} - q_j$. Assume these residuals are indexed so that $u_1 \leq \ldots \leq u_J$. These residuals then define an empirical CDF G_J that can be used in the imperfect discrimination model. To avoid numerical analysis problems when discrimination is sharp, we use a linear spline smoothing of the empirical CDF, with 2J+2 knots placed at each u_j and at midpoints between the u_j, constrained so that the expectation of the smoothed distribution, conditioned on the interval formed by the knots bracketing u_j, equals u_j. For numerical integration, we choose evenly spaced points between successive knots. This construction preserves the mean zero property of the residuals. Let (υ_m,p_m) for m = 1,...,M denote the evaluation points and weights for the numerical integration. Add a parameter α to account for scaling differences between AHEAD and the external survey. Then, the probability of an observation in the imperfect discrimination to unfolding brackets model is

$$(10)\qquad P((b_1,s_1),(b_2,s_2),...,(b_K,s_K)\,|\,\mathbf{x},t) = \sum_{m=1}^{M} p_m \prod_{k=1}^{K} T_k(s_k(\alpha+\mathbf{x}\hat{\beta}-b_k-\upsilon_m)) \;.$$

In the case of normal discrimination, this becomes

$$(11)\qquad P((b_1,s_1),(b_2,s_2),...,(b_K,s_K)\,|\,\mathbf{x},t) = \sum_{m=1}^{M} p_m \prod_{k=1}^{K} \Phi\left(\frac{s_k(\alpha+\mathbf{x}\hat{\beta}-b_k-\upsilon_m)}{\lambda_k}\right) \;.$$

The parameters of this model, α and the λ_j, can be estimated by maximum likelihood. We shall term (11) the *empirical prior normal discrimination* model.

III. An Experiment on Anchoring in the AHEAD Survey

A. The AHEAD Experiment. The AHEAD panel study, in progress to study the economic and health status of the elderly, provides the primary data for this paper.[5] Unfolding brackets are used as follow up to non-response on a variety of economic questions regarding income and assets, and have proven quite effective in reducing item non-response. To test for anchoring effects in these elicitations, an experimental module was introduced in the second wave of the panel, administered in

[5] The survey is being conducted by the University of Michigan Survey Research Center on behalf of the Institute on Aging of the National Institute of Health. The survey questionnaires and data are available at http://www.umich.edu//~hrswww/ .

Fall 1995, that asked questions on savings account balances and on consumption, using seven alternative treatments, randomly assigned to each respondent. Each treatment specified a gating design for questions on savings balances, and another design for monthly consumption. A total of 4855 subjects were administered the experimental module. Figures 1-7 describe the unfolding bracket questions for each of the treatments. If there are no distortions in response due to the brackets, then the proportions of subjects appearing in the various brackets should be independent of treatment. Because of the random assignment of treatments, tests for anchoring can be carried out without considering covariates. However, when modeling the effects of anchoring, one will want to consider the effect of covariates that may affect the magnitude of anchoring effects.

B. Other Data. As external open-ended data supplements for the AHEAD questions on consumption and savings balances, we have analyzed consumption data from the Fall 1994 Consumer Expenditure Survey (CES), and savings data from the 1989 Survey of Consumer Finances (SCF). We have selected measures from these surveys that closely match AHEAD variable definitions, and have converted all dollar values to 1995 dollars using the Consumer Price Index. For example, the mean age of household heads in wave 1 of AHEAD was 76.2. In the subsamples of CES and SCF with heads aged 70+, the mean ages of heads was 77.1 and 74.5, respectively. Marital status reveals some significant difference in the surveys: the proportions married among households with heads aged 70+ are 0.561, 0.398, and 0.570 in AHEAD, CES, and SCF, respectively. Thus, CES includes a larger fraction of unattached individuals.

IV. Consumption

A. Measuring Consumption. The level of consumption, and its relation to the annuitized value of wealth, are the key determinants of the economic well-being of the elderly, determining current and prospective poverty rates. Consumption data can be obtained from consumer expenditure surveys, which are typically panels with expenditure diaries; in household panels as the difference between stated income and the imputed savings required to account for stated changes in asset holdings; or by direct questions. Subjects are likely to monitor some consumption components, such as utilities, food expenditures, and major durable purchases, more closely than others, and inferring total consumption from the more reliably reported components is an alternative to asking directly for total consumption.

B. Data from the Consumer Expenditure Survey. A relatively reliable picture of the distribution of consumption can be obtained from the Fall 1994 Consumer Expenditure Survey of the Bureau of Labor Statistics. This survey collects panel data on detailed expenditure categories, which then can be aggregated to total expenditure, and contains 772 households aged 70+. We have constructed a total consumption measure from this survey that closely matches the definition of consumption used in the AHEAD survey. Figure 8 gives the empirical CCDF of total consumption for these households, as well as separate CCDF's for Singles and Couples. Figure 9 gives the Lorenz curve for household consumption. For all households aged 70+, the median monthly consumption level in 1995 dollars is $1224, and the mean is $1735, with a standard deviation of $1801. The poverty level in 1995 was $768 per month for elderly couples, and $609 for elderly singles. Then, 5.2 percent of couples and 23.0 percent of singles were below the poverty level; in aggregate, 13.8 percent of households and unattached individuals over age 70 are below the poverty level. Note that the poverty comparisons are being made in terms of consumption rather than income. Some households who are below poverty levels of income can by decumulating assets have consumption levels above the poverty line. Then, the 13.8 percent poverty rate defined in terms of consumption may be consistent with the official poverty rate of 16.3 percent among those age 70+; see Hobbs (1996).

Subjects in Wave 2 of the AHEAD panel who were given Treatments 1,2 and 7 were first asked an open-ended question about consumption, with an unfolding bracket follow-up to non-response. Figure 10 gives the distribution of open-ended responses, without correction for the selection effects of non-response. Consumption levels obtained from AHEAD are uniformly below those obtained from the CES, some combination of selection effects, differences in the definition of consumption between CES and the understanding of AHEAD respondents, and response bias in the AHEAD data. Noteworthy is the frequency of focal responses in the AHEAD survey at $500, $1000, $1500, and $2000, and the dispersion of responses in comparison to the CES. Table 2 provides some summary statistics on consumption, based on the CES survey. The CES data show substantially higher mean consumption than AHEAD, $1738 versus $1252, and somewhat lower dispersion, with standard deviations of $1833 versus $2376. This pattern is repeated in the medians, $1224 versus $1000. An examination of the CES and AHEAD consumption distributions indicates that differences occur primarily in the tails. There is a significantly thicker lower tail in the AHEAD distribution, with

CES having 10.8 percent (SE = 1.1 percent) below \$500 and AHEAD having 19.3 percent (SE = 1.1 percent) below this level. This suggests that very low income households have a strong tendency to underestimate consumption, perhaps because they fail to consider items such as consumption in kind that are included in the CES. However, the upper tail of the AHEAD open-ended responses is thinner than CES, with 23.2 percent (ES = 1.5 percent) of CES respondents above \$2000, and 18.4 percent (SE = 1.0 percent) of AHEAD respondents above this level. An economic explanation of this pattern would require that there be income expenditure components that are part of the CES definition of consumption, but are not considered consumption by elderly households.

To determine the influence of demographics on consumption levels, we regressed log consumption on selected demographic variables, using the CES data. The results are given in Table 3. There is an economically, but not statistically, significant decline in consumption with age, a combination of life cycle and cohort effects. Figure 11 shows this relationship. If the disturbances in the regression have median zero, then this can be interpreted at the relation of median consumption to age. We find significant positive effects of education, presumably tied to lifetime earnings and current income, and significant negative effects of living alone and of being a female head.

C. The AHEAD Data. The consumption module asks for total monthly consumption expenditures. The experiment initially asks for an open-ended response under Treatments 1, 2, and 7, with unfolding bracket follow up for non-respondents. We term the subjects who responded to these follow-up brackets the *residual bracket* respondents. The remaining treatments forced bracket responses. Gate starting values were \$500, \$1000, \$2000, and \$5000. Figures 1-7 describe each consumption treatment and the gate designs. Subjects forced to give unfolding bracket responses (Treatments 3-6) had a first gate response rate of 98.2 percent, and a complete bracket response rate of 93.5 percent. Of subjects asked the initial open-ended consumption question (Treatments 1,2,6), 64.2 percent gave a usable response. Non-respondents were followed up with unfolding bracket questions, with a response rate on the first gate question of 89.3 percent and a complete bracket response rate of 50.7 percent. The combined open-ended or first gate response rate was 96.2 percent, and the open-ended or complete bracket response rate was 84.0 percent. Unfolding brackets are therefore *very effective in reducing item non-response when used as a* *follow up to an open-ended question*. However, subjects given an initial open-ended question have a higher rate of incomplete response than those facing only bracket questions, suggesting that once a subject admits to not knowing a quantity, there is more reluctance to giving possibly speculative gate responses. The quality of bracket responses obtained from subjects who would be non-respondents to an open-ended question remains an issue. Residual bracket respondents complete the unfolding bracket sequence at a much lower rate than subjects in general, indicating that incomplete bracket response may in itself be a useful indicator of subject uncertainty.

Analysis is carried out in terms of log consumption and log bracket quantities. Figure 12 gives the CCDF of consumption, for each starting value, for the subjects who give complete bracket responses, including incomplete responses. Higher starting values induce significantly higher responses. Table 4 gives the sample sizes, medians, and means for the bracket responses for each treatment, as well as for the treatments grouped by starting value or by response format (forced bracket versus initial open-ended). Means and medians are also given for the open-ended responses. *The location of the distribution of stated consumption rises sharply with starting value*, as can be seen by comparison of the treatments grouped by starting value in the second panel of the table: a starting value of $500 leads to median consumption of $878, while a starting value of $5000 leads to median consumption of $1455. A regression of the nonparametric mean of log consumption on the log starting value yields a coefficient of 0.235, with a standard error of 0.028, indicating that *a 100 percent increase in the starting value induces a 23.5 percent increase in stated consumption.*

The median consumption among residual bracket responses ($931) is significantly lower than that for the forced bracket respondents ($1129) facing the same starting values. In some combination, open-ended non-response may be associated with true lower consumption levels and with an effect in which an initial open-ended question acts to depress subsequent bracket responses. The mean open-ended response ($1253) is significantly lower than the mean bracket response for the forced bracket respondents ($1572), suggesting an overall tendency for bracket responses to be higher than open-ended responses. This finding is consistent with other studies of anchoring (Kahneman and Jacowitz, 1994; Green *et al*, 1996) which find for quantities that have distributions skewed to the right, anchors in the middle or upper tail of the distribution tend on average to elevate bracket responses above open-ended responses. There is inconsistent evidence that anchors in the lower tail of the distribution have the reverse effect, lowering bracket response below open-ended response. In many, but not all, cases where the quantity has an objective value, the open-ended responses appear to be located closer to the objective value than the bracket responses. Comparing the CES data and the AHEAD open-ended responses with data from forced bracket responses in AHEAD, and taking the CES data to be closest to true consumption, we conclude that subject's beliefs about consumption levels as reflected in the AHEAD open-ended responses are systematically biased downward, and the effect of forced brackets is to elevate responses compared with the open-ended distribution, reducing but not completely offseting the initial bias in beliefs.

The non-parametric and log normal models without anchoring, or by treatment, give similar measures of location. Likelihood ratio tests reject the hypothesis of no anchoring at any conventional level of significance, while the hypothesis of log normality, given the maintained hypothesis of no anchoring, is weakly rejected.[6]

Figure 13 compares the CCDF of the open-ended responses with the estimated CCDF's for pooled forced bracket responses and pooled residual bracket responses. The open-ended responses generally are more concentrated than the bracket responses, a pattern consistent with the implication of anchoring that minority responses are increased. There are significant focal points among open-ended responses, particularly at \$400, \$500, \$1200, \$1500, \$2000, and \$2500. The presence of significant focal points is itself an indication of response bias, as there are no factors operative in the economy that tend to cluster consumers at rounded-off consumption levels. An interesting psychometric question still to be answered is whether individuals tend to do mental accounting on quantities in terms of focal categories, so the focal levels will also play a substantial role in bracket responses, or whether they represent a reporting shorthand for more continuous

---

[6] The completed log likelihood for the saturated model is -4403.90. The completed log likelihood for the nonparametric model with no anchoring effects is -4522.39. Then, a likelihood ratio test statistic for the hypothesis of no anchoring effects is 236.98 with 24 degrees of freedom. This hypothesis is rejected at any conventional significance level. A normal parametric model with no anchoring effect has log likelihood -4525.92. Then, a likelihood ratio test statistic for the hypothesis that consumption is log normal, given the maintained hypothesis of no anchoring, is 7.047, with two degrees of freedom. The hypothesis is rejected at the 5 percent level, but not at the one percent level.

underlying beliefs. There is some psychological evidence for the former explanation, in which case the simple discrimination model for gate response requires elaboration.

The pattern in Figure 13 in which both open-ended and selected bracket responses are lower than forced bracket responses suggests that both a *selection effect*, in which low consumption individuals are more likely to be non-respondents on the open-ended question, and a *psychometric effect*, in which responses to unfolding brackets are higher than to open-ended questions, may be operating. A test for the former effect, which may be confounded by the latter one, can be carried out using a bivariate selection model. First, a binomial probit model for open-ended response is estimated as a function of subject characteristics, including sex, an indicator if the respondent handles the family finances, Cognitive impairment, Age, Marital status, HS Grad, some college , and dummy variables wealth quartile in the first wave of AHEAD. Then, open-ended log consumption responses are regressed on these subject characteristics and on an inverse Mills ratio term constructed from the probit model. A test for endogenous selection effects can be carried out by testing if the coefficient on the inverse Mills ratio variable is zero; this can be done using a conventional T-test. Table 5 gives the estimates of the two models. The probability that a respondent will answer an open-ended question is significantly higher if the respondent handles household finances, is not cognitively impaired, and is relatively young. Respondents with college are more likely to respond, and respondents in the top wealth quartile are less likely to respond. Confidentiality may be a factor in the last effect, but the additional cognitive effort required to accumulate a larger number of accounts and activities may be more important. The level of consumption is significantly increased for respondents who are married, or whose wealth is in the upper two quartiles. The coefficient on the inverse Mill's ratio is insignificant, indicating that there is no systematic bias in open-ended reported consumption explained by endogenous selection. However, the significant downward shift in consumption estimated from either open-ended or residual bracket responses indicates that there may be psychometric biases which tend to shrink open-ended responses and reduce "Yea-saying" in residual bracket responses, relative to the case of forced brackets.

We have estimated the empirical prior normal imperfect discrimination model (11) for the AHEAD bracket respondents, using the CES data to form the external prior G_m and using a five-point numerical integration procedure between the knots in the linear spline smoother of G_m. Table 6 gives the results for three alternative models. The first three models use the least squares parameters $\hat{\beta}$ estimated from the CES data to estimate $\mathbf{x}\beta$, so the only model parameters are the standard errors λ_k in the discrimination probabilities and the scaling factor α. In the first model, the restriction $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$ is imposed, and in the second model, this is relaxed to $\lambda_3 = \lambda_4 = \lambda_5$. The third model requires only $\lambda_4 = \lambda_5$, and estimates β directly from the AHEAD data. These models all indicate substantial discrimination errors, largest (e.g., λ smallest) for the initial bid and decreasing for each successive gate. Likelihood ratio tests show that model 2 is significantly better than model 1, and model 3 is significantly better than model 2. Also, a likelihood ratio test shows that model 1 is much better than a model without anchoring errors; i.e., with perfect discrimination.[7] *We conclude that there is a significant anchoring effect that is captured by the empirical prior imperfect discrimination model, with discrimination errors largest for the first bid and declining as the gate sequence continues.*

V. Savings Balances

A. Savings Balances. The 1991 median net worth of households aged 70+, in 1995 prices, was $92,609, according to U.S. Bureau of the Census (1994, Table G). Savings balances, or interest-earning assets at financial institutions, represent an important component of net worth of the elderly. The same source estimates that the distribution of net worth of households aged 65+ is 41.5 percent home equity, 21.0 percent savings balances, 12.1 percent in other interest-earning assets such as checking accounts, U.S. savings bonds, IRA or KEHOE accounts, and 9.4 percent stocks and mutual funds. The 1989 Survey of Consumer Finance contains 625 households aged 70+. In this population, 31 percent had zero savings balances, and the distribution is highly skewed. Expressed in 1995 prices, the median savings balance is $9130, the mean is $88,881, and the standard deviation is $299,920. Figure 14 shows the complementary CDF for savings balances in SCF, and Figure 15 shows the Lorenz curve for this asset.

Table 7 provides some summary statistics on savings, based on the SCF survey. Table 8 gives a probit model for having positive savings, as a function of

[7] The log likelihood for the perfect discrimination model is -4807.91, so that the likelihood ratio test statistic for model 1 against the perfect discrimination model is 1255, with 2 degrees of freedom.

demographic variables, and a regression of log savings on demographic variables, for the subpopulation with positive savings, with an inverse Mills ratio term to control for selection. The probit model shows that being married, more educated, and in good health are all associated with higher savings balances. The omitted health category is excellent health, so that the coefficients suggest that those in the best and the worst health are most likely to have zero savings. Since net worth is positively correlated with health status, the results suggest that those reporting excellent health tend to keep their assets in less liquid forms than savings balances. The coefficient on poor health is explained by some combination of the positive correlation of health status and net worth, and the drain on assets imposed by major health problems; see Smith (1995). The regression results indicate that savings balances for those with positive savings do not vary significantly with demographic variables. There is no evidence that selection is an issue. Figure 16 shows the profile of savings balances with age predicted by the regression model. The statistically insignificant upturn past age 80 is probably an artifact, although at advanced ages there may be conversion of assets, particularly housing equity, to more liquid form.

B. AHEAD Data. Recall that Figures 1-7 describe each of the savings gate designs; variable numbers refer to the survey instrument. These designs forced a bracket response for subjects who indicated in a preliminary question that they had positive savings. The starting gate values were \$1000, \$2000, \$5000, \$10,000, \$20,000, \$50,000, and \$200,000. Table 9 gives the sample sizes for each treatment. A total of 4855 subjects were administered the module containing the unfolding brackets treatments, with treatments assigned randomly. In the case of couples, both household members were given the questions, with treatments drawn independently for each.

In the sample, 69.6 percent of respondents indicated positive savings balances, and of these, 94.6 answered the first bracket question and 88.4 percent gave completed bracket responses; these response rates did not vary systematically with the first gate amount. Item response rates to open-ended questions on financial assets are typically in the 50-70 percent range, so that the unfolding bracket method is extremely effective in lowering non-response. Further, unfolding brackets reduce the problem of large reporting errors in open-ended responses. On the other hand, there is information loss from using bracket rather than continuous responses, and subject uncertainty that could produce large open-ended response errors could also contaminate bracket responses, and the presentation of gate cues may cause anchoring errors.

Parametric and nonparametric analysis of the savings data is carried out in terms of the log of savings balances and logs of gate amounts. The parametric analysis assumes that savings is log normal, conditioned on treatment. The nonparametric analysis uses the saturated multinomial model. The CCDF of savings balances was estimated from the bracket frequencies in the sample, using the subsample of respondents that completed the sequence of unfolding bracket responses, and using all respondents, with final bracket probabilities obtained from the saturated model.

Figure 17 shows the CCDF's under the different treatments computed for all observations, including partial bracket response. There are economically significant anchoring effects, with the CCDF from higher starting values showing substantially higher means and medians than those from lower starting values.[8] However, the effects of anchoring are not uniform. The starting values of $1000, $2000, and $5000 yield very similar CCDF's, and the starting value of $200,000 does not exert as much pull to the right as the starting value of $50,000. Table 9 gives the medians and means of these distributions; the pattern is consistent with that suggested by examining the CCDF's: the three lowest starting values lead to similar location measures, and the location measures then increase with starting value, except at the highest level. The distribution of savings is highly skewed, so that means are much larger than medians. The differences in location by starting value are strongly statistically significant; a likelihood ratio test rejects the hypothesis of no effect of gate design on response at any conventional level of significance.[9] A regression of the nonparametric mean of log savings on log starting value, weighted

---

[8] Nonparametric medians are estimated by log linear interpolation; this is equivalent to assuming that the density of log savings is uniform within each bracket. Nonparametric means are estimated assuming that savings is bounded between $1 and $800,000, and are sensitive to the assumed upper bound. Standard errors on the nonparametric estimates of the mean are lower than their parametric counterparts. This is primarily due to the assumed upper bound on savings imposed on the nonparametric estimator, eliminating the upper tail that contributes significantly to the variance of the parametric estimates.

[9] The completed log likelihood for the saturated model is -6122.99, and the completed log likelihood for a multinomial model with no anchoring is -6202.59. Then, the likelihood ratio statistic is 159.19 with 42 degrees of freedom.

to reflect the different numbers of observations for each treatment, yields a coefficient of 0.1037, with a standard error of 0.0365. Thus, a 100 percent increase in starting value produces a statistically significant 10.4 percent increase in estimated mean savings. A log normal savings model without anchoring produces mean and median estimates that are qualitatively similar to its nonparametric counterpart; however, the log normal parametric specification is rejected using a likelihood ratio test.[10]

The empirical prior imperfect discrimination model was estimated for the AHEAD unfolding bracket data on savings, using the SCF empirical distribution of savings balances, and using five evaluation points between the knots in the linear spline smoothing of the empirical prior. The results are given in Table 10. Discrimination errors are highest at the first gate (λ_1 lowest), with no significant variation in the λ_k for successive gates (model 2 versus model 1). In general the λ's are larger for the savings data than for the consumption data, corresponding to fewer discrimination errors and less anchoring effects. A likelihood ratio test rejects the hypothesis of perfect discrimination. One caveat is that the discrimination functions and the parameter α are operating both to explain variations across treatments and to explain differences between measured savings balances in SCF and beliefs about savings in AHEAD, and one cannot interpret the discrimination model parameters as arising solely from anchoring. Model 3 gives a significantly better fit than Model 1, indicating that there are some significant differences in the β parameters between the SCF and AHEAD populations. Married couples and heads with more education have significantly higher savings balances, and poor health leads to significantly lower savings balances.

Conclusions

This study has used an experimental module in the AHEAD panel to establish that anchoring can cause significant biases in unfolding bracket questions on quantitative

---

[10] The completed log likelihood for the log normal model without anchoring is -6216.79, and the likelihood ratio statistic for the log normal versus nonparametric models without anchoring is 28.24, with 5 degrees of freedom. When full interactions with treatments are allowed, so that log linear models estimated separately for each treatment are compared with the saturated model, a likelihood ratio test of the joint hypothesis that the log normal specification is correct is rejected; the likelihood ratio statistic is 121.6 with 35 degrees of freedom.

economic variables. In the case of savings, variation in starting values for unfolding brackets from $5,000 to $200,000 induces a 100 percent difference in estimated median savings. The anchoring is even stronger for consumption: increasing the starting value for unfolding brackets from $500 to $5000 induces nearly a doubling of estimated median consumption.

A simple model in which each gate presented to the subject can induce discrimination errors is successful in explaining much of these anchoring effects. Then, variation in unfolding bracket gates, in tandem with the discrimination model or an alternative model of anchoring, promise to be effective in identifying the effects of anchoring, and undoing most of these effects. We recommend that survey researchers who wish to use unfolding bracket elicitations adopt experimental variations in their designs that permit identification and correction of anchoring biases, and exercise caution in imputing economic variables based on stated brackets.

References

Bernheim, D.; Skinner, J.; Weinberg, S. (1996) "What Accounts for the Variation in Retirement Saving Among U.S. Households?" NBER Working Paper.

Bishop, R.; Heberlein. T. (1979) "Measuring values of extra-market goods: are indirect measures biased?" *American Journal of Agricultural Economics* **61**, 926-930.

Bohm, P. (1972) "Estimating willingness to pay: an experiment," European Economic Review, **3**, 111-130.

Boyle, K. (1989) "Commodity Specification and the Framing of Contingent-Valuation Questions," *Land Economics*; 65(1), 57-63.

Boyle, K.; Bishop, R.; Welsh, M. (1985) "Starting Point Bias in Contingent Valuation Bidding Games," *Land Economics*; 61(2), 188-94.

Cameron, T.; Huppert, D. (1991) "Referendum contingent valuation estimates: sensitivity to the assignment of offered values," *Journal of the American Statistical Association* **86**, 910-918.

Coupey, E. (1994) "Restructuring: Constructive Processing of Information Displays in Consumer Choice," *Journal of Consumer Research* **21**, 83-89.

Cowing, T.; Dubin, J.; McFadden, D. (1982) "The NIECS Database and Its Use in Residential Energy Demand Modelling," MIT Energy Laboratory Discussion Paper MIT-EL-82-041WP.

Garling, T. (1992) "The Importance of Routines for the Performance of Everyday Activities," *Scandinavian Journal of Psychology* **33**, 170-177.

Gourville, J. (1996) "Pennies a Day: Increasing Consumer Compliance Through Temporal Re-Framing," Harvard University Working Paper.

Green, D.; Jacowitz, K.; Kahneman, D.; McFadden, D. (1996) "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," *Resource and Energy Economics*, forthcoming.

Gustman, A.; Mitchell, O.; Samwick, A.; Steinmeier, T. (1997) "Pension and Social Security Wealth in the Health and Retirement Survey," NBER Working Paper 5912.

Harrison, G. (1992) "Valuing Public Goods with the Contingent Valuation Method: A Critique," *Journal of Environmental Economics and Management*; 23(3), 248-57.

Hobbs; F. (1996) *65+ in the United States*, U.S. Department of Commerce, Bureau of the Census, Current Population Reports P23-190.

Hogarth, R.; H. Kunreuther (1995) "Decision-Making under Ignorance: Arguing with Yourself," *Journal of Risk and Uncertainty* **10**, 15-36.

Holmes, T.; Kramer, R. (1995) "An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation," *Journal of Environmental Economics and Management* **29**, 121-32.

Horowitz, J. (1988) "Semiparametric M-Estimation of Censored Linear Regression Models," in G. Rhodes et al, eds, *Nonparametric and Robust Inference*, JAI Press.

Hutchinson, G.; Chilton, S.; Davis, J. {1995) "Measuring Non-use Value of Environmental Goods Using the Contingent Valuation Method: Problems of Information and Cognition and the Application of Cognitive Questionnaire Design Methods," *Journal of Agricultural Economics* **46**, 97-112.

Ichimura, H. (1993) "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71-120.

Jacowitz, K. and D. Kahneman (1995) "Measures of anchoring in estimation tasks," Personality and Social Psychology Bulletin, forthcoming.

Kahneman, D.; Slovic, P.; Tversky, A., eds (1982) *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Kahneman, D.; Tversky, A. (1973) "On the Psychology of Prediction," *Psychological Review* **80**, 237-251.
Kahneman, D.; Tversky, A. (1979) "Intuitive Prediction: Biases and Corrective Procedures," *Studies in Management Science* **12**, 313-327.
Kahneman, D.; Tversky, A. (1982) "On the Study of Statistical Intuitions," *Cognition* **11**, 123-141.
Kahneman, D.; Tversky, A. (1984) "Choices, Values, and Frames," *American Psychologist* **39**, 341-350.
Lowenstein, G. (1988) "Frames of Mind in Intertemporal Choice," *Management Science* **34**, 200-214.
Lewbel, A. (1997) "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," Brandeis University Working Paper.
Lewbel, A.; McFadden, D. (1997) "Estimating Features of a Conditional Distribution from Binomial Data," University of California, Berkeley, Working Paper.
Machina, M. (1981) "'Rational' Decision Making versus 'Rational' Decision Modeling?" *Journal of Mathematical Psychology* **24**, 163-175.
Machina, M. (1989) "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Literature* **32**, 1622-1668.
McFadden, D. (1974) "On Some Facets of Betting," in M Balch et al, eds, *Essays on Economic Behavior Under Uncertainty*, North Holland, 99-122.
McFadden, D. (1981) "Econometric Models of Probabilistic Choice," in C. Manski & D. McFadden, eds, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.
McFadden, D. (1986) "The Choice Theory Approach to Market Research," *Marketing Science* 5, 275-297.
McFadden, D. (1994) "Contingent valuation and social choice," *American Journal of Agricultural Economics*, **76**, 689-708.
McFadden, D. (1996) "Why is Natural Resource Damage Assessment So Hard?", Hibbard Lecture, University of Wisconsin.
McFadden, D. (1996) *Lectures in Simulation Assisted Statistical Inference*, University of California, Berkeley Working Paper.
Northcraft, G.; Neale, M. (1987) "Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions," *Organizational Behavior and Human Decision Processes*, **39**, 84-97.
Payne, J., Bettman, J.; Johnson, E. (1992) "Behavioral Decision Research: A Constructive Process Perspective," *Annual Review of Psychology* **43**, 87-131.
Poterba, J.; Summers, L. (1986) "Reporting Errors and Labor Market Dynamics," *Econometrica*, **54**, 1319-38.
Prelec, D. (1990) "Values and principles: Some limitations on traditional economic analysis," in A. Etzioni & P. Lawrence (eds) *Perspectives on Socioeconomics*. London: M.E. Sharpe.
Quattrone, G.; Tversky, A. (1986) "Self-Deception and the Voter's Illusion," in J. Elster, ed, *The Multiple Self*, Cambridge University Press.
Rabin, M. (1996) "Psychology and Economics," *Journal of Economic Literature*, forthcoming.
Schkade, D.; Payne, J.. (1994) "How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation," *Journal of Environmental Economics and Management*; **26**, 88-109.
Schultze, W.; McClelland, G.; Waldman, D.; Lazo, J. (1994) "Sources of Bias in Contingent Valuation," DOE/EPA Workshop on Using Contingent Valuation to Measure Non-Market Values.

Seip, K.; Strand, J. (1992) "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment," *Environmental and Resource Economics* **2**, 91-106.
Shafir, E.; Tversky, A. (1992) "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice," Cognitive Psychology **24**, 449-474.
Shafir, E.; Simonson, I.; Tversky, A. (1993) "Reason-Based Choice," *Cognition* **49**, 11-36.
Silberman, J.; Klock, M. (1989) "The behavior of respondents in contingent valuation: evidence on starting bids," *Journal of Behavioral Economics*, **18**, 51-60.
Smith, J. ()
Smith, V. (1979) "An experimental comparison of three public good decision mechanisms," Scandinavian Journal of Economics, **81**, 198-215.
Svenson, O. (1996) "On the Modeling of Human Choices in Descriptive Behavioral Decision Theory," Stockholm University Working Paper.
Thalor, R. (1980) "Toward a Positive Theory of Consumer Choice," *Journal of Human Behavior and Organization* **1**,39-60.
Thalor, R. (1985) "Mental Accounting and Consumer Choice," *Marketing Science* **4**, 199-214.
Thalor, R. (1990) "Savings, fungability, and mental accounts," *Journal of Economic Perspectives* 4, 193-205.
Thalor, R. (1991) *Quasi-Rational Economics*, Russell Sage Foundation.
Thaler, R.; Johnson, E. (1990) "Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice," *Management Science*, **36**, 643-660.
Thurstone, L. (1927) "A Law of comparative Judgment," *Psychological Review* **34**, 273-286.
Tversky, A.; Kahneman, D. (1971) "Belief in the Law of Small Numbers," *Psychological Bulletin* **76**, 105-110.
Tversky, A.; Kahneman, D. (1974) "Judgment under uncertainty: heuristics and biases" *Science*, **185**, 1124-1131.
Tversky, A.; Kahneman, D. (1981) "The Framing of Decisions and the Psychology of Choice," *Science* **211**, 453-458.
Tversky, A.; Kahneman, D. (1991) "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* **107**, 1039-1061.
Tversky, A.; Shafir, E. (1992) "Choice Under Conflict: The Dynamics of Deferred Decisions," *Psychological Science* **3**, 358-361.
U.S. Bureau of the Census (1994) *Household Wealth and Asset Ownership*, Current Population Report P70-34.
Whittington, D. et al. (1992) "Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application," *Journal of Environmental Economics and Management*; 22(3), 205-25.