Volume Title: Frontiers in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90304-4

Volume URL: http://www.nber.org/books/wise98-1

Publication Date: January 1998

Chapter Title: Consumption and Savings Balances of the Elderly: Experimental Evidence on Survey Response Bias

Chapter Author: Michael D. Hurd, Daniel McFadden, Harish Chand, Li Gan, Angela Menill, Michael Roberts

Chapter URL: http://www.nber.org/chapters/c7306

Chapter pages in book: (p. 353 - 392)

8 Consumption and Savings Balances of the Elderly: Experimental Evidence on Survey Response Bias

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

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

8.1 Psychometric Biases in Economic Survey Data

8.1.1 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 on 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 nonresponse 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

Michael D. Hurd is professor of economics at the State University of New York, Stony Brook, a researcher at RAND, and a research associate of the National Bureau of Economic Research. Daniel McFadden is professor of economics at the University of California, Berkeley, and a research associate of the National Bureau of Economic Research. Harish Chand is a graduate student at the University of California, Berkeley. Li Gan is a graduate student at the University of California, Berkeley. Angela Merrill is a graduate student at the University of California, Berkeley. Michael Roberts is a graduate student at the University of California, Berkeley.

This research was supported by a grant from the Institute on Aging of the National Institutes of Health to the National Bureau of Economic Research. The authors are indebted to Danny Kahneman, whose remarks on anchoring motivated this research.

often too sanguine about the reliability of subjects' statements regarding objective economic data.

8.1.2 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 because of the 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, Dubin, and McFadden (1982) found in an Energy Information Administration panel of houses that 5 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 Panel Study of Income Dynamics and the Health and Retirement Survey.

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 survey of Asset and Health Dynamics among the Oldest Old (AHEAD) was varied by design.

8.1.3 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 question from the date of the interview, or the past 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 nonresponse or a dispersion in responses resulting from different implicit assumptions about 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 nonresponse. 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, the Longitudinal Study of Aging, 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 will 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 *nonpecuniary 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 nonpecuniary 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 balances or monthly social security checks, 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 them to process 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; 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 1998). Table 8.1 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 will show that anchoring bias is a significant issue in consumption and savings variables of key interest for the 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

Table 8.1 **Cognitive Illusions**

Effect	Description
Anchoring	Responses are influenced by cues contained in the question.
Availability	Responses rely too heavily on readily retrieved information, and too little on background information.
Context	Previous questions and interviewer interaction color perception.
Framing/reference point	Question format changes saliency of different aspects of the cognitive task.
Focal	Quantitative information is stored or reported categorically.
Primary/recency	Initial and recently experienced events are the most salient.
Projection	Responses are consonant with the self-image the subject wishes to project.
Prospect	The likelihoods of low-probability events are misjudged and treated either as too likely or as impossible.
Regression	Causality and permanence are attached to past fluctuations, and regression to the mean is underestimated.
Representativeness	High conditional probabilities induce overestimates of unconditional probabilities.
Rule-driven	Motivation and self-control induce strategic responses.
Saliency	The most salient aspects of the question are overemphasized.
Status quo	Current status and history are privileged.
Superstition	Elaborate causal structures are attached to coincidences.
Temporal	Time discounting is temporally inconsistent.

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 estimates. 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 (e.g., 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 bid amount is explicitly random, suggesting that there is more to anchoring than "rational" problem solving. This could happen because subjects are subrational, making cognitive errors and processing information inconsistently, or because they are "superrational," going beyond the substantive question to "model" the mind of the questioner and form superstitious beliefs about the behavior of nature.[1]

Several other effects in table 8.1 may be significant in economic survey responses and are topics for further research. *Focal* effects occur when the cognitive organization of quantitative information is categorical rather than extensive 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. Travel times are usually reported in five-minute intervals, monthly income in multiples of $500, and so forth. 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 effects on the analysis of economic data. Since focal responses concentrate at rounded-off dollar amounts, growth or inflation is 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* because 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 those 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

1. It may seem a contradiction in terms to label superstition "superrational." 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 worldviews (see McFadden 1974).

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, several cognitive distortions related to motivation, self-control, and projection of self-image 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.

8.2 Elicitation Protocols and the Unfolding Bracket Method

8.2.1 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 nonresponse, 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; that is, a "yes" response is followed by a larger bid, a "no" response by a smaller bid. The sequence of bids and responses establishes a bracket for the quantity of interest.

There is considerable evidence that the unfolding bracket method is effective in reducing nonresponse 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 bracket 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. For example, McFadden (1994) finds that unfolding bracket questions on WTP for natural resources contain anchoring distortions; Green et al. (1998) find similar results for both objective estimation and WTP tasks.[2]

8.2.2 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 $\mathbf{x}$ via a linear model

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

where β is a vector of parameters and υ is a disturbance, independent of $\mathbf{x}$, that has mean zero and a cumulative distribution function $G(\upsilon)$.[3] Then $G(\mathbf{x}\beta - q)$ is the complementary cumulative distribution function (CCDF) of q, given $\mathbf{x}$, in the population.

In the case of an unfolding bracket response, an observation will be denoted

$$t,\ \mathbf{x},\ (b_1, y_1),\ (b_2, y_2), \ldots,\ (b_K, y_K),$$

where t is a *treatment*, b_k is the kth 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 $y_1, \ldots, y_{k-1}$. 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, $-\infty \leq q^{\text{bot}} \leq q < q^{\text{top}} \leq +\infty$, and augment the response pairs (b_k, y_k) with the pairs $(q^{\text{bot}}, 1)$ and $(q^{\text{top}}, 0)$. The notation $B = (b', b'')$ will be used for the interval determined by an unfolding bracket response; that is, $(b', 1)$ 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.

8.2.3 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 a subject 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

$$P(y \mid \mathbf{x}) = G(\mathbf{x}\beta - b), \tag{2}$$

2. 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 WTP 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.

3. 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 $p = \mathbf{x}\alpha - \eta$, $S = 0$ if $p \leq 0$, and $q \equiv \log S = \mathbf{x}\beta - \upsilon$ if $p > 0$, with (η, υ) having a bivariate distribution.

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

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

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 $P(b', b'' \mid \mathbf{x})$ 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 $\mathbf{E}_{q|x} r(q, \mathbf{x})$ exists, an estimator that is RCAN under mild regularity conditions is

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

where $i = 1, \ldots, N$ 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 kth moments of q and e^q, respectively.

The case of unfolding bracket responses and nonparametric G presents a semiparametric estimation problem, with equation (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 are 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 $\mathbf{E}_{q|x}r(q, \mathbf{x})$. For each subject in a sample $i = 1, \ldots, N$, define residuals $\upsilon'_i = \mathbf{x}_i\hat{\beta} - b'_i$ and $\upsilon''_i = \mathbf{x}\,\hat{\beta} - b''_i$, 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 nondecreasing order, and let (υ_m, y_m) denote the mth pair in this order, for $m = 1, \ldots, 2N$. Integration by parts gives

$$\mathbf{E}_{q|\mathbf{x}_0}r(q,\mathbf{x}_0) = r(\mathbf{x}_0\beta,\mathbf{x}_0) + \int_{-\infty}^{+\infty} r'(\mathbf{x}_0\beta - \upsilon,\mathbf{x}_0)\cdot[G(\upsilon) - \mathbf{1}(\upsilon > 0)]\,d\upsilon$$

$$(5) \qquad \approx r(\mathbf{x}_0\beta,\mathbf{x}_0) + \sum_{m=1}^{2N} r'(\mathbf{x}_0\beta - \upsilon_m,\mathbf{x}_0)\cdot[G(\upsilon_m) - \mathbf{1}(\upsilon_m > 0)] \cdot\frac{\upsilon_{m+1} - \upsilon_{m-1}}{2} + O(N^{-2}),$$

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

$$(6) \qquad \hat{\mathbf{E}}_{q|\mathbf{x}_0}r(q,\mathbf{x}_0) = r(\mathbf{x}_0\hat{\beta},\mathbf{x}_0) + \sum_{m=1}^{2N} r'(\mathbf{x}_0\beta - \upsilon_m,\mathbf{x}_0)\cdot[y_m - \mathbf{1}(\upsilon_m > 0)] \cdot\frac{\upsilon_{m+1} - \upsilon_{m-1}}{2},$$

where y_m is the average (in case of ties) of the ys 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 the covariates have a continuous component with sufficient variation so that υ_m 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

$$(7) \qquad \hat{\mathbf{E}}_{q|\mathbf{x}_0}r(q) = r(0) + \sum_{m=1}^{2N} r'(b_m)\cdot[\mathbf{1}(b_m < 0) - y_m]\cdot\frac{b_{m+1} - b_{m-1}}{2}$$

will have relatively satisfactory finite-sample properties because of the approximation properties of 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 WTP, 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.

8.2.4 Model for Anchoring to Unfolding Brackets

We develop a simple model for anchoring that combines features of a model proposed in Green et al. (1998) 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 CDFs $T_k(\cdot)$ that are symmetric about zero. Let $s_k = 2y_k - 1$ 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

$$P((b_1,s_1),\ (b_2s_2),\ldots,\ (b_K,s_K)\mid \mathbf{x},t) \tag{8}$$
$$= \int_{q^{\text{bot}}}^{q^{\text{top}}} \prod_{k=1}^{K} T_k(s_k(\mathbf{x}\beta - b_k - \upsilon))\cdot G'(\upsilon)\,d\upsilon.$$

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 individual retirement accounts should be included.

This model is able to capture several of the stylized features of anchoring. First,

$$P((b_1,1)\mid \mathbf{x},t) = \int_{q^{\text{bot}}}^{q^{\text{top}}} T_1(\mathbf{x}\beta - b_1 - \upsilon)\cdot G'(\upsilon)\,d\upsilon \equiv R(\mathbf{x}\beta - b_1),$$

where R is the CDF of the random variable $\upsilon + \eta_1$. This is 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, for example, $((b', 1), (b'', -1))$ and $((b'', -1), (b', 1))$. The respective probabilities

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

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

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

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

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 λs, these empirical distributions are interpolated. This 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 $m = 1, \ldots, M$ 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

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

In the case of normal discrimination, this becomes

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

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

8.3 An Experiment on Anchoring in the AHEAD Survey

8.3.1 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.[4] Unfolding brackets are used as follow-up to nonresponse on a variety of economic questions regarding income and assets and have proved quite effective in reducing item nonresponse. To test for anchoring effects in these elicitations, an experimental module was introduced in the second wave of the panel, administered in 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

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

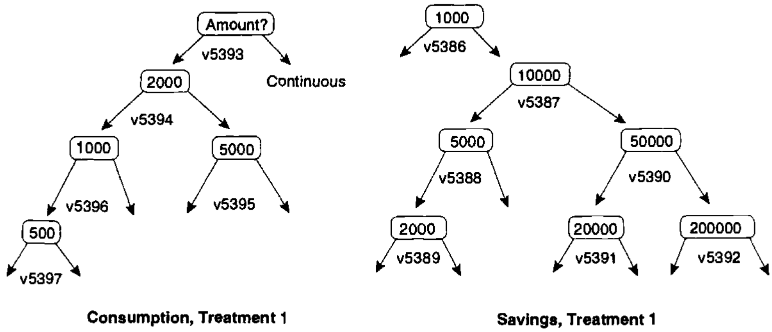


Fig. 8.1 Treatment 1

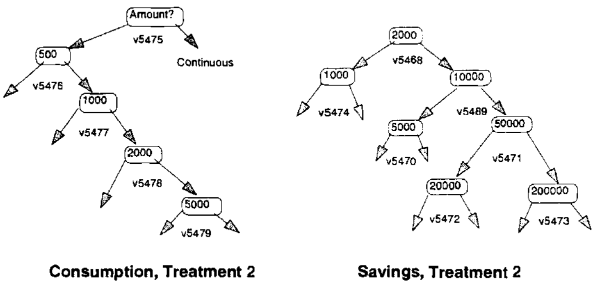
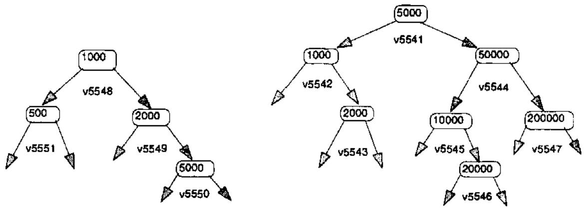


Fig. 8.2 Treatment 2

and another design for monthly consumption. A total of 4,855 subjects were administered the experimental module. Figures 8.1 through 8.7 describe the unfolding bracket questions for each of the treatments; variable numbers (e.g., v5397) refer to the survey instrument. 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.

8.3.2 Other Data

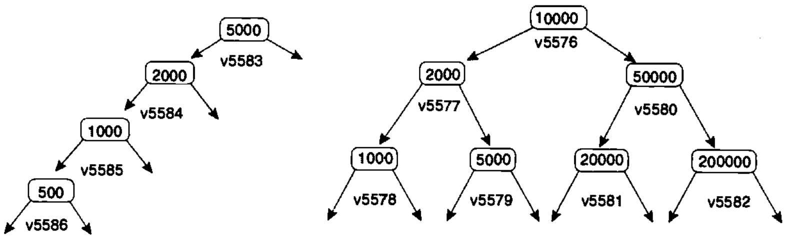
As external open-ended data supplements for the AHEAD questions on consumption and savings balances, we have analyzed consumption data from the fall 1994 Consumer Expenditure Survey (CES) and savings data from the 1989 Survey of Consumer Finances (SCF). We have selected measures from these surveys that closely match AHEAD variable definitions and have converted all dollar values to 1995 dollars using the CPI. For example, the mean age of



Consumption, Treatment 3

Savings, Treatment 3

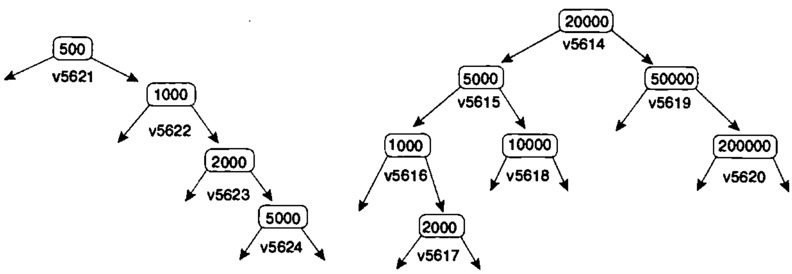
Fig. 8.3 Treatment 3



Consumption, Treatment 4

Savings, Treatment 4

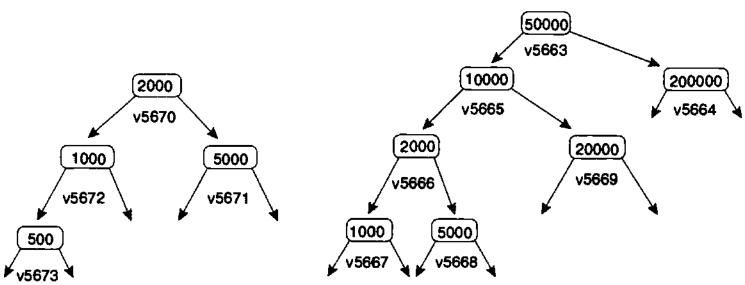
Fig. 8.4 Treatment 4



Consumption, Treatment 5

Savings, Treatment 5

Fig. 8.5 Treatment 5



Consumption, Treatment 6

Savings, Treatment 6

Fig. 8.6 Treatment 6

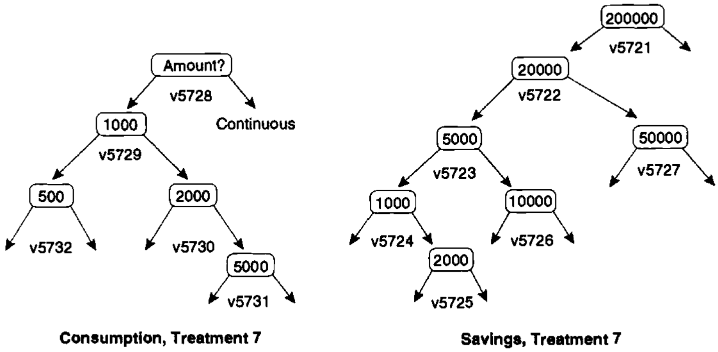


Fig. 8.7 Treatment 7

household heads in wave 1 of AHEAD was 76.2. In the subsamples of the CES and SCF with heads aged 70 or over, the mean ages of heads were 77.1 and 74.5, respectively. Marital status reveals some significant differences between the surveys: the proportions married among households with heads aged 70 or over are 0.561, 0.398, and 0.570 in AHEAD, the CES, and the SCF, respectively. Thus, CES includes a larger fraction of unattached individuals.

8.4 Consumption

8.4.1 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.

8.4.2 Data from the Consumer Expenditure Survey

A relatively reliable picture of the distribution of consumption can be obtained from the fall 1994 CES 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 or over.

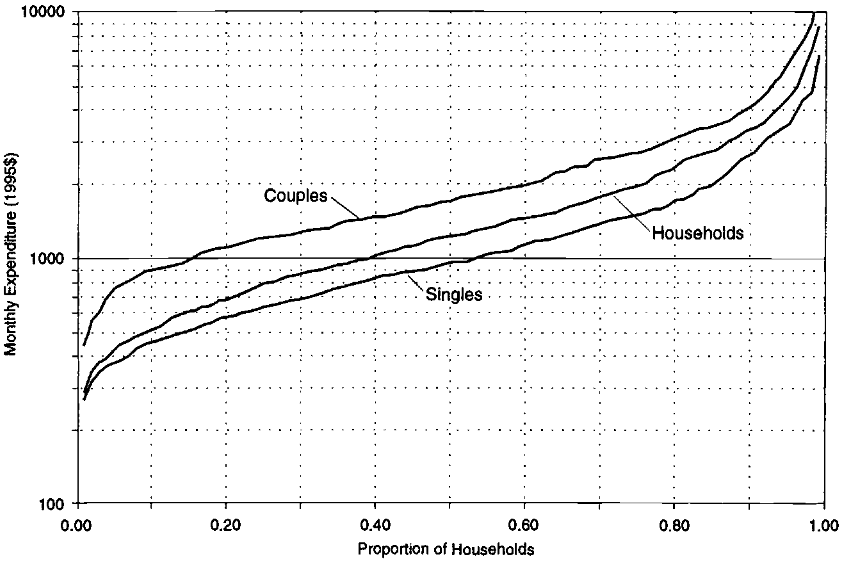


Fig. 8.8 Consumption in 1994 CES: households, singles, and couples aged 70 or over

We have constructed a total consumption measure from this survey that closely matches the definition of consumption used in the AHEAD survey. Figure 8.8 gives the empirical CDF of total consumption for these households, as well as separate CDFs for singles and couples. Figure 8.9 gives the Lorenz curve for household consumption. For all households aged 70 or over, the median monthly consumption level in 1995 dollars is $1,224, and the mean is $1,735, with a standard deviation of $1,801. 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 aged 70 or over (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 nonresponse. Figure 8.10 gives the distribution of open-ended responses, without correction for the selection effects of nonresponse. Consumption levels obtained from AHEAD are uniformly below those

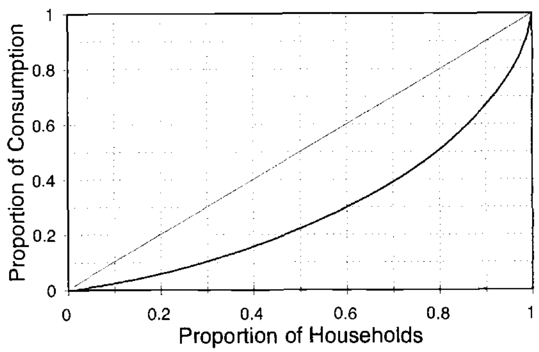


Fig. 8.9 Lorenz curve for consumption in 1994 CES: households aged 70 or over

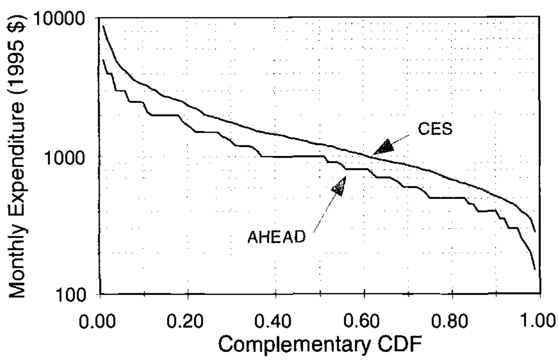


Fig. 8.10 Consumption in AHEAD and 1994 CES: households aged 70 or over

obtained from the CES, probably due to some combination of selection effects, differences in the definition of consumption between the 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, $1,000, $1,500, and $2,000, and the dispersion of responses in comparison to the CES. Table 8.2 provides some summary statistics on consumption, based on the CES survey. The CES data show substantially higher mean consumption than AHEAD, $1,738 versus $1,252, and somewhat lower dispersion, with a standard deviation of $1,833 versus $2,376. This pattern is repeated in the medians, $1,224 versus $1,000. 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 the CES having 10.8 percent (S.E. = 1.1 percent) below $500 and AHEAD having 19.3 percent (S.E. = 1.1 percent) below this level. This suggests that very low income households have a strong tendency to underestimate consumption,

Table 8.2 Consumption Summary Statistics

Variable	Coding	Mean	Standard Deviation
log(Monthly expenditure)	1995$	7.160	0.730
Monthly expenditure	1995$	1,738	1,833
Head age	Years	77.058	5.987
Spouse age	Years	71.166	7.045
married	1 = Yes, 0 = No	0.398	0.490
Head some college	1 = Yes, 0 = No	0.272	0.330
Spouse some college	1 = Yes, 0 = No	0.331	0.471
Home owner	1 = Yes, 0 = No	0.769	0.412
Head sex	1 = Male, 0 = Female	0.462	0.499
Spouse sex	1 = Male, 0 = Female	0.097	0.296
Head minority	1 = Yes, 0 = No	0.105	0.307
Spouse minority	1 = Yes, 0 = No	0.086	0.281
Lives with kids	1 = Yes, 0 = No	0.023	0.151
Head high school graduate	1 = Yes, 0 = No	0.539	0.492
Spouse high school graduate	1 = Yes, 0 = No	0.721	0.449
No spouse	1 = Yes, 0 = No	0.624	0.485
Couples			
Monthly expenditure		2,429	2,444
Home owner		0.886	0.318
Lives with kids		0.062	0.242
No spouse			
Monthly expenditure		1,322	1,154
Home owner		0.699	0.459
Lives with kids[a]		0.000	0.000

Source: Fall 1994 Consumer Expenditure Survey.

[a]Unattached individuals living with kids are considered part of the kid's household.

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 in the CES, with 23.2 percent (S.E. = 1.5 percent) of CES respondents above $2,000 and 18.4 percent (S.E. = 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 8.3. There is an economically, but not statistically, significant decline in consumption with age, a combination of life cycle and cohort effects. Figure 8.11 shows this relationship. If the disturbances in the regression have median zero, then this can be interpreted as 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.

Table 8.3 **Demographics and Consumption**

Variable	Coefficient	Standard Error
Constant	7.8907	0.5256
Head age	−0.0103	0.0070
Pos(Head age − 80)[a]	−0.0176	0.0149
No spouse	−0.3334	0.0648
Head some college	0.3094	0.0587
Head high school graduate	0.1694	0.0540
Head male	0.2524	0.0621
Head minority	−0.0776	0.0743
R^2	0.2768	

Source: Fall 1994 Consumer Expenditure Survey.
Note: Ordinary least squares estimation; dependent variable is log(Monthly expenditure).
[a]Pos(x) = max(0,x).

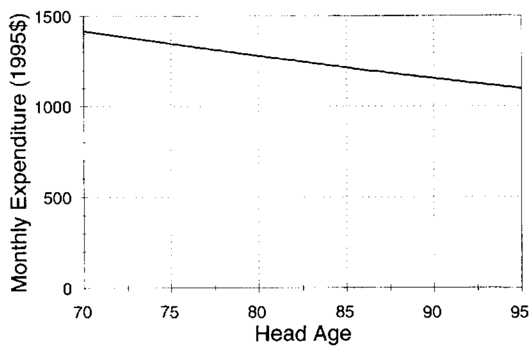


Fig. 8.11 Consumption vs. age in 1994 CES

8.4.3 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 nonrespondents. 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, \$1,000, \$2,000, and \$5,000. Figures 8.1 through 8.7 describe each consumption treatment and the gate designs. Subjects forced to give unfolding bracket responses (treatments 3 through 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, and 6), 64.2 percent gave a usable response. Nonrespondents were followed up with unfolding bracket questions, with a response rate on the first

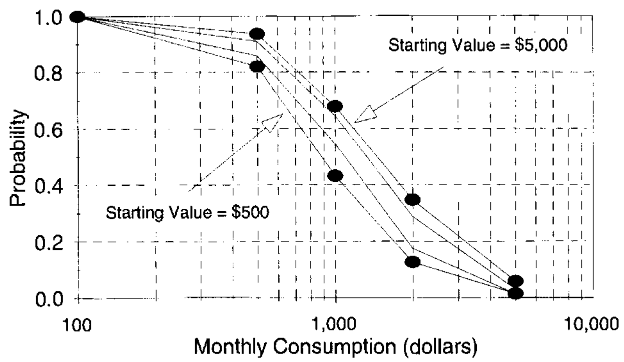


Fig. 8.12 Consumption CCDF by starting value: all bracket responses

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 nonresponse 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 give possibly speculative gate responses. The quality of bracket responses obtained from subjects who would be nonrespondents 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 8.12 gives the CCDF of consumption, for each starting value, for the subjects who give bracket responses, including incomplete responses. Higher starting values induce significantly higher responses. Table 8.4 gives 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 vs. 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 $886, while a starting value of $5,000 leads to median consumption of $1,455. 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.*

Table 8.4 Consumption: Sample Sizes, Medians, and Means

Treatment	Starting Gate Amount	Sample Size	Number with Open-Ended Response	Percentage of Bracket Responses Completed	Medians: Nonparametric[a]	S.E.[b]	Parametric[c]	S.E.[d]	Means: Nonparametric[e]	S.E.[f]	Parametric[g]	S.E.[h]
1	2,000[i]	739	492	53.8	1,061	87	1,128	72	1,732	108	1,513	88
2	500[i]	689	422	51.3	861	53	864	53	1,261	87	1,139	63
3	1,000	627	0	92.8	1,146	39	1,104	37	1,508	49	1,365	40
4	5,000	782	0	94.0	1,455	56	1,486	52	2,161	65	1,979	62
5	500	707	0	92.9	895	31	934	31	1,311	45	1,180	35
6	2,000	594	0	94.1	1,415	53	1,392	51	1,946	61	1,764	57
7	1,000[i]	717	464	47.0	897	62	967	69	1,466	98	1,352	89
2 and 5	500	1,396	422	81.5	886	26	915	27	1,298	40	1,170	31
3 and 7	1,000	1,344	464	79.7	1,090	36	1,066	33	1,497	44	1,364	38
1 and 6	2,000	1,333	492	82.3	1,326	46	1,310	42	1,884	53	1,695	49
Open-ended first (1, 2, 7)		2,145	1,378	50.7	931	35	980	37	1,485	57	1,331	46
Forced (3, 5, 6)		1,928	0	93.3	1,129	25	1,167	25	1,572	30	1,523	29
Pooled (1, 2, 3, 5, 6, 7)		4,073	1,378	81.2	1,077	22	911	18	1,358	31	1,237	22
Open-ended responses			1,378		1,000	9			1,253	64		
Overall		4,855	1,378	84.0	1,163	21	1,170	19	1,696	26	1,534	22

Completed Log Likelihoods	Semiparametric	DF	Normal	DF
No anchoring	−4,522.4	4.0	−4,525.9	2
Saturated	−4,403.9	28.0	−4,426.6	14
Imperfect discrimination	−4,442.6	6.0	−4,442.6	4

[a]Exponential of linearly interpolated CCDF of log consumption, with the CCDF estimated using a "saturated" multinomial model for all respondents.

[b]Standard error is estimated by (median) × $(a - b)/(2$ × (prob. of bracket) × (root N)), where (b,a) is the log consumption bracket containing the estimator. This estimator assumes that log consumption is uniformly distributed within the bracket containing the median.

[c]Exponential of the mean of a log normal distribution fitted by maximum likelihood estimation to bracket frequencies of log consumption.

[d]Standard error is estimated by (median) × (SD) × root (pi/2 × N), where SD is the estimated standard error of log consumption.

[e]Sum of bracket midpoints times estimated bracket probabilities.

[f]Standard error is estimated by square root of (sum of squared bracket midpoints times bracket probabilities minus median squared)/N.

[g]Exponential of (mean) +0.5 × (sigma)^2, where mean and sigma are estimates of the mean and standard deviation of log consumption.

[h]Standard error is estimated by (mean) × (SD) × root (1 + 0.5 × (SD)^2)/(root (N), where SD is the estimated standard error of log consumption.

[i]Subjects were first asked for an open-ended response, with unfolding brackets if there was no response to the open-ended question.

The median consumption among residual bracket responses ($931) is significantly lower than that for the forced bracket respondents ($1,129) facing the same starting values. In some combination, open-ended nonresponse 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 ($1,253) is significantly lower than the mean bracket response for the forced bracket respondents ($1,572), suggesting an overall tendency for bracket responses to be higher than open-ended responses. This finding is consistent with other studies of anchoring (Jacowitz and Kahneman 1995; Green et al. 1998), which find that for quantities with 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, open-ended responses appear to be located closer to the objective value than 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 subjects' 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 offsetting the initial bias in beliefs. The nonparametric and lognormal 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 lognormality, given the maintained hypothesis of no anchoring, is weakly rejected.[5]

Figure 8.13 compares the CCDF of the open-ended responses with the estimated CCDFs 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, $1,200, $1,500, $2,000, and $2,500. 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

5. The completed log likelihood for the saturated model is −4,403.90. The completed log likelihood for the nonparametric model with no anchoring effects is −4,522.39. 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 −4,525.92. Then a likelihood ratio test statistic for the hypothesis that consumption is lognormal, given the maintained hypothesis of no anchoring, is 7.047 with 2 degrees of freedom. The hypothesis is rejected at the 5 percent level, but not at the 1 percent level.

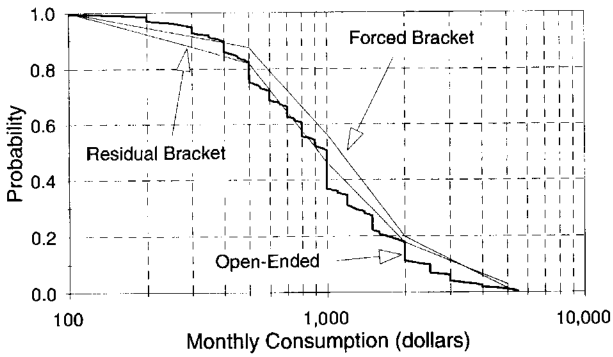


Fig. 8.13 Consumption CCDF for alternative elicitation formats: all responses

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 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 8.13 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 nonrespondents 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, 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 for whether the respondent handles the family finances, cognitive impairment, age, marital status, high school graduate, some college, and dummy variables for 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 whether the coefficient on the inverse Mills ratio variable is zero; this can be done using a conventional *T*-test. Table 8.5 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 some 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 ac-

Table 8.5 Consumption: Selection in Open-Ended Responses

	Probit Model (Dependent Variable: Whether Open-Ended Response)		Regression Model (Dependent Variable: Consumption)	
Explanatory Variable	Estimate	Standard Error	Estimate	Standard Error
Constant	0.084	0.258	−2,458.6	3,434.6
Sex	−0.067	0.039	27.5	145.7
Rfinance	0.165	0.047	−369.5	335.5
Cognition	−0.361	0.044	531.7	655.7
Age	−0.011	0.003	19.5	21.9
Married	0.045	0.046	304.8	120.4
High school graduate	0.075	0.045	−119.0	160.8
Some college	0.119	0.049	199.5	250.9
Wealth quartile 2	−0.038	0.048	141.7	102.5
Wealth quartile 3	−0.062	0.049	284.8	138.2
Wealth quartile 4	−0.156	0.056	825.5	319.0
Inverse Mills ratio			4696.8	4589.2
Log likelihood or R^2	−3,319.08		0.179	
No. of observations	6,722		1,368	

counts 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 Mills 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 that 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 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 8.6 gives the results for three alternative models. The first two models use the least squares parameters β 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; 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

Table 8.6 **Empirical Prior Imperfect Discrimination Model of Consumption: AHEAD Forced and Residual Bracket Responses**

	Model 1		Model 2		Model 3	
Parameter	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Lambda 1	0.687	0.042	0.685	0.042	0.487	0.042
Lambda 2	2.053[a]	0.127	1.505	0.147	1.040	0.091
Lambda 3	2.053[a]	0.127	2.610[b]	0.233	1.844	0.288
Lambda 4 and lambda 5	2.053[a]	0.127	2.610[b]	0.233	3.447	0.639
Alpha	0.268	0.025	0.257	0.025	−0.064	0.284
Married					0.353	0.048
Head some college					0.294	0.053
Spouse some college					0.394	0.052
Home owner					0.013	0.056
Head age					−0.012	0.004
Log likelihood	−4,180.243		−4,174.682		−3,991.677	

Note: The log likelihood with perfect discrimination (same as model 1 with lambdas constrained to be infinite) is −4,807.911

[a]Model 1 parameters constrained to be the same.

[b]Model 2 parameters constrained to be the same.

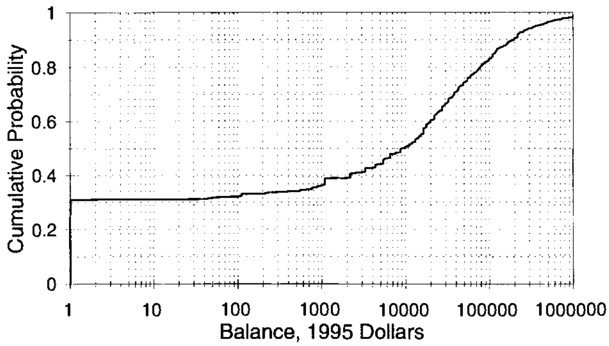


Fig. 8.14 Savings balances in 1989 SCF

better than model 2. Also, a likelihood ratio test shows that model 1 is much better than a model without anchoring errors, that is, with perfect discrimination.[6] *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.*

8.5 Savings

8.5.1 Savings Balances

The 1991 median net worth of households aged 70 or over, in 1995 prices, was $92,609, according to the U.S. Bureau of the Census (1994, table G). Savings balances, or interest-earning assets at financial institutions, represent an important component of the net worth of the elderly. The same source estimates that the distribution of net worth of households aged 65 or over 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 Keogh accounts, and 9.4 percent stocks and mutual funds. The 1989 SCF contains 625 households aged 70 or over. In this population, 31 percent had zero savings balances, and the distribution is highly skewed. Expressed in 1995 prices, the median savings balance is $9,130, the mean is $88,881, and the standard deviation is $299,920. Figure 8.14 shows the CDF for savings balances in the SCF, and figure 8.15 shows the Lorenz curve for this asset.

Table 8.7 provides some summary statistics on savings, based on the SCF. Table 8.8 gives a probit model for having positive savings, as a function of demographic variables, and a regression of log savings on demographic variables, for the subpopulation with positive savings, with an inverse Mills ratio

6. The log likelihood for the perfect discrimination model is −4,807.91, so that the likelihood ratio test statistic for model 1 against the perfect discrimination model is 1,255, with 2 degrees of freedom.

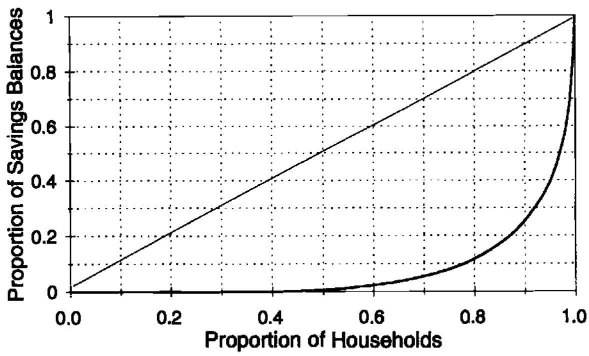


Fig. 8.15 Lorenz curve for savings in 1989 SCF

Table 8.7 Savings Balances Summary Statistics

Variable	Coding	Mean	Standard Deviation
Positive savings balance	1 = Yes, 0 = No	0.688	0.444
Savings balances	1995$	88,880	299,919
Head age	Years	74.522	5.784
Spouse age	Years	69.936	7.241
Married	1 = Yes, 0 = No	0.571	0.444
Head education	Years	11.840	3.819
Spouse education	Years	0.331	0.471
Home owner	1 = Yes, 0 = No	0.731	0.444
Head health			
Good	1 = Yes, 0 = No	0.376	0.485
Fair	1 = Yes, 0 = No	0.276	0.448
Poor	1 = Yes, 0 = No	0.122	0.327
Spouse health			
Good	1 = Yes, 0 = No	0.406	0.492
Fair	1 = Yes, 0 = No	0.235	0.424
Poor	1 = Yes, 0 = No	0.086	0.280
No spouse	1 = Yes, 0 = No	0.421	0.494
Couples			
Positive savings balance		0.754	0.431
Savings balances		133,539	384,681
Home owner		0.843	0.365
No Spouse			
Positive savings balance		0.597	0.491
Savings balances		27,411	61,569
Home owner		0.578	0.495

Source: 1989 Survey of Consumer Finances.

Table 8.8 **Demographics and Savings Balances**

	Probit Model (Dependent Variable: Positive Savings Balance)		OLS on Positive Subsample (Dependent Variable: Log Savings Balances)	
Variable	Coefficient	Standard Error	Coefficient	Standard Error
Constant	−1.246	1.069	7.611	1.947
Head age	0.007	0.001	−0.006	0.036
Pos(Head age − 80)[a]	−0.027	0.040	0.049	0.122
Married	0.266	0.119	0.081	1.075
Education	0.067	0.016	0.049	0.264
Home owner	0.205	0.128	−0.049	0.835
Head health				
Good	0.169	0.054	−0.881	2.044
Fair	0.252	0.062	−0.578	0.383
Poor	0.078	0.074	−0.801	0.579
Inverse Mills ratio			2.269	5.365
Log likelihood or R^2	−343.750		0.155	

Source: Fall 1994 Consumer Expenditure Survey.
[a]$\text{Pos}(x) = \max(0,x)$.

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 8.16 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.

8.5.2 AHEAD Data

Recall that Figures 8.1 through 8.7 describe each of the savings gate designs. These designs forced a bracket response for subjects who indicated in a preliminary question that they had positive savings. The starting gate values were \$1,000, \$2,000, \$5,000, \$10,000, \$20,000, \$50,000, and \$200,000. Table 8.9 gives the sample sizes for each treatment. A total of 4,855 subjects were admin-

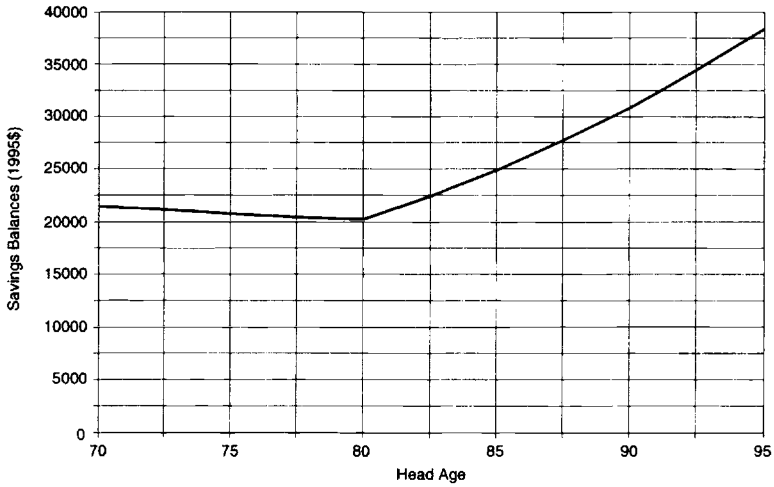


Fig. 8.16 Savings balances vs. age in 1989 SCF

istered the module containing the unfolding bracket 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 nonresponse. 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 analyses of the savings data are carried out in terms of the log of savings balances and logs of gate amounts. The parametric analysis assumes that savings is lognormal, 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 8.17 shows the CCDFs under the different treatments computed for all observations, including partial bracket response. There are economically significant anchoring effects, with the CCDFs from higher starting values showing substantially higher means and medians than those from lower start-

Table 8.9 **Savings: Sample Sizes, Medians, and Means**

Treatment	Starting Gate Amount	Sample Size	Number with Positive Savings[a]	Percentage of Bracket Responses Completed	Medians: Non-parametric[b]	S.E.[c]	Parametric[d]	S.E.[e]	Means: Non-parametric[f]	S.E.[g]	Parametric[h]	S.E.[i]
1	1,000	739	511	91.4	11,750	1,313	11,894	1,191	58,429	5,274	57,438	7,366
2	2,000	689	479	85.8	12,453	1,384	12,239	1,220	50,259	4,455	48,808	5,995
3	5,000	627	425	89.4	11,626	1,130	11,644	1,266	54,482	5,463	52,087	7,143
4	10,000	782	543	88.0	19,145	1,857	16,048	1,417	57,640	4,617	54,865	5,773
5	20,000	707	492	85.8	19,759	2,517	16,129	1,714	72,609	6,215	79,044	10,786
6	50,000	594	416	90.4	24,670	3,212	23,583	2,463	88,143	6,919	91,018	11,628
7	200,000	7.7	511	88.1	19,490	2,504	17,795	1,787	79,898	6,514	80,690	10,246
Overall		4,855	3,377	88.4	16,206	669	15,107	582	65,797	2,156	66,011	3,190

Completed Log Likelihoods	Semiparametric	DF	Normal	DF
No anchoring	−6,202.6	7	−6,216.8	2
Saturated	−6,123.0	49	−6,183.8	14
Imperfect discrimination	−6,188.1	9	−6,213.2	4

[a]Subjects were first asked whether they had positive savings, and affirmative respondents were then presented with unfolding bracket questions.

[b]Exponential of linearly interpolated CCDF of log saving, with the CCDF estimated using a "saturated" multinational model for all respondents.

[c]Standard error is estimated by (median) × $(a - b)/(2 \times$ (prob. of bracket) × (root N)), where (b,a) is the log savings bracket containing the estimator. This estimator assumes that log savings is uniformly distributed within the bracket containing the median.

[d]Exponential of the mean of a log normal distribution fitted by maximum likelihood estimation to bracket frequencies of log savings.

[e]Standard error is estimated by (median) × (SD) × root(2 × pi/N), where SD is the estimated standard error of log savings.

[f]Sum of bracket midpoints times estimated bracket probabilities.

[g]Standard error is estimated by square root of (sum of squared bracket midpoints times bracket probabilities minus median squared)/N.

[h]Exponential of (mean) + 0.5 × (sigma)^2, where mean and sigma are estimates of the mean and standard deviation of log savings.

[i]Standard error is estimated by (mean) × (SD) × root(1 + 0.5 × (SD)^2/(root N), where SD is the estimated standard error of log savings.

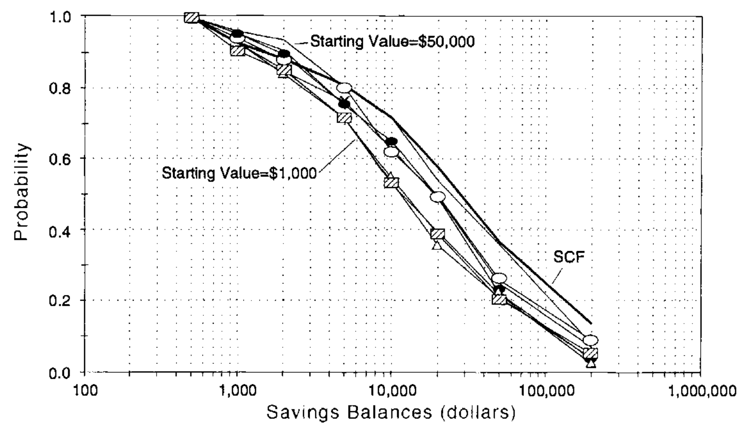


Fig. 8.17 Savings balance CCDF for AHEAD by starting value: all responses
Note: Starting values are $1,000 (*open inverted triangle*), $2,000 (*open triangle*), $5,000 (*open square*), $10,000 (*open circle*), $20,000 (*filled inverted triangle*), $50,000 (*filled triangle*), and $200,000 (*lined square*). Heavy line graphs SCF.

ing values.[7] However, the effects of anchoring are not uniform. The starting values of $1,000, $2,000, and $5,000 yield very similar CCDFs, and the starting value of $200,000 does not exert as much pull to the right as the starting value of $50,000. Table 8.9 gives the medians and means of these distributions; the pattern is consistent with that suggested by examining the CCDFs: 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.[8] A regression of the nonparametric mean of log savings on log starting value, weighted 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 lognormal savings model without anchoring produces mean and median estimates that are qualitatively similar to its nonparametric coun-

7. 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 they 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.

8. The completed log likelihood for the saturated model is −6,122.99, and the completed log likelihood for a multinomial model with no anchoring is −6,202.59. Then the likelihood ratio statistic is 159.19 with 42 degrees of freedom.

Table 8.10 Empirical Prior Imperfect Discrimination Model of Savings Balances: AHEAD Forced Bracket Responses for Subjects with Positive Savings

	Model 1		Model 2		Model 3	
Parameter	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Lambda 1	3.643	0.228	3.641	0.228	2.827	0.182
Lambda 2	10.273	1.820	10.280	1.823	8.499	1.129
Lambda 3	6.570[a]	0.918	6.611	1.030	6.113[b]	0.863
Lambda 4 and lambda 5	6.570[a]	0.918	6.481	1.499	6.113[b]	0.863
Alpha	−1.057	0.092	−1.058	0.092	−2.511	0.341
Married					0.893	0.161
Head education (years)					0.141	0.028
Spouse education (years)					0.143	0.027
Poor health					−1.979	0.245
Log likelihood	−6,293.610		−6,293.608		−6,177.115	

Note: The log likelihood with perfect discrimination (same as model 1 with lambdas constrained to be infinite) is −6,521.557.

[a]Model 1 coefficients constrained to be the same.

[b]Model 3 coefficients constrained to be the same.

terpart; however, the lognormal parametric specification is rejected using a likelihood ratio test.[9]

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 8.10. Discrimination errors are highest at the first gate (λ_1 lowest), with no significant variation in the λ_k for successive gates (model 2 vs. model 1). In general, the λs 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 the 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, indicat-

9. The completed log likelihood for the lognormal model without anchoring is −6,216.79, and the likelihood ratio statistic for the lognormal vs. 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 lognormal specification is correct is rejected; the likelihood ratio statistic is 121.6 with 35 degrees of freedom.

ing 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.

8.6 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 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 $5,000 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. Thus, variation in unfolding bracket gates, in tandem with the discrimination model or an alternative model of anchoring, promises 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 that they exercise caution in imputing economic variables based on stated brackets.

References

Bohm, P. 1972. Estimating willingness to pay: An experiment. *European Economic Review* 3:111–30.

Coupey, E. 1994. Restructuring: Constructive processing of information displays in consumer choice. *Journal of Consumer Research* 21:83–89.

Cowing, T., J. Dubin, and D. McFadden. 1982. The NIECS database and its use in residential energy demand modelling. MIT Energy Laboratory Discussion Paper no. MIT-EL-82-041 WP. Cambridge: Massachusetts Institute of Technology.

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,* forthcoming.

Hobbs, F. 1996. *65+ in the United States.* Current Population Report no. P23-190. Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

Horowitz, J., and G. Neumann. 1989. Computational and statistical efficiency of semiparametric GLS estimators. *Econometric Reviews* 8:223–35.

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* 21:1161–66.

Kahneman, D., and A. Tversky. 1979. Intuitive prediction: Biases and corrective proce-

dures. *Studies in Management Science* 12:313–27.
Lee, L. 1995. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics* 65:381–428.
Lewbel, A. 1997. Semiparametric latent variable model estimation with endogenous or mismeasured regressors. Waltham, Mass.: Brandeis University. Working paper.
Lewbel, A., and D. McFadden. 1997. Estimating features of a conditional distribution from binomial data. Berkeley: University of California. Working paper.
Lowenstein, G. 1988. Frames of mind in intertemporal choice. *Management Science* 34:200–214.
McFadden, D. 1994. Contingent valuation and social choice. *American Journal of Agricultural Economics* 76:689–708.
Northcraft, G., and M. Neale. 1987. Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39:84–97.
Poterba, J., and L. Summers. 1986. Reporting errors and labor market dynamics. *Econometrica* 54:1319–38.
Rabin, M. 1998. Psychology and economics. *Journal of Economic Literature,* forthcoming.
Smith, J. 1995. Racial and ethnic differences in wealth in the Health and Retirement Survey. *Journal of Human Resources* 30 (Suppl.): S158–S183.
Smith, V. 1979. An experimental comparison of three public good decision mechanisms. *Scandinavian Journal of Economics* 81:198–215.
Svensson, O. 1996. On the modeling of human choices in descriptive behavioral decision theory. Working paper, Stockholm University.
Thurstone, L. 1927. A law of comparative judgment. *Psychological Review* 34:273–86.
Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31.
U.S. Bureau of the Census. 1994. *Household wealth and asset ownership.* Current Population Report no. P70-34. Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

Comment James P. Smith

Let me go straight to the bottom line. This is a very good paper. It deals with an important problem, has the appropriate combination of technique and substance, is completely convincing in its main conclusion, and is constructive in offering remedies.

The paper argues (and I believe proves) that serious anchoring effects exists when household surveys use follow-up bracket questions after initial nonresponse to economic questions. Follow-up brackets are a sequence of "more than x or less than y" questions offered to respondents who initially refused or were unable to provide an exact value for, say, their assets or their income. Anchoring occurs when the content of the question itself conveys information about what the probable "correct" answer is. For example, if respondents were

James P. Smith is a senior economist at RAND.
This work was supported by grants 5P01-AG08291 and 1R01-AG12394 awarded by the National Institute on Aging.

asked about the size of their checking accounts, responses may be influenced by whether the first question was set at the $100 level, $1,000 level, or $10,000 level, even if the final set of categories offered will eventually be identical. Since respondents may assume that question designers know more than they do, the entry point may tell respondents something about what the "correct" answer is.

We know from recent research that unfolding brackets are an important survey research tool that can substantially reduce item nonresponse to economic questions. They also significantly improve estimates of missing values of respondent assets (see Juster and Smith 1997). For example, Juster-Smith report that item nonresponse for asset questions is reduced by almost 75 percent and estimates of mean nonhousing wealth increased by 18 percent by the use of brackets in the Health and Retirement Survey.

This paper takes this survey technology a step further by asking whether estimates of missing values are affected by the placement of the initial entry point in the bracket sequence. That is, even if the final set of bracket categories are the same, it may matter if we start respondents off at a very low entry number or a very high one. The reason the authors can test this question is that an experiment was performed in the survey of Asset and Health Dynamics among the Oldest Old (AHEAD) whereby respondents were randomly assigned different entry points to questions on monthly consumption and savings account balances. Although the entry points varied (randomly), all respondents eventually were presented the same set of final bracket categories for both questions. For the consumption measure, there were four different initial entry points: $500, $1,000, $2,000, and $5,000. After going through the full bracket sequence, all respondents will have been offered the option of placing their unknown consumption into the same five bracket categories: 0–$500, $501–$1,000, $1,001–$2,000, $2,001–$5,000, and more than $5,000. Similarly, there were seven initial entry points for savings accounts: 1, 2, 5, 10, 20, 50, 200 (all in thousands of dollars), but all respondents eventually were provided the same eight bracket categories in which their unknown values could be placed.

The evidence presented that entry points do in fact matter is overwhelming. Table 8C.1 summarizes this evidence by listing the authors' nonparametric estimates of how median values vary with initial entry point for the two measures. For example, their estimate of median household consumption is $895 when the initial entry point is $500 and $1,455 when the initial entry point was $5,000. This range represents a 62 percent difference, which is truly scary for those of us who fret about the quality of economic data. With tongue firmly in cheek, I cannot resist suggesting that a simple cure for high measured poverty is simply to use high entry points in bracket sequences on income. Similarly, the authors' estimate of median savings balances are $11,750 when the lowest entry point of $1,000 was used but rises to $19,590 for respondents who received the highest entry bid of $200,000 (a 66 percent range). The simplicity and beauty of their test is that, since respondents were selected randomly with

Table 8C.1 **Estimated Median Values by Initial Entry Point**

Consumption ($)		Savings ($)	
Entry Bracket	Median	Entry Bracket	Median
500	895	1,000	11,750
1,000	1,146	2,000	12,453
2,000	1,415	5,000	11,626
5,000	1,455	10,000	19,145
		20,000	19,759
		50,000	24,670
		200,000	19,490

respect to entry points, there should be no systematic differences across entry bids. That is obviously not the case.

In light of how persuasive their case is that entry point brackets matter, what are the remaining issues? There are two critical ones. First, how much can we and should we generalize from this evidence to other measures of economic well-being? Second, what can we do about the problem?

How far should we generalize to other measures of economic status? Maybe not too much, since these two items—consumption and savings accounts—probably represent worst-case scenarios. Total consumption and savings accounts are among the most difficult to measure economic constructs conceptually. This inherent difficulty in measurement is compounded by some quite imperfectly worded questions in surveys. For these reasons, consumption or savings accounts may exaggerate the extent of the problem.

For example, consumption is typically measured in economic surveys such as the Consumer Expenditure Survey (CES). These consumption surveys are lengthy, detailed, time consuming, and often involve the use of household diaries, and they appropriately worry about a host of thorny problems—the periodicity of measurement, how to treat consumer durables, and the jointness of many consumption items. By contrast, measurement of consumption in AHEAD relies on a single question. And consider the precise wording of the AHEAD consumption question.

> 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 expenses such as food, clothing, transportation, entertainment, and other expenses you and your household may have.

Little wonder then that AHEAD respondents were more than a little unsure of what the interviewer wanted to know and what the answer would be even if they did know. Such vagueness and uncertainly makes respondents particularly sensitive to any clues (including entry bids) that they might obtain from the interviewer. One piece of evidence about the severity of this problem is that mean consumption in AHEAD is only 72 percent as high as that measured in the CES. AHEAD respondents apparently significantly understate total consumption. This understatement may also be partly a consequence of the limited set of items listed in the question after the phrase "such as food." For example, respondents' answers may have been different and larger if medical care was added to the list.

The conceptual and wording problems are just as severe with the savings account question. "Savings account" is an old-fashioned and perhaps outmoded term. Quite frankly, if I were the respondent, I would have no idea what the interviewer was asking. Savings accounts used to be interest-bearing accounts that were distinguished from checking accounts (on which one could write checks). But in today's world of interest-earning checking accounts, the distinction may have lost much of its original meaning. Similarly, the precise wording of the question on savings is of little help.

> I have a few more questions about how people are getting along financially these days. First, do you have any money in SAVINGS ACCOUNTS?

Alongside this question wording, there exists an instruction to interviewers to exclude checking accounts, money markets, mutual funds, and so forth. This may represent another situation in which most respondents are very unsure of exactly what question is being posed. Even if respondents knew what the question meant, they may not have much confidence that they actually know the correct (e.g., most accurate) answer. Once again, such vagueness maximizes the likelihood that respondents will be unusually sensitive to any clues or hints given by the interviewer to help answer the question. I believe that other constructs that economists care about—such as education, income, or even specific asset categories (stocks, house values)—would be less susceptible to the anchoring phenomenon. But to be fair, they are unlikely to be immune from it. Additional testing using the methodology spelled out in this paper should be a high priority for these other central economic constructs.

Finally, what can we do about the problem? Here the paper is quite unusual in not simply pointing out a serious survey methodological problem but also offering a constructive suggestion. The authors argue that the type of random variation in anchoring entry points used in the AHEAD experimental module should be a standard part of all surveys that rely on follow-up brackets. Experimental variation in entry gates allows researchers to statistically identify the biases caused by anchoring and to correct their parametric estimates for these biases.

I think their idea is excellent, but I would take it one step further. A respondent's answer or estimate of the unknown economic value (C^*) is a function of both his own prior beliefs (C_R^*) and any information advertently or inadvertently provided by interviews (C_I^*). In a simple linearized expression of this idea, we can write

$$C^* = a + bC^*_R + e(C^*_I - C^*_R) + f(C^*_I - C^*_R)^2.$$

In this formulation, if interviews (entry points) do not matter, then $e = f = 0$. Interviews will matter only if they provide departures from respondents' initial beliefs in some way, but big departures will probably get less weight ($f < 0$). The notion that large departures may get less weight receives some support from the savings column in table 8C.1. The big difference in median savings appears to depend only on whether the initial entry bid was higher or lower than $10,000, with little variation among bids below or above that threshold. Most people do not have accounts of $200,000, and they may not take seriously entry bids that start that high.

This formulation also suggests that the same set of variations in initial entry points may not be optimal for all respondents. If one could only use a single entry point, values close to the population median may minimize variances in estimation errors. But with modern survey methods, there is no reason not to center the variation in initial entry points around an individual respondent's expected value. To minimize mean square errors, this suggests that the set of random entry points chosen should vary across respondents. For example, the set of entry points chosen should be lower for respondents with asset values below the population median than for respondents with asset values above the median.

How can we know a priori which respondents might have low or high asset values? We must remember that these new economic wealth surveys are longitudinal in nature so that we have considerable information from prior waves about where individual respondents are likely to lie in the distribution. Variation across respondents in the set of entry points can be programmed into the computer-assisted telephone or personal interview technology.

Reference

Juster, F. Thomas, and James P. Smith. 1997. Improving the quality of economic data: Lessons form the HRS and AHEAD. *Journal of the American Statistical Association* 92(440):1268–78.

This Page Intentionally Left Blank