Tail and Center Rounding of Probabilistic Expectations in the Health and Retirement Study[1],[2]

Pamela Giustinelli[3]
Department of Economics
Bocconi University

Charles F. Manski
Department of Economics
Northwestern University

Francesca Molinari
Department of Economics
Cornell University

Abstract

We study rounding of numerical expectations in the Health and Retirement Study (HRS) between 2002 and 2014. We document that respondent-specific rounding patterns across questions in individual waves are quite stable across waves. We discover a tendency by about half of the respondents to provide more refined responses in the tails of the 0-100 scale than the center. In contrast, only about five percent of the respondents give more refined responses in the center than the tails. We find that respondents tend to report the values 25 and 75 more frequently than other values ending in 5. We also find that rounding practices vary somewhat across question domains and respondent characteristics. We propose an inferential approach that assumes stability of response tendencies across questions and waves to infer person-specific rounding in each question domain and scale segment and that replaces each point-response with an interval representing the range of possible values of the true latent belief. Using expectations from the 2016 wave of the HRS, we validate our approach. To demonstrate the consequences of rounding on inference, we compare best-predictor estimates from face-value expectations with those implied by our intervals.

Keywords: Interval data; Subjective Probabilities; Survey data.
JEL Codes: D80, D84, C83.

[1] **Acknowledgements:** We thank Wilbert Van der Klaauw and two anonymous referees for comments that helped us significantly improve the paper, and Maura Coughlin, Adam Karabatakis, Miriam Larson-Koester, and Qihong Ruan for able research assistance. We received useful feedback from seminar participants at the HRS work-in-progress series, Bocconi University, University of Southampton, NYU CUSP, University of Michigan, Purdue University, Laval University, University of Oslo, Statistics Norway, University of Munich, and University of Padova, as well as from participants in the 2016 NYFed and ESRC RCMiSoC Workshop on Subjective Expectations and in the 2018 CESifo and ESRC RCMiSoC Workshop on Subjective Expectations and Probabilities in Economics. Part of this research was conducted while Molinari was on sabbatical leave at the Department of Economics at Duke University, whose hospitality she gratefully acknowledges.
[2] **Funding:** This work was supported by: the National Institute on Aging [grant numbers NIA P01 AG10179, P30 AG012846 for Giustinelli]; the National Science Foundation [grant numbers SES 1131500 through the University of Michigan node of the NSF-Census Research Network for Giustinelli, SES 1129475 for Manski, SES 1824575 for Molinari], and the Michigan Institute for Teaching and Research in Economics (MITRE)'s Undergraduate Student Support program (for Giustinelli).
[3] **Contact:** Pamela Giustinelli, Department of Economics, Bocconi University, Via Roentgen 1, 20136, Milan, Italy. Phone: +39 02 5836 3413. E-mail: pamela.giustinelli@unibocconi.it.

1. Introduction

Judgements about the likelihood of future events are an important input for predictions and decisions by citizens, policy makers, and researchers. From the early 1990s on, surveys designed by economists have increasingly measured respondents' subjective expectations for future events using a 0-100 scale of percent chance.

Research using numerical survey expectations has devoted substantial effort to evaluating how persons respond to the questions posed in specific domains. Manski (2004, 2018), Attanasio (2009), Hurd (2009), van der Klaauw (2012), Armantier *et al.* (2013), Delavande (2014), Schotter and Trevino (2014), Giustinelli and Manski (2018), and Altig et al. (2019) review the large literature from various perspectives.

Questions eliciting expectations on a 0-100 percent-chance scale enable respondents to report beliefs to the nearest 1 percent. But how do respondents use the scale in practice? The accumulated evidence reveals that respondents tend to round their responses. Responses that are not a multiple of 5 or 10 percent occur infrequently. When observed, they tend to occur near the endpoints of the scale to convey very small or large probabilities.

Rounding of survey expectations and other survey data poses a series of challenges for inference. First, rounding generates measurement error. The measurement error induced by rounding is not the classical type where observed data equal true values plus white noise. Hence, econometric analysis based on the familiar errors-in-variables framework is not appropriate. Indeed, no general theoretical predictions can be made on the direction of the bias of estimates obtained using rounded data.

Second, the extent of rounding underlying each response is not directly observable and may vary across respondents and questions. Surveys do not instruct respondents as to what degree of rounding they should apply when answering specific questions; hence, there may not exist consensus rounding norms. Respondents may vary in their rounding practices, which are unknown to data users.

Third, the reasons why respondents round when reporting numerical expectations are incompletely understood. Manski and Molinari (2010) hypothesize that respondents may round to simplify communication and/or to convey partial knowledge. Giustinelli, Manski, and Molinari (2020) design and analyze question sequences that substantiate both reasons for rounding.

Observed response patterns across questions carry information about respondents' rounding practices, although they do not reveal why respondents round. Manski and Molinari (2010) studied respondent-specific response patterns across all expectations questions asked in the 2006 wave of the Health and Retirement Study (HRS). They found strong evidence of rounding, with the extent differing across respondents. They proposed use of a person's response pattern across questions to infer the person's rounding practice, the result being interpretation of reported numerical values as interval data.

In this paper, we significantly expand study of respondent-specific rounding patterns by analyzing responses across all expectations questions asked in the core HRS questionnaire between 2002 and 2014. This enables us to learn important new features of rounding practices.

Section 2 explains the basic themes of our approach to analysis of rounding. Whereas it is common for researchers to focus on a survey item of interest and study responses to this question across respondents, we propose to study each respondent's answers to the entire set of probabilistic expectations questions she was surveyed about. For each respondent, we use the response pattern across these multiple questions to infer the extent to which the respondent rounds responses to specific questions. The key assumption is stability of respondent-specific rounding across responses.

The remainder of the paper applies the approach to the HRS. Users of HRS data should find this application of direct interest. Others should find it helpful as a case study of the general framework presented in Section 2 providing lessons for applications to other datasets.

Section 3 presents the main findings of our data analysis, with Supplementary Appendices reporting further details. We document that respondent-specific rounding patterns across questions in individual waves are quite stable across waves. We discover a tendency by about half of the respondents to provide more refined responses in the tails of the 0-100 scale than the center. In contrast, only about five percent of the respondents give more refined responses in the center than the tails. We find that respondents tend to report the values 25 and 75 more frequently than other values ending in 5. We also find that rounding practices vary somewhat across question domains, which range in the HRS from personal health to personal finances to macroeconomic events.

Based on our examination of rounding practices in Section 3, Section 4 specifies and implements a version of the general framework introduced in Section 2, featuring stability of rounding across waves and heterogeneous rounding across scale locations and question domains.

Our framework interprets each numerical response given by a respondent as an interval and has a two-stage structure. The first stage classifies each respondent into one of a set of mutually exclusive and exhaustive rounding types and places an upper bound on the amount of rounding each respondent is inferred to apply when reporting their expectations. The second stage assigns an interval to each of the respondent's original point responses, which represents the range of values in which the respondent's underlying true belief is plausibly deemed to lie based on the respondent's inferred rounding type.

Our approach accommodates substantial heterogeneity in rounding practices. Within a specific question domain, a respondent's rounding type is a bivariate vector of the form (tail, center) rounding, partitioning the 0-100 scale into two symmetric tails (0-24 and 76-100) and a center (25-75). Thus, in addition to being person specific, the inferred degree of rounding may differ between tails and center and may vary across question domains. The assigned intervals vary across respondents and across values of the observed point responses.

We use our framework to study how rounding tendencies vary with observable characteristics of the respondents. We find that higher levels of educational attainment and of cognition are associated with a tendency to give more refined responses (less rounding) across all scale segments and question domains. We also examine how rounding tendencies vary across cohorts of HRS respondents.

We use survival expectations, working expectations, and stock market expectations observed in the 2016 wave of the HRS to validate our approach. Specifically, for each of these three questions and for each scale segment we compute the proportion of respondents who gave a response in 2016 that is consistent with the respondent's rounding type in the relevant domain and scale segment predicted by our algorithm on the basis of the 2002-2014 data. We find that 93.39% of tail responses and 88.25% of center responses to the survival question in 2016 are consistent with the predictions generated by our algorithm based on the 2002-2014 data. We observe similar (or slightly higher) fractions of valid cases for the working and stock market questions.

While this paper studies rounding as a subject of intrinsic interest, a reader may naturally ask how the interval data that our proposed approach generates from point responses may be used in statistical analyses. This matter has been addressed in the econometric literature studying conditional prediction with interval measurement of outcomes and/or covariates; see Manski and Tamer (2002) and Beresteanu and Molinari (2008).

Section 5 demonstrates how interval data on subjective expectations can be employed as either an outcome variable or a covariate in prediction analyses of substantive interest, and how exclusion restrictions can be incorporated into the analysis to sharpen the inference. One application considers best linear prediction of the labor supply expectations of working HRS respondents, conditional on specified covariates. A second application uses longevity expectations and other covariates to predict hours worked. Both applications bring to bear partial identification analysis. This analysis enables one to bound the maximum bias of point predictions made under the conventional assumption that persons do not round responses.

As far as we are aware, only two previous papers systematically study rounding of responses to probabilistic expectations questions. One is Manski and Molinari (2010), on whose work we build. The other is Kleinjans and van Soest (2014), who develop and estimate a panel-data structural econometric model to analyze response patterns to each of six expectations questions in the HRS. Their analysis aims to investigate the extent to which probability reports are determined by genuine underlying probabilistic beliefs, rounding, a tendency to give so-called "focal" responses of (0, 50, 100), and selective item non-response. Despite the very different approaches taken, they find, as we do, that tendencies to round, give "focal" responses, and not respond tend to be persistent over time.

Some authors have devoted special attention to responses of 0, 50, and 100 percent. Fischhoff and Bruine de Bruin (1999) and Bruine de Bruin *et al*. (2002) hypothesize that some respondents use 50 percent to signal epistemic uncertainty. Lillard and Willis (2001) and Hudomiet and Willis (2013) conjecture that respondents form full subjective distributions for the probability of an event and then report whichever of the values (0, 50, 100) is closest to the mode of their distribution. We analyze each individual's reports of (0, 50, 100) percent jointly with that individual's responses to the entire set of expectations questions asked, and we find that a small fraction of respondents in any wave give responses that are exclusively in the (0, 50, 100) group (between 2% and 4%, see Appendix Table 1).

Some researchers have investigated how respondents' propensity to give "focal" responses to numerical expectation questions varies with the survey technology (e.g., with survey mode, features of the percent-chance scale, etc.). For example, in an online survey with a nationally representative sample of the Dutch population, Bruine de Bruin and Carman (2018) found that elicitation of percent-chance expectations using a visual linear scale with a clickable slider significantly reduced the use of "focal" responses relative to a more traditional open-ended mode, without affecting the predictive validity of responses and survey satisfaction of respondents. In this paper, we use numerical expectations elicited in the HRS using computer-assisted, in-person or phone, interviews by means of traditional open-ended percent-chance questions.

Beyond readers who have interest in expectations data, we anticipate that general survey researchers will find this paper useful. Our study of tendencies to round responses to expectations questions should heighten concern that respondents may round responses to numerical questions in other contexts. Consider, for example, questions asking respondents to state their income or the number of hours they worked in the past week. Respondents may round their responses, with the extent of rounding differing across persons. Examination of a person's response pattern across different numerical questions, in the manner that we do here, may provide a credible way to infer that person's rounding practice. One may then interpret reported numerical values as intervals.

2. An Inferential Framework for Person-Specific Rounding: Combining Stability and Heterogeneity

This section introduces a framework for inferring person-specific rounding in survey reports of percent-chance expectations. Our approach is most directly applicable to respondents who have precise point probabilities and who round their reports to simplify communication rather than to convey partial knowledge. It is also applicable in some cases where respondents have partial knowledge and hold interval rather than precise subjective probabilities for uncertain events. Then the algorithm developed in this paper still yields an interval that encloses a respondent's true beliefs if the algorithm interval completely includes the latent interval.

When rounding reflects a desire to simplify communication, we think it credible to view rounding as a person-specific trait that is stable across multiple responses by a given person. We think it desirable to allow for unrestricted heterogeneity across individuals in their rounding practices. Persons may vary in the manner in which they communicate their subjective beliefs in surveys.

For each respondent, we use the response pattern across multiple questions to infer the extent to which the respondent rounds responses to particular questions. For example, consider a respondent *A* who is asked three expectations questions, which (s)he answers with 44 percent, 35 percent, and 70 percent. By answering 44 percent to one of the three questions, respondent *A* reveals that (s)he rounded to the nearest 1% when answering that question. Under the assumption that person-specific rounding is stable across these questions, one can further infer that *A* rounded to the nearest 1% also when answering the other two questions.

Consider now a respondent B who is asked the same three questions, which (s)he answers with 40 percent, 35 percent, and 70 percent. The most refined response is 35, a multiple of 5 that is not a multiple of 10. Observing that respondent B answered 35 to the second question, one might infer that respondent B rounded to the nearest 5% or by a finer degree to that question. Under the assumption of stability of person-specific rounding across these questions, one can further infer that respondent B rounded to the nearest 5% or by a finer degree to all three questions.

Generalizing beyond these examples, the most basic version of our inferential approach has these features:

(i) for each individual respondent, inspect the respondent's pattern of responses across multiple expectation questions;

(ii) use the most refined response among those considered in (i) to place an upper bound on the amount of rounding the respondent applied in all responses in (i);

(iii) replace each point response considered in (i) with an interval that represents the range of possible values of the true latent belief and whose width depends on the rounding upper bound inferred in (ii).

A key data requirement for applicability of the approach is observability of multiple responses per respondent. The key assumption is stability of respondent-specific rounding across responses.

Here, as elsewhere, an applied researcher may feel comfortable using an assumption if one thinks it credible. The strongest form of person-specific stability assumes that a respondent's rounding practice is the same across all questions. However, one might not find such a strong assumption credible. Weaker forms of stability assume that the rounding practice is the same across specified groups of questions rather than across all questions. Such weaker forms of stability have less identifying power, but they may be more credible.

For example, a researcher may assume that the rounding practice is stable within a question domain but might vary across domains. Manski and Molinari (2010) used this type of stability assumption, considering questions on personal health, personal finance, and the macroeconomy to be distinct domains. Or a researcher may hypothesize that rounding practices vary with question complexity, as reasoned by Kleinjans and van Soest (2014). Or one may think that rounding of probability reports varies with the underlying probabilities of events. Heiss et al. (2019) propose that rounding may vary with the magnitudes of objective probabilities.

In this paper, we find evidence that rounding varies with the magnitudes of subjective probabilities, with greater rounding in the centers than the tails of subjective distributions. We use a uniform definition of tails and center across expectations questions, where the tails correspond to the 0-24 and 26-100 segments of the response scale and the center corresponds to the 25-75 segment. Respondents, however, might round in terms of what they think might be the normal probability range for a certain event. If so, a researcher might find more credible to adapt the definition of tails and center to the empirical support of responses to different expectations questions, by letting the partition into tails and center vary across groups of questions with differing empirical supports.

After grouping the questions answered by a given respondent according to the chosen classification criterion, one can apply the inferential approach described in (i)-(iii) separately to each group of questions. This refinement of the basic approach assumes stability of person-specific rounding within groups of questions, while allowing for heterogeneity of person-specific rounding across groups of questions.

Manski and Molinari (2010) applied the inferential approach, using all expectations reported by participants in the 2006 wave of the HRS. In this paper, we take advantage of the panel structure of the HRS. In addition to observing numerous expectations per respondent in each wave, we observe respondents' expectations over many waves. The richness of this data enables us to implement a version of the inferential approach that assumes stability of person-specific rounding across waves as well as within a wave.

Applicability of the approach in empirical work does not require that one has expectations data as rich as that collected by the HRS. For example, a single time-series of responses to the same question asked over time to the same respondent may suffice, if one is willing to maintain stability of rounding over time. The larger the number of responses observed per respondent, the greater the possibilities of use of the extra information.

3. Exploratory Analysis of Response Patterns across Questions and Waves in the HRS

Since 2002 the HRS has devoted Section P of its core questionnaire to measurement of expectations in the domains of personal health, personal finances, and general economic conditions. Across seven biannual waves spanning 2002 to 2014, expectations have been elicited on a 0-100 percent chance scale. Several questions have been repeated across multiple waves. Appendix Table S1 shows the list of questions organized by domain and the waves in which they were asked. For each question listed in Table S1, Appendix Table S2 reports the number of waves in which the question was asked, the number of respondents to which the question was asked, and the total number of observations for the question.

The number of questions per wave ranges between 22 in 2002 and 38 in 2006. Most questions are in the personal finances domain (between 11 and 23 per wave, 31 overall), followed by the personal health domain (between 3 and 9 per wave, 10 overall), and the domain of general economic conditions (between 2 and 7 per wave, 12 overall). A subset of 12 questions across the three domains were asked in all waves.

The number of responses varies across questions and waves, ranging from about 5,000 to 30,000 responses per question in each wave. The variation across questions stems from the fact that the HRS makes extensive use of skip sequencing. Thus, whether a respondent is asked a specific question depends on the previous answers given by the respondent and on whether the event specified by the question is relevant to the respondent.

The total number of responses generated by a question across the seven waves varies because questions have been added and removed over time. It also varies due to changes in sample composition across waves. The HRS sample has periodically been augmented with new cohorts of respondents who joined the study in specific waves. Respondents exit the study due to attrition or death.

3.1. Temporal Stability of Response Tendencies

In this Sub-section we study response patterns across questions in each wave, alternatively using all questions asked in the wave and the twelve questions asked in all waves. Focusing on the latter questions, we analyze the stability of response tendencies across pairs of waves. Supplementary Appendix SA2 provides further detail, investigating patterns of response to specific questions in Table S3. To ensure comparability with the analysis of the 2006 data by Manski and Molinari (2010), we condition also our analysis on respondents aged 50 or older.

We initially consider seven mutually exclusive and exhaustive response patterns, progressing from the most rounded to the least rounded. The first one (“All NR”) is for respondents who respond to no questions in the wave, coded in the HRS as “Don’t know” or “Refuse.” The second one (“All 0 or 100”) is for respondents who only use the values 0 and 100. The next one (“All 0, 50, or 100”) is for respondents who only use the values (0, 50, 100). The next two categories (“Some multiple of 10” and “Some multiple of 5”) are, respectively, for respondents who answer at least one question with a value that is a multiple of 10 other than (0, 50, 100), and a multiple of 5 that is not a multiple of 10. The final category (“Some 1-4 or 96-99) is for respondents who respond to at least one question with a non-round value in 1-4 or 96-99. “Some other” is a category for respondents who respond at least once with a non-round value in 6-94.

Table 1 reports the fraction of respondents for each wave of the survey whose responses fall into each of the categories just described, both when considering only the twelve questions that were asked across all waves (top panel), and when considering all questions asked in a wave (bottom panel). A detailed description of Table 1 is provided in the Supplementary Appendix (see Sub-section SA3.1). Here we emphasize the main message; the response patterns found by Manski and Molinari (2010) in the 2006 wave of the HRS hold throughout the seven waves between 2002 and 2014. However, these are aggregate patterns that may partly be susceptible to variation across waves in sample composition.

To address this issue, we compute transition matrices of response tendencies across waves. Specifically, for each pair of waves indicated by column, Table 2 reports the fractions of respondents classified as belonging to any rounding category in the first wave who transitioned to: the same rounding category in the second wave (1st row), a finer or coarser adjacent category (2nd row), and a more distant rounding category (3th row). The reported calculations use the twelve questions in common to the seven waves.

We find that between 0.406 and 0.436 of the respondents remain in the same rounding category across any pair of adjacent waves. Between 0.373 and 0.386 transition to an adjacent category. Thus, between 0.788 and 0.813 of respondents transitions to the same or an adjacent category. Even transitions between the first and last waves, with fourteen years separating them, display high persistence, with over 0.78 of the respondents transitioning to the same or an adjacent category.

3.2. Pooling Data across Waves to Probe More Deeply into Response Tendencies

With temporal stability established, we henceforth pool the HRS data across waves. This greatly increases the number of expectations responses observed per respondent, multiplying it sevenfold for respondents interviewed in all waves between 2002 and 2014. Across all questions and waves, the average number of responses per respondent is 106.8. By question domain, this figure ranges from 19.1 for personal health to 66 for personal finances. The complete figures are shown in Table S5 of the Supplementary Appendix.

With such rich respondent-specific data, we can probe more deeply into rounding practices. Specifically, we analyze response patterns separately by question domain, while paying particular attention to the location of responses inside the 0-100 scale. By so doing, we learn important features of respondents' response patterns in specific domains.

In order to investigate whether and, if so, how rounding practices vary across the measurement scale, we found it useful to partition the values of the 0-100 percent chance scale as following. We define the center (C) of the percent-chance scale to be values in the range 26-74 and the tails (T) to be values in the ranges 0-24 and 76-100. The values 25 and 75 form the boundary between the tail and center. We group responses into nine categories, defined by their presence in T or C and by their degree of granularity. The categories are: V1-T ≡ values in 1-24 or 76-99 that are not multiples of 5; V1-C ≡ values in 26-74 that are not multiples of 5; V5-T ≡ {5, 15, 85, 95}; V5-C ≡ {35, 45, 55, 65}; V10-T ≡ {10, 20, 80, 90}; V10-C ≡ {30, 40, 60, 70}; V25 ≡ {25, 75}; V100 ≡ {0, 100}; V50 ≡ {50}. The complete partition is shown in Table S6 of the Supplementary Appendix.

With this categorization, Table 3 shows the distribution of responses across respondents for the twelve questions asked in Section P in all waves. Table S7 in the Supplementary Appendix shows analogous statistics for all expectation questions asked between 2002 and 2014. Comparison of the frequencies of V25 responses (in column 5) with the frequencies of the remaining V5 responses (V5-C in column 9 and V5-T in column 8) reveals that the fraction of {25, 75} responses is always higher than the fraction of responses ending in 5 in the center of the scale (35, 45, 55, 65). For most questions across the three domains, the fraction of {25, 75} responses is higher than the fraction of responses ending in 5 in the tails of the scale (5, 15, 85, 95).

Even more striking is comparison of the frequencies of responses in the tails versus those in the center. The fractions of V10, V5, and V1 responses in the tails are higher than the corresponding fractions in the center for nearly all questions in Tables 2 and S7 (but P47 and P190).

It is important to note that this pattern is not driven by a generalized tendency of respondents to provide more responses in the tails than in the center. Aggregation of the statistics shown in Table S7 across response values in the center (columns 3, 5, 7, 9, and 11), and across response values in the tails (columns 4, 6, 8, and 10), reveals that the fractions of center responses and tail responses are quite balanced for nearly all expectations questions in the personal health and general economic conditions domains. Only in the domain of personal finances is there a tendency by respondents to give more responses in the tails than in the center. Low fractions of center responses seem to occur especially in questions asking the percent chance that the respondent will receive or leave an inheritance of a specified amount. For these questions the majority of observed responses are 0 or 100.

4. Transforming Expectations Responses into Interval Data

Generalizing the inferential approach proposed by Manski and Molinari (2010), this section develops a new algorithm that uses the response tendency of a respondent that we have documented in the previous sections to characterize rounding of responses to particular questions. The algorithm classifies each respondent into one of a set of mutually exclusive and exhaustive rounding types and transforms each original point response into an interval where the true latent belief is deemed to lie.

Our algorithm relies on considerably weaker and more credible assumptions than inference that uses expectations reports at face value. Nevertheless, we cannot be certain that the intervals we construct are accurate. The algorithm is subject to two potential forms of misclassification.

First, a given survey response may be less rounded than the interval assigned by the algorithm; that is, the actual rounding interval may be a subset of the algorithm's interval. Then our use of the data is correct, but it yields inference that is less sharp than it would be if the true degree of rounding were known.

Second, the actual rounding interval may not be completely contained in the algorithm's interval. Then the actual belief may lie outside our interval, making our use of the data incorrect. Still, use of the algorithm substantially lowers the risk and severity of the latter type of error relative to the standard approach that takes survey responses at face value.

4.1. Determination of Respondent Rounding Types

Based on the evidence in Section 3, we allow a respondent's rounding type to vary across question domains and between the tails and center of the measurement scale. Thus, within a specific domain of questions, a respondent's rounding type is a bivariate vector of the form (tail, center) rounding, partitioning the 0-100 scale into two symmetric tails (0-24 and 76-100) and a center (25-75).

We believe that our specific choice of tails and center reasonably reflects the empirical patterns of HRS responses, but judgments need not be uniform. The algorithm can be easily adapted to different definitions of tails and center or extended to accommodate finer partitions of the 0-100 scale (e.g., outer tails, inner tails, center).

The new algorithm refines the earlier one posed by Manski and Molinari (2010) in multiple ways. One refinement is to separate tail from center rounding. Another is to classify persons who only use the response values (0, 25, 50, 75, 100) as rounding to the nearest 25 percent rather than to the nearest 5 percent. The distinction between tail and center rounding is operationally meaningful only for persons who round in a more refined manner; that is, to the nearest 10, 5, or 1 percent. Nevertheless, we find that it simplifies exposition to separate responses in the tail and center of the probability scale for all persons.

A further difference between the two algorithms is that here we use a tighter criterion for assignment of a person to a more refined rounding type. To explain the tighter criterion, consider categorization of a respondent as one who rounds to the nearest 10 percent (or to a more refined degree). Manski and Molinari assigned a respondent to this rounding type if all responses are multiples of 10 and at least one response is not a value in (0, 50, 100). We use here a tighter criterion that requires observation of at least two responses that are multiples of 10 other than (0, 50, 100), of which one must be in the domain under consideration and the other may be in a different domain and may also be a less rounded response.

Adding the new requirement reflects our desire for further credibility when assigning a person to a more refined rounding type. We want enhanced credibility because misclassification into an overly refined rounding category yields an inferential error, as the person's latent beliefs may not entirely lie within the overly refined interval. Misclassification of a person into a rounding category less refined than their actual one does not yield an inferential error, as the less refined interval includes the actual one as a subset.

The main criteria for classification of respondents into (tail, center) rounding types, denoted by ($\mathcal{M}y_n$-T, $\mathcal{M}x_n$-C) with $y_n \in \{1, 5, 10, 100\}$ and $x_n \in \{1, 5, 10, 25, 50\}$, are as follows.

- **Center rounding type** Define $x_1 = 1$, $x_2 = 5$, $x_3 = 10$, $x_4 = 25$, and $x_5 = 50$. Respondent j is classified as rounding to the nearest x_n percent in the center, and denoted $\mathcal{M}x_n$-C, within question domain *l*, if one of the following two conditions holds: (i) they are observed to give at least two answers in the center that are multiples of x_n percent but not of $x_{n'}$ for any *n' < n* within domain *l*; or (ii) they are observed to give one answer in the center that is a multiple of x_n percent (but not of $x_{n'}$ for any *n' < n*) within domain *l* AND at least one answer in the center that is a multiple of $x_{n'}$ for any *n' ≤ n* within a second domain *l'* distinct from *l*.
- **Tail rounding type** Define $y_1 = 1$, $y_2 = 5$, $y_3 = 10$, and $y_4 = 100$. Respondent *j* is classified as rounding to the nearest y_n percent in the tails, and denoted $\mathcal{M}y_n$-T, within question domain *l* if one of the following two conditions holds: (i) they are observed to give at least two answers in the tails that are multiples of y_n percent but not of $y_{n'}$ for any *n' < n* within domain *l*; or (ii) they are observed to give one answer in the tails that is a multiple of y_n percent (but not of $y_{n'}$ for any *n' < n*) within domain *l* AND at least one answer in the tails OR center that is a multiple of $x_{n'}$ for any *n' ≤ n* within a second domain *l'* distinct from *l*.

To illustrate, consider a respondent who has answered four expectations questions in the domain of personal finances, either within the same wave or over multiple waves. Two of the observed responses belong to the tails, {5, 85}, and two to the center, {30, 60}. As the set of responses includes two multiples of 5 percent in the tails and two multiples of 10 percent in the center, our algorithm classifies this respondent as one rounding to the nearest 5 percent, *or to a finer degree*, in the tails ($\mathcal{M}$5-T) and to the nearest 10 percent, *or to a finer degree*, in the center ($\mathcal{M}$10-C).

The Supplementary Appendix SA4.1 provides additional and more complex examples. It also presents the complete algorithm in a formal and compact way in Table S8 (Panels A and B).

4.2. Empirical Distribution of Rounding Types and Association with Observable Characteristics

We apply the algorithm to all HRS respondents who responded to at least one expectations question in any question domain and in any wave between 2002 and 2014. Table 4 reports the empirical distributions of rounding types for each domain of questions, separately for the tails (top panel) and the center (middle panel). Table S9 shows the joint empirical distribution of tail and center rounding types for each domain.

The more aggregated statistics shown in the bottom panel of Table 4 reveal that, depending on the domain, between 40.40% and 61.03% of respondents are inferred to apply finer rounding in the tails than in the center. Between 28.49% and 38.73% of respondents apply the same degree of rounding in the tails and in the center. Between 2.90% and 6.71% of respondents apply coarser rounding in the tails than in the center.

The rounding type of a small minority of respondents could not be determined either in the tails or in the center or both. Most undetermined cases occur when, for a given respondent, we do not observe any answer in the relevant domain and scale segment. Among respondents for whom we observe at least one answer in the relevant domain and scale segment, all cases of undetermined tail rounding type disappear and only a few cases of undetermined center rounding type remain. The latter are respondents for whom we only observe one answer in the center in the relevant domain and no answers in the center in the remaining two domains.

We now investigate how rounding types vary with observable respondent characteristics. We summarize the data using parametric bivariate ordered probit regression, which embodies the basic ordinal property that our rounding categories display across different degrees of rounding.

Table 5 presents estimated coefficients of three bivariate ordered probit regressions, one per question domain. The outcome variables are the respondent's bivariate vectors of tail and center rounding categories in each domain. As predictors, we use binary variables for standard respondent's demographics, including gender (male, with female omitted), educational attainment (high school, some college, bachelor, and graduate, with less than high school omitted), and race (black and other, with white omitted).

Given our assumption of time stability of rounding practices, we cannot include time-varying covariates in the regression. We nevertheless include information on individual's cohort and cognitive functioning by incorporating in our bivariate ordered probit regressions: (i) a cohort indicator based on whether each respondent's cross-wave average age lies in the categories 60-69, 70-79, and 80+ years, with 50-59 the omitted category and (ii) each respondent's cross-wave average cognitive score. See Fisher et al. (2012) and Crimmins et al. (2011) for a description and an empirical assessment of the HRS cognitive measures.

The cognitive score has a range of 0-35. In our data, the respondent-specific cross-wave average cognitive score has a mean of 23 and a standard deviation of 4.11 across respondents. The respondent-specific cross-wave standard deviation in cognitive score has a mean of 2.9 across respondents. The fact that the standard deviation of the cross-wave average score is larger than the average cross-wave standard deviation in the score lessens our concerns for using a time-fixed measure of cognitive functioning in our bivariate ordered probit regressions. Nonetheless, the time variation in cognitive score and its association with rounding warrant study in future research.

The model permits the error terms of the latent variables underlying the inferred tail and center rounding categories to be correlated with each other. The correlation parameter, ρ, is estimated along with the other coefficients. The rounding categories are ordered from least coarse to coarsest. Thus, positive associations indicate a tendency to round more coarsely.

Estimated coefficients with standard errors are reported in Table 5. Table 6 reports predicted probabilities of selected tail and center rounding types for persons with specified covariate values.

We find that higher levels of educational attainment and of person-specific average (cross-wave) cognitive score are associated with a tendency to give more refined responses across all scale segments and question domains. The patterns for the other predictors are more varied.

For example, respondents in the oldest cohort category (80+) tend to give more rounded responses than respondents belonging to the youngest one (50-59) across all scale segments and questions domains.

On the other hand, respondents in the two intermediate cohort groups (i.e., 60-69 and 70-79) belong to rounding categories that may be more refined, coarser, or statistically indistinguishable from those characterizing respondents from younger cohorts, depending on the specific domain or scale segment.

Male respondents tend to round more coarsely than female respondents in the personal health and personal finances domains, but only in the tails. On the other hand, male respondents tend to round less coarsely than women respondents in the center in the domain of general economic conditions. While respondents belonging to the residual race category (including Hispanic, Asian, and Pacific Islander) tend to round more coarsely than white respondents, the differential rounding tendencies of black respondents relative to white respondents vary across question domains and scale segments.

The large, positive, and statistically significant estimates of the correlation parameter ρ reveal that rounding tendencies are positively correlated across scale segments. Hence, respondents who give coarser responses in the tails are more likely to do so in the center.

Parameter estimates for a specification without cognitive score are shown in Table S10 of the Supplementary Appendix.

4.3. Using Survey Responses and Rounding Types to Form Expectations Intervals

It is natural to wonder the extent to which failing to account for rounding might lead to inaccurate conclusions when analyzing data. A simple numerical illustration pertaining to the analysis of the effect of longevity expectations on hours worked shows that ignoring rounding may yield highly inaccurate conclusions.

Suppose that two respondents both round their response to the longevity expectation question to the closest multiple of 25. Suppose that one respondent views their probability to live past age 75 to be forty percent while the other respondent views it to be sixty percent, with the latter working significantly more hours as a consequence. With rounding, both respondents report their probability to live past age 75 as fifty percent. The notable difference in hours-worked outcomes with apparently the same expectations may be misinterpreted as caused by unobserved heterogeneity in labor-leisure preferences, when the actual cause is different longevity expectations.

Next, consider a scenario where the first respondent views their probability to live past age 75 to be thirty-seven percent while the other respondent views it to be thirty-eight percent, with the latter working slightly more hours. With rounding, the first respondent reports a probability to live past age 75 of twenty-five percent, and the second respondent reports fifty percent. The slight difference in outcomes with an apparent large difference in expectations may be misinterpreted as evidence of minimal effect of expectations on labor supply.

These examples, while stylized, illustrate that ignoring rounding might lead to "boundary mistakes;" that is, to significantly underestimating or overestimating an effect of interest. We therefore propose an algorithm that uses the information contained in each respondent's reporting behavior across the survey, as analyzed in the preceding sections, to transform observed percent-chance point reports into intervals.

Here we present the construction of interval data within the context of the illustration introduced in Section 4.1. The Supplementary Appendix SA4.3 presents the complete algorithm formally, discusses more complex cases, and reports the distributions of interval width for the responses given to specific questions.

In the example introduced in Section 4.1, the respondent is observed to answer with {5, 30, 60, 85} to four expectations questions concerning personal finances and is classified to be of rounding type ($\mathcal{M}$5-T, $\mathcal{M}$10-C) in that domain. Because the respondent is classified to round to the nearest 5 percent in the tails, the algorithm assigns to each of the respondent's point responses in the tails an interval of width 5 centered around the point response. Specifically, the algorithm assigns the interval [2.5, 7.5] to response 5 (i.e., 5 $\mp$ 2.5) and the interval [82.5, 87.5] to response 85 (i.e., 85 $\mp$ 2.5). Similarly, as the respondent is classified to round to the nearest 10 percent in the center, the algorithm assigns interval [25, 35] to the 30 percent response (i.e., 30 $\mp$ 5) and the interval [55, 65] to the 60 percent response (i.e., 60 $\mp$ 5).

In general, construction of intervals around point responses near the thresholds which separate the center from the tails (25 and 75 percent) requires specific “boundary conditions.” Such conditions are not binding in this example. We explain them in the Appendix.

By construction, each interval contains the point response because the former is centered around the latter. Moreover, the interval is assumed to cover the unobserved true latent belief with certainty. However, no assumption is made about the location of the true latent belief inside the interval.

Our algorithm relies on considerably weaker and hence more credible assumptions than inference using expectations reports at face value. At the opposite extreme, one could be ultraconservative, maintaining that each point response is consistent with any amount of rounding. One would then replace all reported expectations with a [0, 100] interval. Obviously, doing this empties the data of any information content.

Our choice of assumptions used to identify respondents’ rounding types and bound their unobserved true beliefs strikes a balance between those two extremes and is informed by the respondents’ response patterns across HRS questions and waves, which we have documented in this paper. A researcher entertaining a different set of assumptions about how survey respondents round their expectations reports could easily apply our framework by simply replacing our assumptions with theirs. In general, stronger and/or more numerous assumptions will yield (weakly) narrower intervals.

4.4 Validation of the Algorithm

The panel structure of the HRS provides a unique opportunity for assessing the validity of our approach. Specifically, we use respondents’ expectations observed in 2016 to validate the algorithm we presented earlier in this section. The validation procedure consists of the following steps.

(i) Take any expectation question asked in 2016.

(ii) For each respondent who was asked the question, compare their response in 2016 with the respondent's rounding type in the relevant domain and scale segment inferred by applying the algorithm to the 2002-2014 data.

(iii) If the granularity of the response observed in 2016 to a given expectation question is consistent with the rounding type inferred by the algorithm for the domain to which the question belongs and the scale segment in which the response falls, flag the case as valid. Invalid cases are those where the response given in 2016 is more refined than expected based on the type classification.

(iv) Count the fraction of valid cases out of the total number of cases.

We formally describe the criteria we used to judge consistency between individual 2016 responses and the rounding types inferred by the algorithm based on the 2002-2014 data in Sub-section SA4.4 of the Supplementary Appendix. Such criteria should be apparent in Table 7, which shows validation results for the "survival past 75" question. Table S13 in the Supplementary Appendix shows analogous statistics for the "working past 62" and the "stock market goes up" questions.

A researcher might also be concerned about the opposite mistake, where a respondent rounds more coarsely when responding in 2016 than predicted by the algorithm. Unfortunately, we cannot validate our algorithm with respect to this latter type of mistake because doing so would require observing the unrounded response.

Each panel of Tables 6 and S13 displays the cross-tabulation between the granularity of the response observed in 2016 (by row) and the respondent's response type inferred by the algorithm for the domain to which the question belongs and the scale location in which the response falls (by column). Each cell reports the absolute frequency for the corresponding granularity-type combination. The cells corresponding to the valid cases are marked in green, while the cells corresponding to invalid cases are marked in red.

We find that 93.39% of tail responses and 88.25% of center responses to the survival question in 2016 are consistent with the predictions generated by our algorithm based on the 2002-2014 data. The corresponding figures for the working question are 97.05% and 95.71%. And those for the stock market question are 94.29% and 95.9%. To further understand the properties of our algorithm, we investigate how the share of valid type-assignments for the survival question varies with the number of questions that respondents answered over the 2002-2014 time period. To do so, we divide the sample among individuals who answered no more than 6 questions (11.57% of the sample), exactly 7 questions (62.12% of the sample), and at least 8 questions (26.31% of the sample). We find, respectively for the three subsamples, that 85%, 93.43%, 97.11% of tail responses and 73.38%, 86.76%, 97.02% of center responses are consistent with the predictions generated by our algorithm, thereby indicating a positive association with number of questions answered. These figure are reported in Table S14.

5. Illustrative Applications

This section demonstrates how interval data on subjective expectations can be employed as either an outcome variable or a covariate in prediction analyses of substantive interest. Section 5.1 presents an application where the objective is linear prediction of the labor supply expectations of working HRS respondents, conditional on specified covariates. Section 5.2 studies non-parametric prediction of hours worked of male HRS respondents, using longevity expectations as a covariate. In both cases we examine how accounting for rounding in probabilistic expectations affects the conclusions that one can draw in empirical analysis.

Both applications use the existing body of methodological research on partial identification with interval data, developed in Manski and Tamer (2002) and Beresteanu and Molinari (2008) and reviewed in Molinari (2020). Partial identification analysis determines the inferences possible with interval data. Moreover, it reveals the maximum bias of estimates obtained under the conventional assumption that persons do not round their reports.

To understand computation of maximum bias, consider the simple situation in which one uses a person's response to an HRS question to estimate his true belief. In the absence of assumptions relating responses to true beliefs, true beliefs could be anywhere in [0, 100]. Hence, maximum bias in any response p is in the range [-p, 100 - p]. Under the maintained assumption that our rounding algorithm is correct, true belief could be anywhere in our interval [L, U], so our algorithm places maximum bias in the interval [L - p, U - p]. Similarly, the more complex partial identification analyses that we perform in this section formally show the potential implications for prediction if one takes HRS responses at face value, ignoring respondent-specific rounding.

Our main analysis uses only the interval data as constructed in Section 4. However, we also explain how additional exclusion restrictions can be incorporated in the analysis, when credible, to yield stronger conclusions.

5.1 Predicting Labor Supply Expectations of Older Workers

As the American population ages and a larger fraction of "baby boomers" approach retirement age, it of interest to analyze how subjective expectations of HRS respondents for working full-time past age 62 vary with several covariates, including age, gender, coupledness status, household wealth, race, and education.

In each of the HRS waves analyzed in this paper, respondents younger than 62 at the time of the interview were asked, *"Thinking about work in general and not just your present job, what do you think the chances are that you will be working full-time after you reach age 62?"*. See question P17 in Table S3 for the response distribution in each wave and in Table 3 for the response distribution with data pooled across waves. We compare the conclusions drawn when the elicited expectations are taken at face value, as is commonly done in the related literature (e.g., Honig, 1996, 1998), and when our algorithm is used to characterize rounding. We analyze data from each of the seven waves of the HRS from 2002 to 2014, pooling the data across waves. This yields a sample of size 24,052 after dropping respondents who are younger than fifty and those for whom we do not observe some covariates.

When we take the elicited expectations of working past age 62 at face value, we report the results of best linear prediction under square loss. In this case, we assume that nonresponse is random and drop respondents who answered “Don’t know” or “Refuse” to the probability chance question posed in P017. The pooled sample has size 23,811.

When we use our algorithm to interpret the elicited expectations as intervals under the assumptions set forth in Section 4, we repeat the same exercise of best linear prediction under square loss, considering all points in the interval outcome variable of each respondent to be feasible values of the quantity of interest. In this case, the resulting best linear predictor’s parameter vector is not point identified. Rather, it is *partially identified*, meaning that there is a *set* of values (rather than a single value) for the parameter vector that are consistent with the available data and maintained assumptions. This set of values is called the parameters’ *identification region*. We estimate the identification region and report confidence intervals for it using the method proposed by Beresteanu and Molinari (2008) and the Stata package by Beresteanu et al. (2010). Beresteanu and Molinari (2008, Section 4) and Beresteanu et al. (2012, Section 3.2) give a detailed discussion of the method.

The results of our analysis are reported in Table 8. The first column shows the estimates and confidence intervals when elicited expectations are taken at face value. The results suggest an increased expectation to work full-time past age 62 for individuals who are closer to age 62, who are males, who have lower wealth, and who are more highly educated, while a reduced expectation to work past age 62 for wealthier individuals and for non-whites.

The second through fifth columns report set estimates and confidence intervals when elicited expectations are interpreted as interval data according to our algorithm. The only difference between the empirical exercises reported in the two sets of columns (2-3 and 4-5) is that the set estimation in columns 2-3 maintains the assumption of random nonresponse to the expectation question as in the point estimation in column 1, whereas the set estimation in columns 4-5 follows a more conservative approach by replacing missing observations with [0, 100] intervals. We show both sets of results as some researchers may find the assumption of random nonresponse credible, whereas others may not find it credible. Comparison of the estimates across columns 2-3 and 4-5 quantifies the identifying power of the random nonresponse assumption.

In our description, we focus on the results in columns 4-5. The results reveal that the strength of the conclusions that can be drawn is weaker when we interpret elicited expectations as intervals than when we take them at face value. This is to be expected, as there is an intrinsic trade-off between the strength and the credibility of inference. Despite this, our analysis –under considerably weaker assumptions– continues to find that males and individuals with higher education have higher expectations, while blacks have lower expectations, to work past age 62. Interestingly, the interval data that we construct remains sufficiently informative to allow us to learn the sign of several coefficients of the best linear predictor.

One may be willing to augment the analysis with additional exclusion restrictions. A first type of restriction states that beliefs are statistically independent of an instrument z conditional on other covariates x. Formally, $P(v|x,z) = P(v|x)$. This assumption can be incorporated into the analysis easily using the results reported in Manski (2003, Chapter 2), as we show in Supplementary Appendix SA5.1.

Here we propose a different exclusion restriction, which appears to be new in the analysis of interval data. Let v_i denote individual i's true/unrounded subjective expectation to work past age 62, and let $[v_i^L, v_i^U]$ denote that individual's interval delivered by our algorithm. The assumption that we consider states that conditional on $[v^L, v^U]$, the distribution of v is independent of an observed covariate z. Formally, $P(v|v^L, v^U, z) = P(v|v^L, v^U)$ (of course, the assumption could also be imposed conditional on x). Proposition A1 in Supplementary Appendix SA5.1 derives sharp bounds on $E(v|z = z_0) - E(v|z = z_1)$ under this assumption, where z_0 and z_1 are two possible realizations of the covariate z.

For concreteness, suppose that z is an indicator variable taking value 1 for males and 0 for females. In this context, our exclusion restriction requires that conditional on the interval implied by the algorithm for the probability of working past 62, the distribution of underlying (true/unrounded) subjective probabilities of working past 62 is the same across female and male respondents. Then using the same data as in Columns 4-5 of Table 8, under the exclusion restriction we find that $E(v|z = male) -$

$E(v|z = female) \in [3.95, 5.09]$. By comparison, the bounds on this quantity using the interval data alone equal $[1.21, 14.80]$. The point identified value if one assumes that the responses are not rounded equals 4.57. This simple example illustrates that exclusion restrictions can have substantial identification power in this context.

5.2 Longevity Expectations and Hours Worked

Individuals' life horizon and the related mortality risk are key ingredients of economic models of life-cycle behaviors. This raises the question of whether life horizon and mortality risk as perceived by individuals are empirically important determinants of their labor supply, saving and investment decisions, etc. (e.g., Hamermesh (1985)). Previous work has examined the effect of subjective survival probabilities on retirement and Social Security claiming behaviors of older Americans (e.g., Hurd et al. (2004), Delavande et al. (2006)). Here we focus on the relationship between subjective survival probabilities and hours worked.

In all waves of the HRS, respondents under the age of 65 were asked to report their longevity expectations by means of the following question: *"What is the percent chance that you will live to be* 75 *or more?"* (question P28). The sample distribution of responses to P28 in each wave is displayed in Table S3. Table S12 reports the sample frequencies of the width of the algorithm intervals, constructed around respondents' point responses to question P28 in the 2014 HRS wave.

We focus on working male individuals aged 50 through 64, who were asked to report their percent chance of living past 75. Our outcome variable is weekly hours worked. Hours worked were measured in question J172 as following: *"How many hours a week do you usually work [on this job/in this business]?"* This question was asked only of respondents who answered "yes" to question J20, *"Are you doing any work for pay at the present time?"*.

The predictors used are interval-valued longevity expectations (where we replace missing observations with [0, 100] intervals), age, and coupledness status. As in the first application, the exercise is best prediction of the outcome variable given covariates. We again are interested in comparing the conclusions that can be drawn when rounding is addressed with those obtained when rounding is ignored. Econometrically, the key difference between this application and the earlier one is that now the interval-valued variable is used as a covariate. In this case, the inferential problem is more difficult than when the interval-valued variable is used as an outcome of a regression model, because the estimator is no longer linear in the interval data.

Manski and Tamer (2002) study the problem of inference on regressions with interval data on a regressor. That is, the problem is one of inferring, say $E(y \mid \upsilon, x)$, when υ is unobserved but is known to lie in some interval $[\upsilon^L, \upsilon^U]$ with probability 1. The latter assumption is called Interval (I). Under assumption (I), two additional ones – Monotonicity (M) and Mean Independence (MI), Manski and Tamer (2002) derive the identification region for $E(y \mid \upsilon, x)$ and discuss estimation methods.

We estimate the model using the inferential approach of Chernozhukov, Lee, and Rosen (2013) and the Stata package by Chernozhukov et al. (2015). We again present results for the pooled data, which yield a sample of size 13,717 after dropping respondents with missing covariates. As in the application of Section 5.1, when we take the elicited longevity expectations at face value, we drop respondents who answered “Don’t know” or refused to answer the probability chance question posed.

In the interest of space, we present results graphically in Figure 1 rather than in a table. Each panel of Figure 1 reports on the x-axis the subjective percent-chance that a respondent will survive to age 75. The y-axis reports the mean weekly hours of work predicted in two ways. One uses a linear regression model estimated by least squares, taking the longevity expectations data at face value (“OLS”). The other (“Bounds”) uses interval expectations to account for rounding, where the intervals are those described in Section 4. Additionally, the graphs display 95% confidence intervals for both the OLS and Bounds estimates. Different panels show estimates for different sub-samples, corresponding to different age-coupledness status combinations.

Taking the longevity expectations data at face value, we find that they have a positive but economically insignificant association with hours worked, while hours worked decrease substantially with age and if the respondent is not coupled. When we allow for rounding, as illustrated in the plots in Figure 1, we confirm that predicted mean hours worked increase quite weakly in the perceived likelihood of living past 75, while they decrease markedly as age increases, and for individuals who are not part of a couple.

6. Conclusion

We have studied rounding in numerical reports of probabilistic expectations. Our analysis of the responses to all expectations questions asked in the HRS core questionnaire between 2002 and 2014 confirms the earlier findings of Manski and Molinari (2010) based on analysis of the 2006 wave of data. Moreover, the present analysis establishes important new findings. We present a general inferential approach that interprets expectations reports as interval data. We then implement a specification of the general approach that explicitly incorporates the documented patterns of responses across waves, question domains, and location within the measurement scale.

The main tenet of the analysis is that observed response patterns across questions and waves carry information about individual respondents' rounding practices. Observed response patterns, however, do not reveal whether individual respondents round their reports to simplify communication or to convey partial knowledge. Consistent with the first interpretation, we have assumed that respondents have well-formed latent point beliefs. If instead the relevant latent objects were sets or ranges of beliefs, the algorithm would still work as intended as long as the algorithm's interval completely includes the latent interval.

If respondents round to convey partial knowledge about the likelihood of future events of the kind HRS expectations questions refer to, it would be better to allow them to express their ambiguity directly. This could be achieved by allowing respondents to give either a single percent-chance value or a range as they see fit. Then range measures of subjective expectations may be analyzed using existing econometric tools for interval data. See Manski and Molinari (2010), Giustinelli and Pavoni (2017), Delavande et al. (2019), and Giustinelli, Manski, and Molinari (2020) for data collection and analysis of this type.

References

Altig, D. and J.M. Barrero and N. Bloom and S.J. Davis and B.H. Meyer and N. Parker. 2019. Surveying Business Uncertainty. NBER Working Paper 25956.

Armantier, O., W. Bruine de Bruin, S. Potter, G. Topa, W. van der Klaauw, and B. Zafar. 2013. Measuring Inflation Expectations. *Annual Review of Economics* 5, 273-301.

Attanasio, O. 2009. Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review* 99(2), 87-92.

Beresteanu, A. and F. Molinari. 2008. Asymptotic Properties for a Class of Partially Identified Models. *Econometrica* 76(4), 763-814.

Beresteanu, A. and Molchanov, I. and Molinari, F. 2012. Partial Identification Using Random Set Theory. *Journal of Econometrics* 166(1): 17-32.

Beresteanu, A. and Molinari, F. and Steeg Morris, D. 2010. Asymptotics for Partially Identified Models in STATA, Computer Program available at https://molinari.economics.cornell.edu/programs.html

Bruine de Bruin, W. and K.C. Carman. 2018. Measuring Subjective Probabilities: The effect of response mode on the use of focal responses, validity, and respondents' evaluations. *Risk Analysis* 31(10): 2128-2143.

Bruine de Bruin, W., P.S. Fischbeck, N.A. Stiber, and B. Fischhoff. 2002. What Number is "Fifty-Fifty?": Redistributing Excessive 50% Responses in Elicited Probabilities. *Risk Analysis* 22(4), 713-723.

Chernozhukov, V. and Kim, W. and Lee, S. and Rosen, A.M. 2015. Implementing Intersection Bounds in Stata. *Stata Journal* 15(1):21-44.

Chernozhukov, V. and Lee, S. and Rosen, A.M. 2013. Intersection Bounds: Estimation and Inference. *Econometrica* 81(2): 667-737.

Crimmins, E.M., J.K. Kim, K.M. Langa, and D.R. Weir 2011. Assessment of Cognition Using Surveys and Neuropsychological Assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 66B(S1), i162–i171.

Delavande, A. 2014. Probabilistic Expectations in Developing Countries. *Annual Review of Economics* 6, 1-20.

Delavande, A. and Perry M. and Willis, R.J. 2006. Probabilistic Thinking and Social Security Claiming. MRRC Working Paper, 129.

Delavande, A. and J. Ganguli and F. Mengel. 2019. Measuring Uncertainty Attitudes and Their Impact on Behaviour in General Social Surveys. Working Paper, University of Essex.

Fischhoff, B. and W. Bruine de Bruin. 1999. Fifty–Fifty = 50%? *Journal of Behavioral Decision Making* 12, 149-163.

Fischhoff, B. and A.M. Parker and W. Bruine de Bruin and J. Downs and C. Palmgren and R. Dawes and C.F. Manski. 2000. Teens Expectations For Significant Life Events. *Public Opinion Quarterly* 64(2), 189-205.

Fisher, G.G., H. Hassan, W.L. Rodgers, and D.R. Weir. 2012. Health and Retirement Study – Imputation of Cognitive Functioning Measures: 1992–2010 Early Release. http://hrsonline.isr.umich.edu/modules/meta/xyear/cogimp/desc/COGIMPdd.pdf.

Giustinelli, P. and C.F. Manski. 2018. Survey Measures of Family Decision Processes for Econometric Analysis of Schooling Decisions. *Economic Inquiry* 56, 81-99.

Giustinelli, P. and N. Pavoni. 2017. The Evolution of Awareness and Belief Ambiguity in the Process of High School Track Choice. *Review of Economic Dynamics* 25, 93-120.

Giustinelli, P. and C.F. Manski and F. Molinari. 2020. Precise or Imprecise Probabilities? Evidence from survey response on late-onset dementia. Working Paper.

Hamermesh, D. 1985. Expectations, Life Expectancy, and Economic Behavior. *Quarterly Journal of Economics* 100: 389-408.

Heiss, F. and M.D. Hurd and T. Rossmann and J. Winter and M. van Rooij. 2019. Dynamics and Heterogeneity of Subjective Stock Market Expectations, Rationality & Competion Discussion Paper 157.

Honig, M. 1996. Retirement Expectations by Race, Ethnicity and Gender. *The Gerontologist* 36.

Honig, M. 1998. Married Women's Retirement Expectations: Do Pensions and Social Security Matter? *American Economic Review Papers and Proceedings* 88, 202-206.

Hudomiet, P. and R.J. Willis. 2013. Estimating Second Order Probability Beliefs from Subjective Survival Data. *Decision Analysis* 10(2), 152-170.

Hurd, M.D. 2009. Subjective Probabilities in Household Surveys. *Annual Review of Economics* 1, 543-562.

Hurd, M.D. and J.P. Smith and J.M. Zissimopoulos. 2004. The Effects of Subjective Survival on Retirement and Social Security Claiming. *Journal of Applied Econometrics* 19: 761-775.

Kleinjans, K.J. and A. van Soest. 2014. Rounding, Focal Point Answers and Nonresponse to Subjective Probability Questions. *Journal of Applied Econometrics* 29, 567-585.

Lillard, L. and R.J. Willis. 2001. Cognition and Wealth: The Importance of Probabilistic Thinking. Michigan Retirement Research Working Paper MRRC WP UM00-04, University of Michigan.

Manski, C.F. 2003. *Partial Identification of Probability Distributions*, Springer-Verlag.

Manski, C.F. 2004. Measuring Expectations. *Econometrica* 72, 1329-1376.

Manski, C.F. 2018. Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise. *NBER Macroeconomics Annual*, 32, 411-471.

Manski, C.F. and F. Molinari. 2010. Rounding Probabilistic Expectations in Surveys. *Journal of Business and Economic Statistics* 28(2), 219-231.

Manski, C.F. and E. Tamer. 2002. Inference on Regressions with Interval Data on a Regressor or Outcome. *Econometrica* 70(2), 519-546.

Molinari, F. 2020. Microeconometrics with Partial Identification. Working Paper.

Schotter, A. and I. Trevino. 2014. Belief Elicitation in the Laboratory. *Annual Review of Economics* 6, 103-128.

van der Klaauw, W. 2012. On the Use of Expectations Data in Estimating Structural Dynamic Models. *Journal of Labor Economics* 30(3), 521-554.

Tables and Figures Appendix

Table 1: Response Tendencies in the 2002-2014 HRS

Wave	N	All NR	All 0 or 100	All 0, 50, or 100	Some multiple of 10*	Some multiple of 5**	Some 1-4 or 96-99	Some other
		Response pattern						
Based on the 12 questions asked in all waves								
2002	16032	0.022	0.101	0.101	0.392	0.320	0.054	0.011
2004	18250	0.015	0.062	0.084	0.418	0.353	0.056	0.013
2006	17191	0.027	0.072	0.077	0.409	0.336	0.065	0.014
2008	16060	0.021	0.068	0.063	0.417	0.340	0.072	0.018
2010	20400	0.010	0.053	0.050	0.426	0.350	0.092	0.020
2012	19360	0.015	0.051	0.058	0.445	0.328	0.083	0.020
2014	17647	0.012	0.065	0.062	0.458	0.295	0.090	0.018
Based on all questions asked in each wave								
2002	16032	0.014	0.023	0.039	0.324	0.459	0.119	0.022
2004	18250	0.010	0.019	0.032	0.337	0.467	0.108	0.026
2006	17191	0.025	0.019	0.023	0.263	0.513	0.117	0.039
2008	16060	0.021	0.025	0.019	0.290	0.511	0.101	0.033
2010	20400	0.009	0.029	0.022	0.316	0.442	0.144	0.038
2012	19360	0.014	0.027	0.021	0.317	0.443	0.139	0.038
2014	17647	0.012	0.026	0.022	0.329	0.427	0.142	0.042

NOTE: N = sample size, NR = nonresponse, * ≡ {10, 20, 30, 40, 60, 70, 80, 90}, ** ≡ {5, 15, 25, 35, 45, 55, 65, 75, 85, 95}. The following 12 questions were asked in all HRS waves between 2002 and 2014: P47: mutual fund increase in value; P28: live to be 75 or more; P29: live to be X or more; P5: leave inheritance ≥ \$10,000; P6: leave inheritance ≥ \$100,000; P59: leave inheritance ≥ \$500,000; P7: leave any inheritance; P16: work for pay in the future; P17: work full time after age 62; P18: work full time after age 65; P32: move to nursing home in 5 years; P20: finding a job in few months if unemployed.

Table 2: Transitions of Response Tendencies across Waves

Transition waves:	2002 to 2004	2004 to 2006	2006 to 2008	2008 to 2010	2010 to 2012	2012 to 2014	2002 to 2014
	Frequency (based on the 12 questions asked in all waves)						
% transitions to:							
same category	0.406	0.420	0.406	0.415	0.436	0.433	0.389
adjacent category	0.386	0.383	0.383	0.385	0.377	0.373	0.392
more distant category	0.209	0.197	0.212	0.201	0.187	0.194	0.218
N (100%)	14183	16126	15231	13732	18260	16923	8348
same or adjacent	0.792	0.803	0.788	0.800	0.813	0.806	0.782

NOTE: The percentages shown in the table are calculated from transition matrices of response tendencies defined in terms of the following categories: All NR; All (0, 100); All (0, 50, 100); Some multiple of 10 different not in (0, 50, 100); Some multiple of 5 but not of 10; Some 1-4 or 96-99, Some other. The following 12 questions were asked in all HRS waves between 2002 and 2014: P47: mutual fund increase in value; P28: live to be 75 or more; P29: live to be X or more; P5: leave inheritance ≥ \$10,000; P6: leave inheritance ≥ \$100,000; P59: leave inheritance ≥ \$500,000; P7: leave any inheritance; P16: work for pay in the future; P17: work full time after age 62; P18: work full time after age 65; P32: move to nursing home in 5 years; P20: finding a job in few months if unemployed.

Table 3: Responses by Question and across Waves in the 2002-2014 HRS
(12 questions asked in all waves only)

Question: percent chance that…	N total obs.	NR	V50	V100	V25	V10 T	V10 C	V5 T	V5 C	V1 T	V1 C
		Percentage of responses in:									
Personal Health											
P28: Live to be age 75 or more	56497	0.038	0.219	0.204	0.082	0.270	0.120	0.042	0.010	0.013	0.001
P29: Live to be age X or more	118404	0.050	0.211	0.191	0.075	0.236	0.156	0.049	0.013	0.018	0.001
P32: Move to nursing home in 5 year	74696	0.059	0.120	0.426	0.039	0.206	0.062	0.060	0.003	0.023	0.001
General Economic Conditions											
P47: Mutual funds up /next year	105714	0.157	0.247	0.093	0.076	0.185	0.193	0.025	0.014	0.008	0.001
Personal Finances											
P5: Leave inheritance ≥ $10K	116769	0.046	0.083	0.518	0.028	0.228	0.051	0.028	0.001	0.017	0.000
P6: Leave inheritance ≥ $100K	95625	0.014	0.100	0.490	0.037	0.228	0.072	0.035	0.002	0.022	0.000
P7: Leave any inheritance	19716	0.020	0.053	0.763	0.013	0.098	0.021	0.020	0.001	0.012	0.000
P16: Work for pay in the future	66855	0.018	0.055	0.672	0.021	0.139	0.037	0.035	0.001	0.021	0.000
P17: Work full time after age 62	36603	0.011	0.144	0.333	0.055	0.268	0.120	0.043	0.006	0.020	0.001
P18: Work full time after age 65	37062	0.011	0.144	0.280	0.058	0.282	0.130	0.057	0.008	0.028	0.001
P20: Find job in few months/unemployed	8206	0.012	0.211	0.184	0.061	0.277	0.174	0.050	0.012	0.019	0.001
P59: Leave inheritance ≥ $500K	73872	0.011	0.090	0.490	0.034	0.216	0.073	0.046	0.003	0.037	0.000

NOTE: NR ≡ nonresponse; V50 ≡ {50}, V100 ≡ {0, 100}, V25 ≡ {25, 75}, V10-T ≡ {10, 20, 80, 90}, V10-C ≡ {30, 40, 60, 70}, V5-T ≡ {5, 15, 85, 95}, V5-C ≡ {35, 45, 55, 65}, V1-T ≡ non-round values in 1-24 or 76-99, V1-C ≡ non-round values in 26-74.

Table 4: Distribution of Rounding Types by Scale Location and Question Domain

Rounding Type	Percent Personal Health	Percent Personal Finances	Percent General Economic Conditions
Tail Rounding			
$\boldsymbol{\mathcal{M}}$1-T	10.94	22.73	9.29
$\boldsymbol{\mathcal{M}}$5-T	25.84	31.75	24.93
$\boldsymbol{\mathcal{M}}$10-T	46.13	36.81	49.01
$\boldsymbol{\mathcal{M}}$100	13.31	7.95	7.98
None/$\boldsymbol{\mathcal{U}}$ndetermined	3.78	0.76	8.79
Total	100	100	100
Center Rounding			
$\boldsymbol{\mathcal{M}}$1-C	0.38	0.41	0.55
$\boldsymbol{\mathcal{M}}$5-C	5.87	6.77	7.72
$\boldsymbol{\mathcal{M}}$10-C	52.21	63.11	59.67
$\boldsymbol{\mathcal{M}}$25	12.53	10.77	10.33
$\boldsymbol{\mathcal{M}}$50	16.60	11.39	12.52
None/$\boldsymbol{\mathcal{U}}$ndetermined	12.41	7.55	9.21
Total	100	100	100
Tail versus Center Rounding			
Tails finer than center	**45.42**	**61.03**	**40.40**
Tails same as center	32.60	28.49	38.73
Tails coarser than center	*6.71*	*2.90*	*5.94*
No/$\boldsymbol{\mathcal{U}}$ndet. T and/or C	15.27	7.58	14.93
Total	100	100	100
Sample size	28,044	28,252	28,172

NOTE: For each domain (T=tail and C= center), $\boldsymbol{\mathcal{M}}$1 denotes a respondent who rounds to the nearest 1 percent in that domain; $\boldsymbol{\mathcal{M}}$5 denotes a respondent who rounds to the nearest 5 percent or finer in that domain; and so on. $\boldsymbol{\mathcal{U}}$ndetermined denotes respondents who could not be classified to belong to any of the preceding types.

Table 5: Bivariate Ordered Probit Model Predicting Rounding Type

	Personal Health		**Personal Finances**		**Gen. Econ. Conditions**	
	Tail Type	Center Type	Tail Type	Center Type	Tail Type	Center Type
Male	0.0047	-0.0497	-0.0032	-0.0154	-0.0070	-0.0693
	(0.0149)	(0.0155)	(0.0142)	(0.0153)	(0.0151)	(0.0157)
Aged 60-69 cohort	-0.1961	-0.1436	-0.0116	0.0145	-0.1090	-0.1049
	(0.0180)	(0.0194)	(0.0174)	(0.0189)	(0.0185)	(0.0195)
Aged 70-79 cohort	-0.1639	0.0481	0.1466	0.1987	-0.0941	0.0232
	(0.0199)	(0.0206)	(0.0189)	(0.0204)	(0.0199)	(0.0208)
Aged 80+ cohort	0.1092	0.4465	0.4934	0.5658	0.1718	0.3209
	(0.0266)	(0.0261)	(0.0246)	(0.0258)	(0.0266)	(0.0266)
High school	-0.0842	-0.0864	-0.1277	-0.1579	-0.0614	-0.1115
	(0.0224)	(0.0221)	(0.0208)	(0.0219)	(0.0226)	(0.0227)
Some college	-0.0642	-0.0758	-0.1688	-0.1948	-0.0588	-0.1487
	(0.0362)	(0.0379)	(0.0342)	(0.0372)	(0.0364)	(0.0389)
Bachelor	-0.2027	-0.2432	-0.2677	-0.3073	-0.1726	-0.2692
	(0.0288)	(0.0301)	(0.0277)	(0.0296)	(0.0292)	(0.0305)
Graduate	-0.2818	-0.3658	-0.3367	-0.3549	-0.2438	-0.3454
	(0.0319)	(0.0337)	(0.0307)	(0.0332)	(0.0320)	(0.0341)
Black	0.0188	0.1148	-0.1507	-0.0798	-0.0562	-0.0456
	(0.0220)	(0.0226)	(0.0203)	(0.0220)	(0.0219)	(0.0228)
Other race	0.1136	0.1374	0.0604	0.0173	0.0887	0.0477
	(0.0303)	(0.0322)	(0.0289)	(0.0310)	(0.0314)	(0.0322)
Avg. Cog.	-0.0261	-0.0339	-0.0368	-0.0373	-0.0202	-0.0370
	(0.0022)	(0.0023)	(0.0020)	(0.0022)	(0.0022)	(0.0023)
Rho	0.2595		0.3848		0.2897	
	(0.0081)		(0.0087)		(0.0093)	
N	22,447		24,541		22,593	

NOTES: (i) Respondents with undetermined tail or center rounding type are excluded from this analysis. (ii) Predictors are dummies for gender, cohort (based on age averaged across waves), education, and race, plus average cognition score across waves. (iii) Omitted dummies are 'Female,' 'Aged 50-59 cohort,' 'No degree,' and 'White.' (iv) 'Rho' is the parameter capturing the correlation between the error terms of the tail and center latent equations. (v) Standard errors are in parentheses.

Table 6. Predicted Probabilities of Rounding Types for Selected Covariate Profiles

Panel A. Personal Health –(Female, White, Bachelor Degree) Respondents

		Average Cognition Across Waves						
		Mean -1 SD	Mean	Mean +1 SD		Mean -1 SD	Mean	Mean +1 SD
		Prob. of Type ($\mathcal{M}$5-T, $\mathcal{M}$10-C)				**Prob. of Type ($\mathcal{M}$10-T, $\mathcal{M}$10-C)**		
Cohort Group	50-59	0.1846	0.2036	0.2198	50-59	0.3118	0.3123	0.3064
	60-69	0.2136	0.2289	0.2402	60-69	0.2971	0.2897	0.2767
	70-79	0.2008	0.2194	0.2347	70-79	0.2784	0.2768	0.2696
	80+	0.1433	0.1658	0.1878	80+	0.2494	0.2623	0.2701
		Prob. of Type ($\mathcal{M}$100-T, M25-C)				**Prob. of Type ($\mathcal{M}$100-T, $\mathcal{M}$50-C)**		
	50-59	0.0199	0.0157	0.0121	50-59	0.0312	0.0221	0.0153
	60-69	0.0135	0.0103	0.0077	60-69	0.0192	0.0133	0.0090
	70-79	0.0151	0.0119	0.0091	70-79	0.0256	0.0180	0.0124
	80+	0.0247	0.0207	0.0170	80+	0.0583	0.0433	0.0316

Panel B. Personal Finances –(Female, White, Bachelor Degree) Respondents

		Average Cognition Across Waves						
		Mean -1 SD	Mean	Mean +1 SD		Mean -1 SD	Mean	Mean +1 SD
		Prob. of Type ($\mathcal{M}$5-T, M10-C)				**Prob. of Type ($\mathcal{M}$10-T, $\mathcal{M}$10-C)**		
Cohort Group	50-59	0.2634	0.2724	0.2731	50-59	0.2483	0.2248	0.1976
	60-69	0.2632	0.2722	0.2728	60-69	0.2440	0.2209	0.1942
	70-79	0.2453	0.2621	0.2715	70-79	0.2583	0.2415	0.2191
	80+	0.1887	0.2162	0.2402	80+	0.2665	0.2665	0.2586
		Prob. of Type ($\mathcal{M}$100-T, $\mathcal{M}$25-C)				**Prob. of Type ($\mathcal{M}$100-T, $\mathcal{M}$50-C)**		
	50-59	0.0072	0.0049	0.0032	50-59	0.0107	0.0065	0.0038
	60-69	0.0071	0.0048	0.0031	60-69	0.0107	0.0065	0.0038
	70-79	0.0102	0.0071	0.0048	70-79	0.0175	0.0110	0.0067
	80+	0.0196	0.0149	0.0109	80+	0.0443	0.0298	0.0194

Panel C. General Economic Conditions –(Female, White, Bachelor Degree) Respondents

		Average Cognition Across Waves						
		Mean -1 SD	Mean	Mean +1 SD		Mean -1 SD	Mean	Mean +1 SD
		Prob. of Type ($\mathcal{M}$5-T, $\mathcal{M}$10-C)				**Prob. of Type ($\mathcal{M}$10-T, $\mathcal{M}$10-C)**		
Cohort Group	50-59	0.2031	0.2170	0.2273	50-59	0.3733	0.3724	0.3647
	60-69	0.2201	0.2315	0.2387	60-69	0.3625	0.3562	0.3435
	70-79	0.2157	0.2298	0.2401	70-79	0.3509	0.3495	0.3415
	80+	0.1671	0.1858	0.2027	80+	0.3524	0.3658	0.3725
		Prob. of Type ($\mathcal{M}$100-T, $\mathcal{M}$25-C)				**Prob. of Type ($\mathcal{M}$100-T, $\mathcal{M}$50-C)**		
	50-59	0.0111	0.0088	0.0068	50-59	0.0165	0.0116	0.0080
	60-69	0.0086	0.0067	0.0051	60-69	0.0119	0.0082	0.0056
	70-79	0.0094	0.0074	0.0051	70-79	0.0145	0.0101	0.0070
	80+	0.0166	0.0138	0.0112	80+	0.0317	0.0233	0.0167

NOTES: (i) ($\mathcal{M}$ 5-T, $\mathcal{M}$10-C) denotes rounding to the nearest 5 percent or a finer degree in the tails and rounding to the nearest 10 percent or a finer degree in the center. ($\mathcal{M}$ 10-T, $\mathcal{M}$10-C) denotes rounding to the nearest 10 percent or a finer degree in both the tails and the center. ($\mathcal{M}$ 100-T, $\mathcal{M}$25-C) denotes rounding to any degree in the tails and to the nearest 25 percent or a finer degree in the center. ($\mathcal{M}$ 100-T, $\mathcal{M}$50-C) denotes rounding to any degree in both the tails and the center. (ii) Predicted probabilities are evaluated at the mean value of average cognition across waves (denoted Mean), at the mean minus one standard deviation value of average cognition across waves (denoted Mean – 1 SD), and at the mean plus one standard deviation value of average cognition across waves (denoted Mean + 1 SD). Predicted probabilities are evaluated at cohort dummies, based on the person's average age across waves falling in each of the categories 50-59, 60-69, 70-79, and 80+.

Table 7 Validation: Percent Chance of Living to Be 75 or More

Panel A. Tail responses – Absolute frequencies reported in each cell

	Inferred tail rounding type in health domain based on algorithm and 2002-2014 data					
Granularity of tail response to survival past 75 in 2016		$\mathcal{M}$1-T	$\mathcal{M}$5-T	$\mathcal{M}$10-T	$\mathcal{M}$50-T	$\mathcal{U}$ndet-T
	Multiple of 1	37	20	39	3	0
	Multiple of 5	46	119	81	8	0
	Multiple of 10	173	492	944	70	0
	0 or 100	117	255	668	270	0

NOTES: Sub-sample size = 2,507 (after dropping 8 observations for which rounding type missing). Percentage of consistent cases in the tails = 93.39% (green-colored cells).

Panel B. Center responses – Absolute frequencies reported in each cell

	Inferred center rounding type in health domain based on algorithm and 2002-2014 data						
Granularity of center response to survival past 75 in 2016		$\mathcal{M}$1-C	$\mathcal{M}$5-C	$\mathcal{M}$10-C	$\mathcal{M}$25	$\mathcal{M}$50-C	$\mathcal{U}$ndet-C
	Multiple of 1	1	3	1	1	1	1
	Multiple of 5	1	18	31	6	4	0
	Multiple of 10	2	84	512	36	94	25
	25 or 75	7	53	251	64	37	13
	50	5	69	812	153	244	50

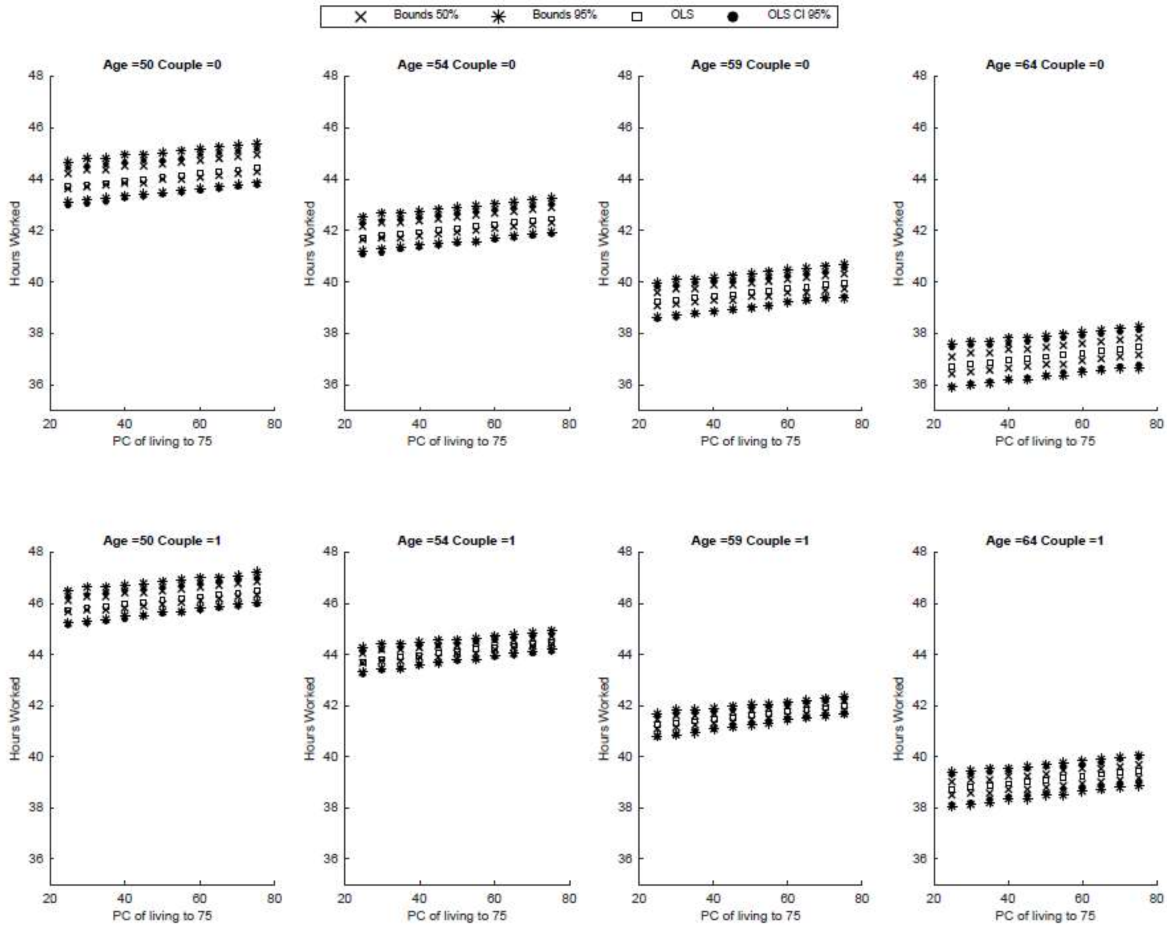
NOTES: Sub-sample size = 2,579 (after dropping 3 observations for which rounding type missing). Percentage of consistent cases in the center = 88.25% (green-colored cells).

Table 8: BLP Prediction of Retirement Expectations:
Point Estimates vs. Set Estimates with Pooled HRS 2002-2014 Data

	OLS Estimates I	**Set Estimates I**		**Set Estimates II**	
	(MCAR imposed)	**LB**	**UB**	**LB**	**UB**
Age	0.1638	-0.4036	0.7212	-0.4944	0.8110
	(0.0306, 0.2970)	(-0.5177, 0.8353)		(-0.5944, 0.9110)	
Coupled	-2.694	-8.5773	3.2792	-9.6014	4.4009
	(-4.1348, -1.2533)	(-9.6555, 4.3573)		(-10.6521, 5.4517)	
Male	8.2172	2.1835	13.9580	1.1365	14.7641
	(7.0017, 9.4327)	(1.2710, 14.8705)		(0.2024, 15.6982)	
Negative wealth	6.1812	-1.6447	13.5758	-4.1145	15.4530
	(4.3986, 7.9637)	(-3.2409, 15.1720)		(-5.7203, 17.0588)	
Below median wealth	6.2116	-1.5954	13.5862	-3.9164	15.2990
	(4.4898, 7.9333)	(-2.9980, 14.9888)		(-5.4242, 16.8065)	
Above median wealth	-0.4701	-9.3489	8.1918	-11.5634	9.8589
	(-2.5209, 1.5808)	(-10.9746, 9.8176)		(-13.1321, 11.4276)	
Black	-9.8655	-16.0655	-3.3521	-17.2459	-2.2001
	(-11.5115, 8.2196)	(-17.2151, -2.20253)		(-18.3527, -1.0933)	
Other race	-4.8209	-11.5792	2.1776	-13.2752	4.2223
	(-6.8371, -2.8046)	(-12.9955, 3.5940)		(-14.7696, 5.7167)	
High school	10.5356	3.0627	17.337	0.2633	19.1237
	(8.7016, 12.3696)	(1.5481, 18.8521)		(-1.1983, 20.5853)	
Some college	13.4775	4.7073	21.5118	1.9292	23.2121
	(10.7289, 16.2260)	(2.7421, 23.4770)		(0.0495, 25.0918)	
Bachelor degree	17.0926	7.9728	25.3006	5.2205	27.0224
	(14.6899, 19.4953)	(6.0970, 27.1764)		(3.5435, 28.6994)	
Graduate degree	19.1551	9.7651	27.6084	7.0036	29.3428
	(16.3555, 21.9546)	(7.8350, 29.5384)		(5.0635, 31.2829)	
Constant	26.0763	-5.8898	59.6645	-10.5647	65.9696
	(18.3266, 33.8259)	(-12.6411, 66.4158)		(-16.5846, 71.9895)	
N	23,811	23,811		24,052	

NOTE: OLS and SetBLP estimates I calculated after dropping DK/RF responses to the point PC question. SetBLP estimates II include DK/RF responses to the point PC question. 95% confidence intervals in parenthesis. OLS CIs clustered at the HH level. SetBLP estimates calculated using 501 bootstrap repetitions. Beresteanu and Molinari (2008)'s confidence sets based on directed Hausdorff. Omitted dummies are '0 wealth,' 'white,' and 'no degree.'

Figure 1: BLP Prediction of Hours Worked Per Week: Point Estimates vs. Set Estimates with Pooled HRS 2002-2014 Data



NOTE: OLS and SetBLP estimates of hours worked per week as a function of longevity expectations, age, and coupledness status. SetBLP estimates are obtained using Chernozhukov et al. (2013, 2015)'s inferential approach. Each graph plots the estimates as a function of longevity expectations for different age groups-coupledness status combinations.