Improving inference using probabilistic expectations

Roger Prudon
Student ID: 387698

July 2, 2017

Supervisor: Prof. dr. R. Lumsdaine
Second assessor: B. Keijsers MSc

Bachelor Thesis Double Degree Programme Econometrics & Economics
Erasmus School of Economics
Erasmus University Rotterdam

Abstract

The focus of this paper is on the use of probabilistic expectations in economic modeling and the effect of rounding on such modeling. A bootstrap algorithm is proposed to obtain empirical parameter estimate bounds, confidence intervals and diagnostic distributions for partially identified models. The algorithm leads to stricter parameter bounds and confidence intervals than those obtained through standard estimation methods. Using the results from the algorithm the effect of focal point answers and probability weighting are assessed for predicting wealth based on the subjective survival probability. The analysis showed that deleting respondents who gave focal point answers did not improve inference and the results for using probability weighting were inconclusive.

Contents

1 Introduction

One of the main interests of (micro-)economics is to predict choice behavior. To do so, economics students are taught to look at what people do, instead of looking at what people state (Manski, 2017). For example, when assessing whether a person prefers apples or pears, an economist should always look at which of the two a person buys (given that they are priced the same). Economic reasoning should thus rely on so-called revealed preferences and not on stated preferences. The issue with revealed preference analysis is that it relies on strong assumptions, in particular the assumption that the expectations of people are rational (i.e., objectively correct). As researchers have great difficulty defending this assumption, revealed preference analysis has been going through a credibility crisis (Manski, 2004).

From the 1990's onward, some economists argue that combining stated and revealed preferences improves the predictive power and credibility of economic analysis. In particular the use of self-reports of expectations in the form of subjective probabilities has become increasingly popular to relax or validate the assumptions of revealed preference analysis. To obtain these subjective probabilities, several large-scale surveys now include sections on probabilistic expectations and much research has been done on the use of these probabilities (Manski, 2004).

The use of subjective probabilities however gives rise to some issues. Manski and Molinari (2010) argue that the response patterns of people to probabilistic expectation questions is such, that it is likely that people round their answers. The degree of rounding differs between people and there are no established conventions for rounded survey answers. If people do in fact round, this implies that the general assumption of white noise is violated. Manski and Molinari (2010) therefore propose to transform the answers to probabilistic questions to intervals, and use these intervals for any inference. Doing so using the standard methods to deal with interval data, renders many results, obtained when using the answers at face value, insignificant.

This thesis will therefore investigate whether current techniques can be improved, leading to the following research question:

"Can inference using probabilistic expectations be improved through improving the predictive power of these expectations and/or through obtaining stricter, credible confidence intervals?"

To answer the research question two areas will be considered. First of all, insights from behavioral economics will be applied to subjective probabilities. In particular the evaluation of probabilities through probability weighting and the effect of focal point answers will be analyzed. Second of all, the estimates of the current methodology for interval data will be evaluated and diagnostics will be developed to compare models.

Section 2 reviews the literature on the various areas of interest. Section 3 describes the data which will be used and section 4 discusses the methodology. In section 5 the empirical analysis is presented, followed by some concluding comments in section 6.

2 Literature Review

This section reviews the literature on the various fields of research of interest. Literature on the methodology will be discussed in the methodology section. Section 2.1 discusses literature on probabilistic expectations and section 2.2 reviews the literature on probability weighting.

2.1 Probabilistic expectations

As described in the introduction, the use of probabilistic expectations has increased significantly since the early 1990's. The idea of using probabilistic expectations however appears to originate in 1966 with Juster (1966) who argued that stated preferences on binary choice decisions should not be interpreted in the same way as revealed preferences of the same binary choice decision. He stated: *"Consumers reporting that they 'intend to buy A within X months' can be thought of as saying that the probability of their purchasing A within X months is high enough so that some form of 'yes' answer is more accurate than a 'no' answer."*. Therefore directly asking for the probability of buying A would be more informative.

For the next quarter century the idea of Juster however did not appeal to researchers and revealed preference analysis was standard practice (Manski, 2004). Revealed preference analysis is based on observations of decisions made by a random sample. Based on these observations probabilistic choice models can be made (McFadden et al., 1973). As pointed out in the introduction, revealed preference analysis is based on the assumption that the expectations of people are rational. As researchers had difficulties with defending this assumption, revealed preference analysis experienced a credibility crisis. Starting in the 1990's the idea of Juster (1966) started to become widely accepted and used to improve the credibility of revealed preference analysis.

With the increased interest in probabilistic expectations, various large scale surveys now include sections focused specifically on obtaining these expectations and substantive research has been done on these surveys. These include but are not limited to the Health and Retirement Study (Juster & Suzman, 1995; Hurd & McGarry, 1995) and the Survey of Economic Expectations (Dominitz & Manski, 1996, 1997). Manski (2017) illustrates more thoroughly how the use of probabilistic expectations has developed since the 1990's.

The starting point of this research is the paper of Manski and Molinari (2010), which investigated the effect of rounding by considering the response patterns to the 2006 Health and Retirement Study (HRS). As Manski and Molinari (2010), and a considerable part of all research on probabilistic expectations, focus on a particular probabilistic expectation: the subjective survival probability, section 2.1.1 will examine the literature on these survival probabilities. Section 2.1.2 will look into the literature on rounding and section 2.1.3 discusses focal point answers.

2.1.1 Subjective survival probability

The subjective survival probability is the probability someone thinks he or she has to survive until a certain age. This subjective probability is of special interest as the time someone expects to remain alive is of great influence on any inter-temporal choice according to the life-cycle model (Fisher, 1930). The life-cycle model assumes that the choice for spending in the current time period does not depend on the current income, but on the average lifetime income. Thus, if someone believes he or she will live for a longer time, the average lifetime income is lower (assuming a constant retirement age) and he or she will thus spend less in the current period.

Research on the subjective survival probability has focused on two aspects of the life-cycle model. First of all, the relationship between subjective survival probabilities and actual mortality rates has been examined. Hurd and McGarry (2002) find that subjective survival probabilities do on average forecast actual mortality. Respondents in their dataset who survived reported an approximately 50% greater subjective survival probability at baseline. Subjective survival probabilities do however suffer from systematic biases (Elder, 2007). Many respondents fail to account for changes in the actual survival probabilities with age and increases in longevity do not appear to be incorporated in subjective survival probabilities. Elder (2007) however does conclude that subjective survival probabilities include information not found in self-reported health status or objective measures of health and are thus a useful predictor of mortality.

The second aspect research focuses on is the relationship between subjective survival probabilities and actual inter-temporal choices such as savings, retirement and consumption. According to the life-cycle model a longer expected lifespan should lead to higher savings until retirement, a later retirement age and less consumption in the current time period. Salm (2010) finds that respondents have a tendency to consume more in the current period if they have a lower subjective survival probability. Hurd, Smith, and Zissimopoulos (2004) show that people with very low subjective survival probabilities tend to retire earlier. To assess whether respondents with higher subjective survival probabilities save more before retirement Bloom, Canning, Moore, and Song (2006) examine the relationship between total wealth of people younger than 65 and their subjective survival probabilities. They find a positive effect between the subjective survival probabilities and the wealth of couples, but this effect is not significant for singles. Furthermore they find no effect on the length of the working life.

Overall the predictive strength of subjective survival probabilities differs strongly between both the various inter-temporal choices and between the analysis of different researchers. These differences can have various causes. The samples of the various articles mentioned above differs significantly in terms of the ages of the respondents in the sample. Furthermore, the methodology differs strongly as either Ordinary Least Squares (OLS), instrumental variable estimation, or probit/logit modeling is used.

2.1.2 Rounding

As mentioned in the introduction, one of the issues of using probabilistic expectations is that response patterns are such that people appear to round their answers. Manski and Molinari (2010) show that for the 2006 HRS 99% of the answers to probabilistic questions were either multiples of 5 or in the 1-4 and 96-99 range. To determine the degree of rounding Manski and Molinari (2010) look into the response patterns to multiple probabilistic questions. These response patterns were used to categorize respondents into groups using the following algorithm:

1. All NR: respondent did not answer any question posed to them
2. All 0 or 100: respondent only answered 0 or 100
3. All 0, 50 or 100: respondent only answered 0, 50 or 100
4. Some M10: respondent only answered with multiples of 10, but not with multiples of 5
5. Some M5: respondent only answered with multiples of 5, and at least once with a multiple of 5 which was not a multiple of 10
6. Some 1-4 or 96-99: respondent answered either with multiples of 5, or with values in the 1-4 and 96-99 range. At least one response in the 1-4 or 96-99 range
7. Some other: respondent answered at least once with a value which was not a multiple of 5 and not in the 1-4 or 96-99 range

Based on these groups, the degree of rounding was assessed and the answers to the probabilistic questions were transformed to interval data using the following algorithm:

1. If $v_{jk} =$(All NR), then $[v_{jkL}, v_{jkU}] = [0, 100]$
2. If $r_{jm} =$(All 0 or 100), then $[v_{jkL}, v_{jkU}] = [max(0, v_{jk} - 50), min(v_{jk} + 50, 100)]$
3. If $r_{jm} =$(All 0, 50 or 100), then $[v_{jkL}, v_{jkU}] = [max(0, v_{jk} - 25), min(v_{jk} + 25, 100)]$
4. If $r_{jm} =$(Some M10), then $[v_{jkL}, v_{jkU}] = [max(0, v_{jk} - 5), min(v_{jk} + 5, 100)]$
5. If $r_{jm} =$(Some M5), then $[v_{jkL}, v_{jkU}] = [max(0, v_{jk} - 2.5), min(v_{jk} + 2.5, 100)]$
6. If $r_{jm} =$(Some 1-4 or 96-99) and $v_{jk} \in [0, 5] \cup [95, 100]$, then $[v_{jkL}, v_{jkU}] = v_{jk}$
7. If $r_{jm} =$(Some 1-4 or 96-99) and $v_{jk} \in [10, 90]$, then $[v_{jkL}, v_{jkU}] = [v_{jk} - 2.5, v_{jk} + 2.5]$
8. If $r_{jm} =$(Some other), then $[v_{jkL}, v_{jkU}] = v_{jk}$

To verify their algorithm, Manski and Molinari (2010) used a follow-up question asking whether or not the answers were exact. Whenever respondents indicated their answer was not exact, they were asked to what extent they rounded. The analysis showed that the acquired intervals were on average more narrow than those obtained through the algorithm. However, for the group of people who indicated that they rounded, the algorithm yielded on average intervals of approximately the same width as the stated intervals. Tendencies to round are also found by Ruud, Schunk, and Winter (2014) and Kleinjans and van Soest (2010).

Rounding has a large impact on statistical analysis (Ruud et al., 2014) as the white noise assumption no longer holds. Given the presence of rounding and its effect on statistical analysis, methods should be used which incorporate interval data (Manski & Molinari, 2010). The methodology used in this paper will be discussed in the methodology section.

2.1.3 Focal point answers

A second issue with probabilistic expectations is the proportion of focal point answers. For probabilities these are often described as 0, 50 and 100. The analysis of Manski and Molinari (2010) showed large proportions of 0, 50 and 100 answers. Research has been done to the (un-)informativeness of these responses and the conclusions drawn vary. Bruine de Bruin and Carman (2012) conclude that answers of 50% in general imply a respondent does not know a probability and the answers of these respondents are thus uninformative. Kleinjans and van Soest (2010) conclude that incorporating focal point answers could change the conclusions made through analysis with probabilistic expectations. Lastly, Gouret (2017) shows in his analysis that even though a large part of focal point answers is uninformative, focal point answers are still a minority in the uninformative response group. This implies for any analysis that one has to weight the loss of informative focal point answers against the gain through reduced noise.

2.2 Probability weighting

The Allais paradox was originally presented to show that expected utility theory cannot always predict the decisions of people (Allais, 1990). People appear to be more sensitive to changes in probabilities at the extremes, and less sensitive in the mid-range. To account for this Kahneman and Tversky (1979) propose the following inverse-S shaped probability weighting function:

$$w(p) = \frac{p^c}{(p^c + (1-p)^c)^{1/c}} \quad \text{, with c } \in (0,1) \tag{1}$$

As seen in (1), the probability weighting function depends on the parameter c. Previous research on diminishing sensitivity to probabilities has shown values for c between .56 and .71 (Tversky & Kahneman, 1992; Camerer & Ho, 1994; Wu & Gonzalez, 1996).

The effect of the weighting function is that the weight assigned to a probability is larger than the probability for small probabilities, and smaller for large probabilities. More importantly is the derivative of the weighting function as the derivative shows the effect of small changes in the probability. The derivative is larger than one at the extremes and smaller than one in the middle region. This implies that the effect of changes in probabilities at the extremes has an effect on the weighting which is larger than the change in the probability itself. Changes in the probability in the middle region have an effect on the weighting which is smaller than the change in the probability itself. The exact effects for the specific cases discussed in this paper will be examined in the results section.

Originally, probability weighting was used to model decision making of people under risk or uncertainty. If one wants to use the stated probabilistic expectations of people to predict their actions, transforming the expectations using a probability weighting function could improve the predictive strength. An extra advantage of doing so is that if the stated expectation is around 50%, transforming the probabilistic expectation interval makes it more narrow. More narrow intervals could lead to results becoming significant. However, if the stated expectation is close to the extremes, 0% and 100%, the intervals become wider which could lead to results becoming insignificant. The effect on significance therefore depends on the exact case at hand.

For the analysis in this paper the subjective survival probability and its interval bounds are transformed through equation (1) before being used as data.

3 Data

The starting point of the research concerns the replication of the results of Manski and Molinari (2010). For this purpose the HRS[1] is used. The HRS is a nationally (USA) representative panel

[1] As part of agreement for usage of data: "The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan."

survey of individuals aged 51-61 at the time of the survey and their partners. The first wave of the HRS was administered in 1992, and subsequent waves followed biennially. In the subsequent waves respondents were not deleted if their age surpassed 61.

This paper will use both the 8th and 9th wave of the HRS, abbreviated from here on as the 2006 HRS[2] and 2008 HRS[3] respectively. The 2006 HRS is used in order to replicate the results of Manski and Molinari (2010). The 2008 HRS will be used as it contains more probing question on the degree of rounding. After deleting the respondents of the HRS which were asked no questions, the 2006 HRS and 2008 HRS had respectively 17191 and 16060 respondents.

The module of the HRS containing the probabilistic expectation questions, module P, starts with asking the probability that it will rain or snow tomorrow. This is done as weather forecasts often include some probabilistic measures. Asking a probabilistic question about the weather should thus encourage respondents to think in probabilistic terms. Following the analysis of Manski and Molinari (2010), the remaining questions in the module will be split up into three groups; questions about personal finance, personal health and general economic conditions.

Besides the probabilistic expectation questions the research will use 3 other variables. First of all the age and gender of respondents will be used. These two variables were obtained through the HRS. The third variable, wealth, was obtained through the RAND HRS[4]. The RAND HRS is a cleaned version of the HRS which includes additional derived variables. Other control factors such as martial status and race will not be incorporated as the main aim is to replicate the analysis of Manski and Molinari (2010) and extend on their methodology.

The following sections will examine the response patterns to the probabilistic questions and the rounding tendencies of the respondents. Next the responses to the subjective survival probability question are analyzed. The last part of the data section concerns the wealth, age and gender of the respondents.

3.1 Response patterns to questions

This section replicates the results of table 1 of Manski and Molinari (2010) and applies the analyses to both the 2006 HRS and 2008 HRS. To gain intuition on the responses, table 1 shows the responses to a selection of the questions. The selection was based on the selection of questions used by Manski and Molinari (2010). As some of the questions present in the 2006 HRS were no longer asked in the 2008 HRS, the selection of questions in the 2008 HRS is smaller. The question numbers in both HRS editions are the same, i.e. question 4 in the 2006 HRS is identical to question 4 in the 2008 HRS. In line with the analysis of Manski and Molinari (2010) the questions were split up into three categories; personal finance, personal health and general economic conditions. The questions in table 1 concerning personal finance are 4, 5, 14, 15, 18, 70, 30 and 31. Questions 28, 103 and 32 concern personal health and questions 34, 110, 47 and 114 concern general economic conditions. Besides the fraction of responses, both tables also state the number of respondents to whom the question was asked. These numbers differ strongly between the various questions as the HRS makes extensive use of skip-sequencing[5].

[2]Health and Retirement Study, (2006 HRS Core (Final)) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2006)

[3]Health and Retirement Study, (2008 HRS Core (Final)) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2008)

[4]RAND HRS Data, 2008. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (August 2016).

[5]Skip-sequencing implies that whether or not a question is asked depends on the answer to a different question or on certain variables such as age and gender

Table 1: Responses by question in the 2006 HRS and 2008 HRS

2006 HRS

		Fraction of responses								
Question	N	NR	0	1-4	50	96-99	100	M10	M5	Other
3:Rain or snow tomorrow	17191	0.029	0.218	0.015	0.150	0.004	0.047	0.468	0.066	0.003
Personal finance										
4:income keep up with cost of living	17191	0.069	0.173	0.012	0.182	0.002	0.063	0.387	0.108	0.003
5:leave inheritance ≥$10,000	17191	0.053	0.159	0.004	0.067	0.008	**0.447**	0.209	**0.052**	0.001
14:lose job during next year	4797	0.020	0.461	0.026	0.107	0.001	0.018	0.274	0.090	0.003
15:find equally good job	4797	0.017	0.173	0.014	0.152	0.004	0.143	0.383	0.112	0.003
18:work full time after age 65	5148	0.016	**0.276**	0.020	0.126	0.003	0.095	0.348	0.114	0.002
70:medical expenses use savings	16754	0.064	0.254	0.011	**0.137**	0.001	**0.056**	0.357	0.118	0.002
30:give help of $5000 or more	16754	0.027	0.382	0.008	0.114	0.002	0.118	0.263	0.084	0.001
31:receive help of $5000 or more	16754	0.027	0.646	0.020	0.044	0.000	0.016	0.173	0.072	0.001
Personal health										
28:live to be 75 or more	6713	0.040	**0.053**	0.004	**0.222**	0.005	0.152	0.375	0.144	0.004
103:live independently at 75	2558	0.015	0.012	0.004	**0.214**	0.004	**0.136**	0.433	0.180	0.002
32:move to a nursing home in 5 years	10044	**0.075**	**0.463**	0.021	**0.101**	0.000	0.007	0.231	0.100	0.002
General economic conditions										
34:U.S. have economic depression	16754	**0.078**	**0.066**	0.006	**0.238**	0.002	0.060	0.404	0.142	0.004
110:Social Sec. will be less generous	16754	0.065	0.048	0.003	0.231	0.005	**0.120**	0.387	0.139	0.002
47:mutual fund increase in value	16754	**0.240**	0.042	0.003	**0.231**	0.001	**0.036**	0.339	0.106	0.003
114:mutual fund increase in real terms	16680	0.281	0.068	0.003	0.182	0.000	0.028	0.334	0.099	0.003

2008 HRS

		Fraction of responses								
Question	N	NR	0	1-4	50	96-99	100	M10	M5	Other
Personal finance										
5:leave inheritance ≥$10,000	16060	0.050	0.153	0.004	0.067	0.010	**0.431**	0.236	**0.046**	0.002
18:work full time after age 65	4159	0.026	**0.181**	0.023	0.135	0.004	0.100	0.397	0.131	0.004
70:medical expenses use savings	15728	0.065	0.264	0.013	**0.127**	0.001	**0.051**	0.357	0.119	0.003
Personal health										
28:live to be 75 or more	5567	0.038	**0.041**	0.004	**0.207**	0.005	0.156	0.394	0.148	0.006
103:live independently at 75	5032	0.038	0.014	0.003	**0.177**	0.006	**0.115**	0.450	0.191	0.005
32:move to a nursing home in 5 years	10106	**0.061**	**0.433**	0.020	**0.089**	0.000	0.007	0.281	0.106	0.002
General economic conditions										
34:U.S. have economic depression	15727	**0.060**	**0.044**	0.005	**0.194**	0.006	0.137	0.409	0.141	0.004
110:Social Sec. will be less generous	15727	0.064	0.049	0.002	0.223	0.006	**0.111**	0.395	0.147	0.003
47:mutual fund increase in value	15727	**0.197**	0.057	0.004	**0.216**	0.001	**0.028**	0.374	0.119	0.004

Note: N=Number of people to who the question was asked, NR=nonresponse, M10=multiple of 10 but not 0, 50 or 100, M5=multiple of 5 but not of 10, bold printed is statistically different between 2006 HRS and 2008 HRS

The NR column gives the fraction of people who did not respond to a question, even though the question was asked. This could imply that they answered refused to answer or don't know. Even though literature proposes that don't know and refusal responses have different causes and could thus have a different effect on analysis (Shoemaker, Eichholz, & Skewes, 2002), both answer possibilities will be categorized as non-responses. Distinguishing between the two answer possibilities could be a point of further research.

The categories 0, 1-4, 50, 96-99 and 100 show the fraction of people who responded with either the precise values 0, 50 and 100 or the fraction of people who responded in the ranges 1-4 or 96-99. M10 responses are responses which are multiples of 10 but not 0, 50 or 100, for example 30. Likewise, M5 responses are responses which are multiples of 5 but not of 10, such as 15. Responses other than those which fall in the previously mentioned categories are categorized as Other. These are thus not multiples of 5, greater than 5 and smaller than 95.

The responses to the various categories do not differ much between the 2006 HRS and 2008 HRS. Bold printed fractions indicate that the proportion of responses differs significantly between the 2006 HRS and 2008 HRS, at a 5% significance level. It should be noted that given a 5% significance level, 1 in 20 proportions will on average be falsely categorized as significantly different if in reality none are. Given that 81 proportions were tested, 4 on average would be falsely categorized as significantly different. A total of 18 proportions are significantly different but the differences are not large and the overall pattern is the same for both years. This is as expected as a large part of the sample is identical for both surveys and the questions themselves are identical. Restricting the 2008 HRS sample to those respondents who were in the 2006 HRS does not alter the differences. The general patterns discussed in the remainder of the data section therefore apply for both questionnaires, unless differences are pointed out explicitly.

The only questions with high nonresponse fractions are question 47 and 114 of the 2006 HRS and 2008 HRS (question 114 was not asked in 2008). However, for both questions and both years more than 95% of the people who did not respond answered don't know implying that the high level of nonrespondents is most likely not due to a reluctance to answer but due to a lack of knowledge. All other questions have nonresponse fraction below 0.08 implying high willingness to answer the survey and in particular the probabilistic expectations questions.

The majority of responses are multiples of 5. At most 4% of all responses fall in the 1-4, 96-99 or Other categories. Of these categories 1-4 responses are observed most frequently with the fraction of responses between 0% and 2.6%. The fraction of responses in the 96-99 category lies between 0.0% and 1.0% and the fraction of responses in the Other category lies between 0.1% and 0.6%. Within the different categories which are multiples of 5; 0, 50, 100, M10 and M5, the fractions of responses varies between 1.2% and 64.6% without a clear pattern.

Considering only the response to a particular question, inferring anything about the degree of rounding is troublesome. A 0 or 100 response, occurring frequently, could either imply that a respondent is fairly sure or absolutely sure of an event. Likewise, a response of 50 could indicate total ambiguity or a fairly sure belief that the event will happen with a 50% probability.

In order to test whether the degree of rounding varies between responses Manski and Molinari (2010) use the following joint hypothesis based on the fraction of 50, M10 and M5 responses:

1. all persons giving a 50, M10 or M5 response round to the nearest 5%.
2. all persons have latent subjective probabilities for events, and the cross-sectional distribution of beliefs is locally uniform. That is, for each nonextreme value x that is a multiple of 5, similar fractions of persons believe there to be an $x\%$ and $(5 + x)\%$ chance that the event will occur.

In order to test this joint hypothesis it should be noted that the response categories 50, M10 and M5 consist of respectively one, eight and ten values. If the hypothesis were to hold the ratio between the fractions of responses in the 50, M10 and M5 categories should be approximately 1:8:10. Looking at table 1 it becomes evident that this is not the case. M10 and 50 responses occur much more often than the stated ratio with M10 responses occurring at least twice as often as M5 responses and 50 responses occurring at least as frequently as M5 responses for almost every question. Based on Student's t-tests with a 5% significance level the joint hypothesis is rejected. Manski and Molinari (2010) argue that it is plausible that the non-extreme subjective probabilities vary smoothly across the population and that the distribution of these probabilities is thus locally uniform. They therefore maintain part two of the joint hypothesis as discussed above as credible assumption and reject part one of the hypothesis. Given the rejection of the first part of the hypothesis, and the observed fractions in table 1, it appears to be the case that the response category a response belongs to is an indication of the degree of rounding.

3.2 Rounding tendencies respondents

The analysis above is based on the response to specific questions and gives an indication that the degree of rounding varies between responses. To determine whether respondents also differ systematically in their degree of rounding table 2 and 3 describe the response pattern of the respondents over all the questions in the expectation module of the 2006 HRS and 2008 HRS.

This section is a replication of table 2 of Manski and Molinari (2010), again applied to both the 2006 HRS and 2008 HRS. Manski and Molinari (2010) indicate that there were 39 probabilistic questions in the 2006 HRS, of which 38 were used for the analysis of the response patterns. These numbers differ from the actual number of probabilistic expectations questions (41). Preliminary analysis showed that Manski and Molinari (2010) most probably omitted the last two questions. These questions, 100 and 101, both ask for the probability that the

Table 2: Respondent tendencies in the 2006 HRS, all items (38 in total)

		Response pattern						
	Sample size	All NR	All 0 or 100	All 0, 50 or 100	Some M10	Some M5	Some 1-4 or 96-99	Some other
				Males				
All	6775	**0.022**	0.014	0.019	0.252	0.532	**0.118**	0.041
Age 50-54	595	0.008	0.005	0.010	0.224	0.556	0.145	0.052
Age 55-59	966	0.014	0.016	0.013	0.234	0.551	0.130	0.041
Age 60-64	859	0.012	0.012	**0.020**	0.231	0.559	0.125	0.043
Age 65-69	1396	0.019	0.012	0.018	0.268	0.519	**0.125**	0.040
Age 70-74	1147	0.019	0.014	0.019	0.272	0.536	0.104	0.036
Age 75-79	823	0.024	0.013	0.012	0.273	0.524	0.111	0.043
Age 80-84	590	0.031	0.019	0.042	0.232	**0.541**	0.105	0.031
Age 85+	399	**0.085**	0.035	0.035	0.261	0.434	0.093	0.058
				Females				
All	9899	**0.026**	0.025	**0.026**	0.270	0.501	**0.115**	**0.036**
Age 50-54	987	0.013	0.014	0.013	0.272	0.502	0.149	0.037
Age 55-59	1443	0.015	0.019	0.015	0.246	0.525	0.141	0.039
Age 60-64	1434	0.009	0.017	0.010	0.230	0.550	0.136	0.048
Age 65-69	1833	**0.025**	0.018	**0.028**	0.291	0.482	**0.124**	0.032
Age 70-74	1511	0.026	0.019	0.025	0.273	0.517	0.104	0.036
Age 75-79	1069	0.019	0.029	0.027	0.295	0.511	0.087	0.033
Age 80-84	817	0.045	0.038	**0.045**	0.285	0.476	0.076	0.034
Age 85+	805	0.088	0.068	**0.067**	0.282	0.402	0.066	0.026

Note: NR= nonresponse, M10 = multiple of 10 but not 0, 50 or 100, M5 = multiple of 5 but not of 10, bold printed is statistically different between 2006 HRS and 2008 HRS respondent would move back to Mexico. As moving back to Mexico does not fall into any of the three categories as used by Manski and Molinari (2010), they were omitted from the analysis. This yielded results almost identical to those of Manski and Molinari (2010). Therefore, questions 100 and 101 will be omitted from any further analysis. The remaining difference of one between the number of probabilistic expectations questions in the 2006 HRS and the number of questions used is due to the first question about the weather. This question was only used to create a mindset of probabilities. It will therefore also not be used in further analysis.

The response pattern categories of table 2 and table 3 are similar to the response categories of table 1. However, as all questions are incorporated, the response pattern categories are mutually exclusive. The categories go from the most rounded response patterns on the left to the least rounded response patterns on the right. Respondents are categorized according to the algorithm discussed in section 2.1.2 of the literature review.

It should be noted that there are minor differences between table 2 of this paper and table 2 of Manski and Molinari (2010). The same methods were used to categorize the respondents, but a change was made to the 2006 HRS data. The gender of at least one respondent in the age 50-54 range was altered from female to male. Both the All and the Age 50-54 rows for males and females therefore differ slightly from those in table 2 of Manski and Molinari (2010). As the other rows are however also different than those of table 2 of Manski and Molinari (2010) more changes must have been made. Given that the gender of one person was changed, the age or gender of more respondents could have been altered. As the number of respondents per gender and age group is identical to those reported by Manski and Molinari (2010), these changes should however be such that at the group level they canceled out, which is possible but not very

Table 3: Respondent tendencies in the 2008 HRS, all items (25 in total)

		Response pattern						
	Sample size	All NR	All 0 or 100	All 0, 50 or 100	Some M10	Some M5	Some 1-4 or 96-99	Some other
				Males				
All	6360	**0.016**	0.019	0.021	0.286	0.519	**0.101**	0.037
Age 50-54	226	0.018	0.013	0.018	0.235	0.535	0.124	0.058
Age 55-59	1043	0.010	0.012	0.010	0.230	0.580	0.113	0.045
Age 60-64	774	0.012	0.009	**0.006**	0.221	0.602	0.107	0.043
Age 65-69	1252	0.014	0.019	0.018	0.314	0.503	**0.099**	0.033
Age 70-74	1172	0.013	0.013	0.019	0.328	0.500	0.099	0.028
Age 75-79	867	0.017	0.013	0.028	0.291	0.513	0.105	0.033
Age 80-84	592	0.034	0.041	0.039	0.314	**0.459**	0.083	0.030
Age 85+	434	**0.032**	0.062	0.053	0.325	0.403	0.081	0.044
				Females				
All	9359	**0.022**	0.032	**0.019**	0.292	0.506	**0.100**	**0.030**
Age 50-54	597	0.022	0.015	0.013	0.228	0.566	0.119	0.037
Age 55-59	1378	0.015	0.017	0.009	0.243	0.539	0.129	0.047
Age 60-64	1279	0.009	0.016	0.011	0.242	0.567	0.117	0.038
Age 65-69	1743	**0.011**	0.020	**0.011**	0.332	0.499	**0.099**	0.028
Age 70-74	1573	0.020	0.034	0.020	0.302	0.513	0.086	0.025
Age 75-79	1169	0.023	0.036	0.029	0.324	0.478	0.086	0.023
Age 80-84	790	0.035	0.057	**0.023**	0.320	0.466	0.078	0.020
Age 85+	830	0.064	0.083	**0.045**	0.324	0.389	0.076	0.019

Note: NR= nonresponse, M10 = multiple of 10 but not 0, 50 or 100, M5 = multiple of 5 but not of 10, bold printed is statistically different between 2006 HRS and 2008 HRS likely. A different explanation is that the answers of some of the respondents were changed, which resulted in them having a different response pattern. A last possibility is that in fact questions 100 and 101 were not omitted from the analysis by Manski and Molinari (2010). This would however imply that two other questions were deleted and there is no indication which ones this could be. The difference between table 2 and the results of Manski and Molinari (2010) is however so small that it does not have a large effect on the resulting analysis.

Table 2 and 3 show similar general trends. A small fraction (1,5% to 3%) of the respondents do not answer any questions. This is in accordance with the small fraction of nonresponse answers seen in table 1. Approximately similar, small, fractions of respondents only answer with 0 or 100 and 0, 50 or 100. Females are slightly more likely to have these response patterns than males. Age does not seem to have an effect on the fractions until age 80. Above age 80 the response patterns All NR, All 0 or 100 and All 0, 50 or 100 are observed more frequently.

As expected due to the results in table 1, the majority of respondents fall in the Some M10 and Some M5 categories with respectively approximately 25% and 50% of the respondents falling in these categories. Around 10% of all respondents respondent with a value in the 1-4 or 96-99 range at least once and 3% to 4% of the respondents belong to the Some other category.

One would expect that due to the relatively big difference in the number of questions of the 2006 HRS and 2008 HRS, there would be a significant difference in the response patterns. By construction, if more questions are asked a respondent can only be categorized in a stricter category, i.e., if a respondent is categorized in the Some M10 category while using ten questions, incorporating an eleventh question would results in the respondent either remaining in the Some M10 category or moving to either the Some M5, Some 1-4 or 96-99 or Some other category. The respondent could never move to the All NR, the All 0 or 100 or the All 0, 50 or 100 category. One would therefore expect the fractions of the response patterns of the categories on the left of the table to be higher for the 2008 HRS relative to the 2006 HRS. For the categories on the right side of both tables one would expect the opposite. This difference is however not observed as the differences between the 2006 HRS and 2008 HRS are very small and not always in the expected direction. To further test this, the 2008 HRS data was restricted to those respondents who were also in the 2006 HRS. This restriction however did not change the response patterns in a systematic way. The differences between the 2006 HRS and 2008 HRS remained similar.

The proportions which were significantly different between the 2006 HRS and 2008 HRS at a 5% significance level are again printed in bold. A total of 13 out of 126 proportions were significantly different. Given the fact that on average 6 or 7 would be falsely categorized as such if in reality none were, and the fact that the differences are not large, one can conclude that the results for the 2006 HRS and 2008 HRS are very similar.

From the preliminary analysis of the data, it can be concluded that respondents differ systematically in their response patterns. These response patterns can be used to obtain intervals using the algorithm of Manski and Molinari (2010) as described in section 2.1.2.

3.3 Subjective survival probability

As mentioned, the subjective survival probability will be used extensively in the analysis. The HRS contains two questions on the subjective survival probability. One adjusts the age to survival as the age of the respondent increases, whereas the other one is fixed for all respondents. Following the analysis of Manski and Molinari (2010) the research will focus on the second subjective survival probability as obtained through question 28, which is formulated as follows:

"What is the percent chance that you will live to be 75 or more?"

This question was only asked to respondents aged 65 or less, resulting in 6713 and 5567 respondents for the 2006 HRS and 2008 HRS respectively. Figure 1 shows the responses.

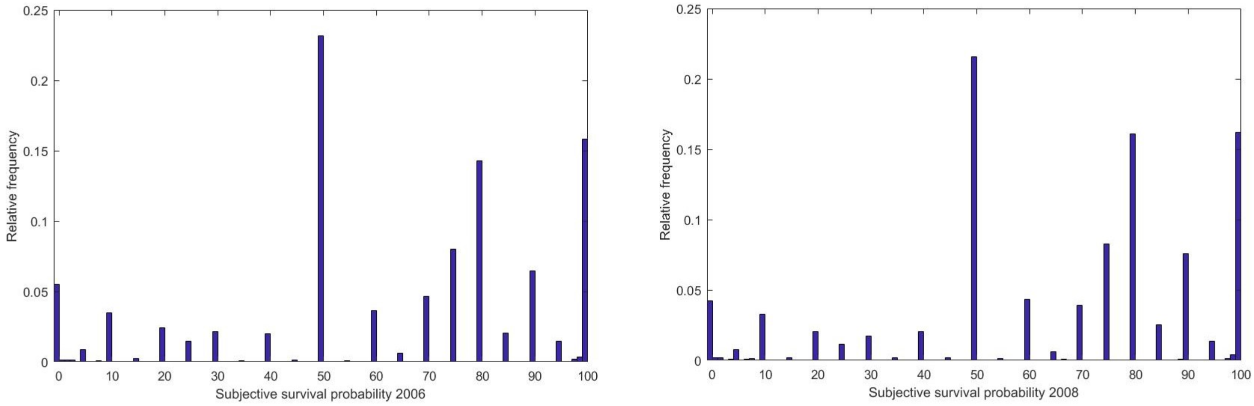


Figure 1: Answers to question 28 of the 2006 HRS (left) and 2008 HRS (right)

The majority of the responses are multiples of 10, with large proportions of respondents answering 50, 80 and 100. To analyze these responses further, the next subsection looks into probing follow-up questions as asked in the 2008 HRS.

3.3.1 Probing for rounding and focal point answers

As discussed in the literature review, the degree of rounding differs between respondents. The analysis in the previous sections has shown how Manski and Molinari (2010) use the response pattern of respondents to determine to what degree they round. Assessing whether this degree of rounding is in fact accurate is however not possible without acquiring the actual degree of rounding. As the 2006 HRS did not contain any probing questions, Manski and Molinari (2010) posed the subjective survival question and follow-up questions to the respondents of the American Life Panel (ALP). The ALP is an Internet survey of American adults. It is important to note that the samples of the ALP and the HRS are very different. The follow-up questions asked whether or not a person rounded, and if he or she did round, to what degree the respondent rounded. As described, the 2008 HRS did include questions to determine the degree of rounding. Follow-up questions were asked after questions 17, 28, 29 and 47. The focus of this section will be on question 28, but the results for the other questions are similar.

The way follow-up questions were asked, and their precise formulation is shown in Appendix B. Several conclusions can be drawn from the answers to the probing questions. First of all, approximately 50% of the respondents stated that their answers were exact. For the M10 respondents this percentage is significantly lower, at around 30%. Furthermore, respondents who answered 50 were randomly assigned to two different probing questions. One of these probing questions asked whether they believed the probability of surviving was equal to the probability of dying or whether they were unsure. The other question asked whether their response was an exact answer or an approximation. As the respondents were randomly assigned one would expect the proportion of respondents answering that the probability was equal, to be approximately the same as the proportion of respondents answering that their answer was exact. This should be the case as both these answers indicate an exact 50% survival probability. The actual proportions for these two answers however differ significantly. This implies that the way a probing question is asked has a big impact on the answers.

Second of all, the respondents who indicated that their response of 0 or 100 was an approximation, gave highly unlikely exact answers. Almost all exact answers given are multiples of 5, indicating that these are probably still rounded. Furthermore, approximately 25% of the people stating that their subjective survival probability was 0%, indicated that their exact subjective survival probability was in fact greater than 50% after being probed. These results show that probing respondents for exact probabilities is troublesome as it cannot be determined whether the original answer or the answers to the probing questions are more accurate.

Lastly, to assess the degree of rounding, the responses to the questions asking for a range were analyzed. It should first of all be noticed that the proportion of non-responses to these questions was significantly higher than the proportion of non-responses to the other probing questions. 10% to 15% of the respondents refused to give a range. Of the people who gave a range, approximately 10% gave a range such that their initial answer was outside the bounds of the range. The high non-responses proportions and the illogical ranges given could indicate that people find it difficult to assess their own degree of rounding, and give ranges. Nevertheless, the ranges were analyzed to determine the average width of the intervals as shown in figure 2. The average width of the intervals was approximately 20% and the average widths of the response pattern categories do not differ significantly at a 5% significance level based on Student's t-tests.

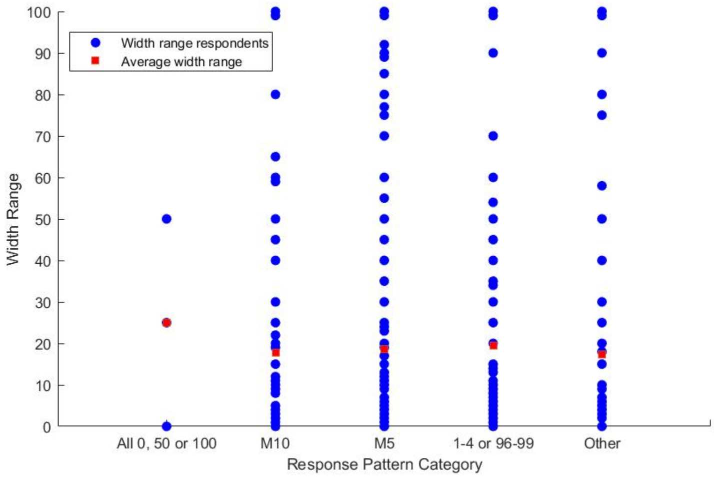


Figure 2: Width of the ranges given after probing

Given the responses to the probing questions, not much can be concluded about the degree of rounding. The algorithm used by Manski and Molinari (2010) will therefore be assumed to be correct. Further research on how to acquire more accurate rounding degrees is required.

3.4 Wealth, age and gender

The data on wealth was acquired through the RAND HRS. The HRS itself does not include a question which asks for the total wealth of the respondents. It does however include a series of questions on the assets of the respondents. These concern among others savings, debts, real-estate, stocks, etc.. The RAND HRS combines the answers of all these questions to determine total wealth. The wealth of the respondents ranges from -\$85,000 to \$3,645,000 with a mean value of \$379,030 and a median of \$145,590. Wealth will be used and reported per \$1,000.

Age and gender were acquired through the HRS. The number of respondents in the various age groups can be seen in table 2 and 3. As the analysis focuses on the subjective survival probability, figure 3 shows the distributions of age for the sample that was asked question 28.

Given the setup of the HRS and the fact that question 28 was only asked to respondents younger than 65, the mass of the age distributions lies between 51 and 65. The HRS was originally setup to be representative for the 51-61 age group. However due to the fact that many people of the previous waves are still in the sample, and new people are added in each wave to ensure representativeness in the 51-61 age group, the age range for which the HRS is representative increases with every wave. As seen in the age distributions, the 2006 HRS and 2008 HRS are therefore representative for the age group 51-65. To ensure representativeness

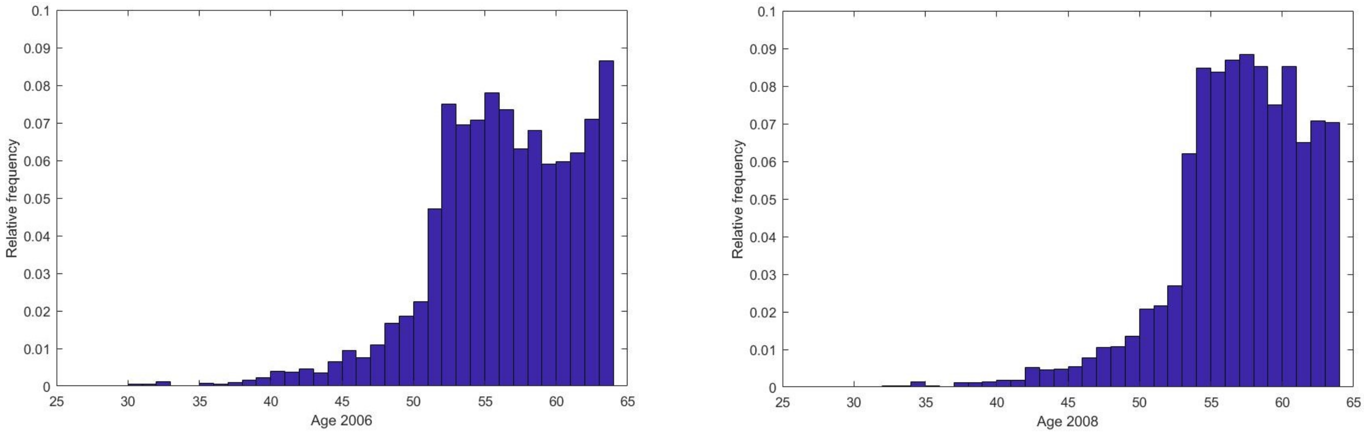


Figure 3: Distribution of age for the sample of the subjective survival probability question

one should omit the respondents below age 51 from the sample. This is however not done by Manski and Molinari (2010) and as the primary goal of this paper is to replicate their results and extend on the methodology, it will not be done for the analysis in this paper either. The analysis sample is therefore not representative for any age group.

The 2008 sample consists of 2496 males and 4217 females. The 2006 sample consists of 2091 males and 3476 females. This could indicate that females are overrepresented in comparison to the population. To examine this one should take the average longevity differences between men and women into account. For the age group 45-65 the ratio of men to woman in the USA is however approximately one (Howden & Meyer, 2010). The difference in longevity only starts to influence this ratio after the age of 65. Therefore the ratio between men and woman in the sample should also be approximately one. As this is clearly not the case, the analysis sample is not representative for the entire population.

4 Methodology

The main analysis concerns Best Linear Prediction (BLP) under square loss as shown in equation (2).

$$y = x'\beta + \epsilon \tag{2}$$

In which y is the dependent variable, x is a vector of k independent variables and β are the coefficients. In case both the dependent and independent variables would be properly defined OLS could be used to estimate the parameters using the following equation:

$$\beta = [E(x'x)]^{-1}E[x'y] \tag{3}$$

However, once either the dependent or one of the independent variables is interval data different methods need to be used as the estimator in equation (3) is no longer point identified.

Sections 4.1 and 4.2 discuss these methods for regressions with interval-data as dependent and independent variable respectively. Section 4.3 examines the diagnostics which will be used.

4.1 Interval-data as dependent variable

If y in equation (2) is only known to lie in an interval (y_L, y_U), β is no longer point identified. Instead, there exist a certain identification region $H[\beta]$ in which the true β lies. Beresteanu and Molinari (2008) propose a method to estimate this identification region. They show that the identification region is a set of parameter values that solve the following set of equations:

$$H[\beta] = \{b \in \mathbb{R}^k : b = [E(x'x)]^{-1}E[x'\tilde{y}] \quad , \quad \tilde{y} \in [y_L, y_U] \text{ with probability } 1\} \tag{4}$$

Instead of having to consider all points in (y_L, y_U), Beresteanu and Molinari (2008) show that equation (4) can be estimated for each individual parameter β_d using the residuals $\tilde{x}_{id}$, obtained after projecting x_d on the other independent variables x, as follows:

$$\hat{H}_n[\beta_d] = \frac{1}{\sum_{i=1}^{n} \tilde{x}_{id}^2} \left[\sum_{i=1}^{n} min\{\tilde{x}_{id}y_{iL}, \tilde{x}_{id}y_{iU}\}, \sum_{i=1}^{n} max\{\tilde{x}_{id}y_{iL}, \tilde{x}_{id}y_{iU}\} \right] \tag{5}$$

In which n is the sample size. The intuition of this equation is as follows; x_d is first projected on the other covariates to filter out any common effect. Once the common effect is filtered out, it is sufficient to estimate the following equation:

$$y = \tilde{x}_d' \beta_d + \epsilon_d \tag{6}$$

In which β can be estimated through equation (4) while replacing x with $\tilde{x}_d$ as follows:

$$H[\beta_d] = \{b_d \in \mathbb{R}^1 : b_d = [E(\tilde{x}_d'\tilde{x}_d)]^{-1}E[\tilde{x}_d'\tilde{y}] \quad , \quad \tilde{y} \in [y_L, y_U] \text{ with probability } 1\} \tag{7}$$

To obtain the bounds of $H[\beta_d]$ it should be noted that the first part of equation (7), $[E(\tilde{x}_d'\tilde{x}_d)]^{-1}$, is always positive and can be estimated through the fraction in equation (5). Given that the first part of equation (7) is always positive, the lower bound of the identification region is thus obtained by some $\tilde{y}$ which minimizes the second part of equation (7), $E[\tilde{x}_d'\tilde{y}]$. This second part can be estimated as $\sum_{i=1}^{n} \tilde{x}_{id}\tilde{y}_i$. Therefore one has to find $\tilde{y}_i$ which minimizes $\tilde{x}_{id}\tilde{y}_i$ for every i. As $\tilde{x}_{id}\tilde{y}_i$ is linear in $\tilde{y}_i$, the minimum of $\tilde{x}_{id}\tilde{y}_i$ always occurs at one of the extremes of (y_{iL}, y_{iU}). Which one of the extremes it is depends on the sign of $\tilde{x}_{id}$. The minimum of $E[\tilde{x}_d'\tilde{y}]$ can thus be estimated through $\sum_{i=1}^{n} min\{\tilde{x}_{id}y_{iL}, \tilde{x}_{id}y_{iU}\}$. Combining this with the estimator of the first part of equation (7) leads to the formula of the lower bound of $H[\beta_d]$ as shown in (5).

To obtain the upper bound of the identification region similar reasoning can be followed whilst maximizing instead of minimizing the second part of equation (7). The mathematical proof of the fact that the estimated identification region converges in probability to the population identification region can be found in the appendix of Beresteanu and Molinari (2008).

4.2 Interval-data as independent variable

If one of the independent variables in equation (2), denoted as v from here on, is known to lie in a certain interval (v_L, v_U), similar reasoning holds as in the case of interval-data as dependent variable as discussed above. β is no longer point identified, but there exists an identification region $H(\beta)$ in which the true β lies. The estimation method for interval-data as dependent variable as proposed by Beresteanu and Molinari (2008) can however not be used to estimate the identification region. The reason for this is that given the dependent variable and the point identified regressors, equation (4) is not linear in v.

Manski and Tamer (2002) therefore propose a Modified Minimum Distance (MMD) estimator for regressions with interval data as independent variable. This estimator requires the following three assumptions:

- Interval (I): $P(v_L \leq v \leq v_U) = 1$
- Monotonicity Parametric Regression (MPR): For each value of x and each $b \in H(\beta)$, $E(y|x, v)$ is weakly increasing in v
- Mean Independence (MI): $E(y|x, v, v_U, v_L) = E(y|x, v)$

Assumption I implicates that the true value v is always within the bound (v_L, v_U). Assumption MPR forces the coefficient of v to be positive. However, if this is not the case one can however multiply the regressor v with -1. Therefore the assumption only ensures that the linear specification of BLP is correct. Lastly, assumption MI states that, given that the true value v is known, the bounds v_L and v_U have no effect on the expected value of y.

Given these assumptions and the fact that BLP has the monotone index form, Manski and Tamer (2002) show that the identification region satisfies the following inequalities:

$$\hat{H}_n[\beta] = \{b \in H[\beta] : f(x_i, v_{iL}, \beta) \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq f(x_i, v_{iU}, \beta), i = 1, ..., N\} \tag{8}$$

In which $\hat{\eta}_N(x_i, v_{iL}, v_{iU}) = E(y|x, v_L, v_U)$ is a consistent estimator. The proof of the fact that these inequalities ensure identification is beyond the scope of this paper and can be found in Manski and Tamer (2002).

To solve the inequalities in (8) Manski and Tamer (2002) propose an estimator which minimizes the distance between $\hat{\eta}_N(x_i, v_{iL}, v_{iU})$ and either $f(x_i, v_{iL}, \beta)$ or $f(x_i, v_{iU}, \beta)$. Cerquera, Laisney, and Ullrich (2014) however show that the estimation method of Manski and Tamer (2002) is outperformed by a geometric approach. This geometric approach uses the eight combinations of the maximums and minimums of x, v_L and v_U to obtain the intersections of all triples of planes defined by these eight pair of restrictions. By taking the convex hull of all of the intersections which satisfy all inequalities, the identification region is estimated. As this method outperforms the method of Manski and Tamer (2002), it will be used in this paper. To estimate the identification region, $\hat{\eta}_N(x_i, v_{iL}, v_{iU})$ will be estimated through OLS estimation. Three of the eight planes are combined at random to obtain the intersections. If the intersections satisfy the remaining five restrictions, they are in the identification region. this process is repeated until all triples of restriction combinations are tested.

4.3 Diagnostics

To assess the estimated models several diagnostics will be used. The parameter estimates will be evaluated based on their confidence intervals (CIs), and the entire model will be evaluated based on in- and out-of-sample performance. In-sample-performance will be based on R-squared and information criterion such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Out-of-sample performance will be evaluated through the Mean-Squared-Prediction-Error (MSPE). For partially identified models different techniques are required to calculate these diagnostics. These techniques will be elaborated on in the following sections. As a last extension to the diagnostics, the empirical distribution of the parameter estimates will be computed in order to acquire (possibly) stricter CIs.

4.3.1 Confidence intervals for partially identified models

To estimate the CI of a parameter, one has to acquire the distribution of the estimators. Under the normality assumption of the error terms the parameter estimators of OLS are normally distributed and thus the CI can be estimated in a straightforward fashion.

Given that the distribution of the parameter estimates of a partially identified model are known, two different approaches are proposed in recent literature to construct the CI. The first approach, proposed by Imbens and Manski (2004) and Stoye (2009), constructs CIs that (asymptotically) uniformly cover each point in the identification region with a prespecified probability. The second approach, proposed by Chernozhukov, Hong, and Tamer (2007), Horowitz and Manski (2000) and Beresteanu and Molinari (2008), constructs CIs that (asymptotically) cover each point in the identification region with a prespecified probability. Following the analysis of Manski and Molinari (2010), the first approach will be implemented. To do so, Imbens and Manski (2004) propose the following CI:

$$\hat{CI_\alpha^\beta} = \left[\hat{\beta}_L - \hat{C}_N\hat{\sigma}_L/\sqrt{N}, \hat{\beta}_U + \hat{C}_N\hat{\sigma}_U/\sqrt{N}\right] \text{ s.t. } \Phi\left(\hat{C}_N + \sqrt{N}\frac{\hat{\beta}_U - \hat{\beta}_L}{max(\hat{\sigma}_L, \hat{\sigma}_U)}\right) - \Phi(\hat{C}_N) = \alpha \quad (9)$$

In which $\hat{\beta}_L$ and $\hat{\beta}_U$ are the lower and upper bound of the identification region and $\hat{\sigma}_L$ and $\hat{\sigma}_U$ the variances of these lower and upper bound parameter estimates.

The intuition behind this CI is that, as the difference between $\hat{\beta}_L$ an $\hat{\beta}_U$ increases relative to the variance of the bounds, $\hat{\sigma}_L$ and $\hat{\sigma}_U$, the true parameter value β is increasingly likely to be within ($\hat{\beta}_L$, $\hat{\beta}_U$). Asymptotically, the difference between the upper and lower bound is large relative to the variances of the bounds, and therefore noncoverage risk becomes one-sided at either bounds and vanishes otherwise. Therefore asymptotically, a 90% CI for the identification region is a 95% CI for the true parameter value (Stoye, 2009).

Acquiring the distribution of the parameter estimators of a partially identified model is however troublesome. Therefore, common practice is to use a bootstrap approach to estimate the CI (Beresteanu & Molinari, 2008; Manski & Molinari, 2010). The bootstrap algorithm estimates one-sided 95% CIs for both the identification region bounds, resulting in a 90% CI for the identification region. The result of Stoye (2009) however ensures that the CI for the true parameter value has at least 95% coverage, as long as the difference between the bounds is large relative to the variance of the bounds. The bootstrap algorithm can be found in Appendix A.

As the CIs as described above are only calculated using the bounds of the interval, they are conservative (Imbens & Manski, 2004). To assess how conservative they are the empirical distribution of the parameters will be evaluated, as will be discussed in the following section.

4.3.2 Empirical distribution diagnostics and parameters

To evaluate the in-sample performance of a model, the most commonly used criteria are R-squared, AIC and BIC. Models used for predictive purposes are often assessed out-of-sample as well, for example through the MSPE. As these diagnostics are based either on the error terms or likelihood of a model, they cannot be computed directly for partially identified models.

A bootstrap approach will be used to obtain an empirical distribution of the diagnostics. To do so, one however has to assume some distribution for the interval data. This means that if the variable is defined on an interval, one has to assume that the true value of the variable comes from this interval based on a certain distribution. As the interval data in this case concerns subjective probabilities, it seems plausible that the distribution is uniform. As far as known to the author, an approach like the one proposed in this paper has not yet been published. No research has been done on the possible distributions of the interval range. A uniform distribution implies that every value in the interval range has an equal probability of being the true value of the variable. A further point of research is to compare various distributions. Distributions of special interest are the truncated normal distribution and heavy-tailed distributions. It is expected that more weight at the tails results in wider ranges of the diagnostics as extreme outcomes are more likely to occur. The assumption of a distribution also makes it possible to obtain the empirical distribution of the parameters and their CIs.

Given the assumed distribution of the interval defined variable, the empirical distribution of the diagnostics and parameters will be calculated using the following algorithm:

1. Split up the data set in an estimation and evaluation part
2. (a) If the interval-data is the dependent variable y, draw y_i^* from $U(y_{iL}, y_{iU})$ for each i=1, ... n with U the uniform distribution
   (b) If the interval-data is the independent variable x_d, draw x_{di}^* from $U(x_{diL}, x_{diU})$ for each i=1, ... n with U the uniform distribution
3. Estimate the parameters using OLS on the estimation part of the bootstrap data. Acquire the R-squared, AIC, BIC, MSPE, parameter estimates and CIs for this instance

4. Repeat step 2 and 3 b times to acquire the empirical distribution of the diagnostics, parameter estimates and CIs

The obtained empirical distribution of the parameters and CIs can be used to reassess the effect rounding has on regressions. The empirical distributions of the other diagnostics can be used to compare models, for example by evaluating best and worst case scenarios, mean diagnostic values and the overall shape and location of the distributions.

By construction, the empirical CIs acquired through the bootstrap algorithm are not robust. They are computed by taking the maximum and minimum values of all of the 95% CIs obtained per repetition. This implies that if the draws from the uniform distributions are extreme for one of the repetitions, the CI for this repetition is extreme and thus the empirical CI obtained is affected strongly. The CI of the extreme repetition will become one of the bounds of the empirical CI. In order to make robust empirical CIs, one might consider using bigger CIs per draw and instead of taking the maximum and minimum of these CIs, build a CI such that an x% of the larger CIs of the draws is within the bounds. An example of this would be to acquire 98% CIs at step 3 of the algorithm. By then looking for bounds such that 97% of all the CIs are within the bounds, a 95,06% empirical CI is created. Different combinations of the confidence levels in step 3 (in the example 98%) and the percentage of CIs within the bounds (in the example 97%) can be compared to determine the most strict, and robust empirical CI.

The analyses in this paper will use the conservative empirical CIs obtained through combining all of the 95% CIs of step 3.

5 Empirical analysis

The empirical analysis is structured as followed. In section 5.1 the regression analysis of Manski and Molinari (2010) is replicated for both the 2006 HRS and 2008 HRS. In section 5.2 probability weighting is implemented on the regressions analysis of section 5.1 and the theoretical estimates of the parameter identification regions and of the CIs are compared to the empirical distributions obtained through the bootstrap algorithm. Section 5.3 will look into the implications of using probability weighting and the bootstrap algorithm if the independent variable is interval-data. To do so, the theoretical parameter identification regions and CIs are compared to the empirical estimates obtained through the bootstrap algorithm. The predictive power of the subjective survival probability with and without probability weighting will be compared through the empirical distributions of the diagnostics.

5.1 Interval-data as dependent variable

Table 4 reports the regression results for the 2006 HRS and 2008 HRS. The subjective survival probability is regressed on age and gender. To assess the effect of rounding, the first column reports the OLS results for which the responses are taken at face value. Columns A-E report the results for which the responses are first transformed to interval data using the algorithm proposed by Manski and Molinari (2010). The results of the 2006 HRS in table 4 are a replication of the results of table 4 of Manski and Molinari (2010).

The differences between columns A-E depend both on the selection of questions used to determine the response pattern and on the sample used. Column A and B both use the entire sample of respondents. However, column A uses all questions to determine the response pattern whereas column B only uses the questions on personal health. As described in the data section, using more questions leads to stricter response patterns and therefore more narrow intervals.

Table 4: Point estimates vs set estimates conditioning on age and gender in the 2006 HRS and 2008 HRS

	HRS2006										
		Set estimates**									
	Point estimates*	(A)		(B)		(C)		(D)		(E)	
	MCAR imposed	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
Age	0.252	-0.969	1.369	-0.506	0.944	-0.182	0.647	-0.167	0.604	-0.147	0.573
	(0.114, 0.390)	(-1.076, 1.477)		(-0.628, 1.060)		(-0.296, 0.755)		(-0.286, 0.715)		(-0.262, 0.685)	
Male	-5.481	-19.648	9.984	-14.154	3.915	-10.910	-0.113	-10.669	-0.645	-10.340	-0.991
	(-6.959, -4.003)	(-20.806, 11.218)		(-15.412, 5.126)		(-12.123, 1.094)		(-11.871, 0.610)		(-11.613, 0.315)	
Const.	50.987	-13.483	119.042	11.012	93.056	28.435	75.510	31.063	74.785	33.002	73.776
	(43.121, 58.852))	(-20.016, 125.167)		(4.306, 99.856)		(22.259, 82.114)		(24.616, 81.527)		(26.397, 80.295)	
N	6442	6713		6713		6442		6385		6313	

	HRS2008										
		Set estimates**									
	Point estimates*	(A)		(B)		(C)		(D)		(E)	
	MCAR imposed	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
Age	0.362	-0.601	1.248	-0.348	1.009	-0.071	0.775	-0.045	0.750	-0.033	0.721
	(0.205, 0.519)	(-0.741, 1.393)		(-0.487, 1.163)		(-0.206, 0.919)		(-0.179, 0.890)		(-0.168, 0.867)	
Male	-5.079	-16.497	7.579	-13.642	4.251	-10.272	0.436	-9.945	0.098	-9.588	-0.060
	(-6.644, -3.514)	(-17.680, 8.929)		(-14.966, 5.578)		(-11.609, 1.743)		(-11.249, 1.436)		(-10.886, 1.225)	
Const.	46.635	-4.784	100.344	9.005	86.026	22.887	71.028	24.390	69.624	26.099	68.967
	(37.627, 55.643)	(-13.478, 108.330)		(-0.241, 93.889)		(14.297, 78.835)		(16.394, 77.325)		(17.448, 76.684)	
N	5354	5567		5567		5354		5313		5268	

Note: *95% CIs based on normal approximation, **95% Imbens-Manski CIs based on the bootstrap with 1000 replications, (A) rounding tendencies based only on personal health questions, (B) response tendencies based on all questions, (C) rounding tendencies based on all question and nonrespondents discarded, (D) rounding tendencies based on all question and nonrespondents and respondents with all 0 and 100 discarded, (E) rounding tendencies based on all question and nonrespondents and respondents with all 0 and 100 or all 0, 50 and 100 discarded.

Columns C, D and E use all questions to determine the response patterns. However, certain groups of respondents are omitted from the analysis. Column C omits the respondents who did not reply to question 28. Column D omits both the respondents who did not reply to question 28 and those who had the response pattern All 0 or 100. Lastly, Column E omits the respondents who did not respond to question 28, the respondent with response pattern All 0 or 100 and the respondents with response pattern All 0, 50 or 100. As discussed in the data section, deleting these respondents could improve the analysis as uninformative focal point answers are omitted. However, informative answers could also be deleted.

Using the responses at face value, age has a positive, significant effect on the subjective survival probability. Being male on the other hand has a negative and significant effect on the subjective survival probability. Even though the parameter estimates differ slightly between the 2006 HRS and 2008 HRS, the sign and magnitude are comparable. Every year a respondent is older increases the subjective survival probability on average by 0.25% or 0.36% for the 2006 HRS and 2008 HRS respectively. Furthermore, being male reduces the subjective probability on average by 5,5% and 5.1% for the 2006 HRS and 2008 HRS. Lastly, the constant implies that the baseline subjective survival probability is approximately 50%. However, as the majority of the sample is aged 50 and above and there are no respondents aged below 25, a part of the constant could be incorporated in the age coefficient. As age starts at 25 this implies that the constant part of age is approximately 6.3% and 9.0% for the 2006 HRS and 2008 HRS respectively.

The results are as expected. Given that subjective survival probabilities predict actual mortality fairly well, a lower longevity for males should and does result in a negative effect of being male. The effect also seems to be close to the actual difference in mortality of approximately 5% (Howden & Meyer, 2010). The positive effect of age has two possible explanations. First of all, the actual survival probability on average increases with age and as the subjective survival probability predicts actual mortality fairly well it should increase with age as well. Second of all, people with lower subjective survival probabilities have a higher probability of dying. Given that the HRS consists of subsequent waves in which only people in the 51-61 age group are added (and possibly their partners with different ages), it is likely that especially in the 61-65 age group more people with low subjective survival probabilities have died than people with high subjective survival probabilities. This increases the average subjective survival probability of the older age groups causing a positive effect of age on the subjective survival probability. Lastly, the magnitude of the constant should ensure a baseline such that the subjective survival probabilities are approximately equal to actual survival rates. Given that the actual survival rate to age 75 for the age group 51-65 is approximately 75% (Arias, Heron, & Xu, 2016), the constant plus the average effect of age should be close to 75. The regressions show an effect of respectively approximately 65% and 70%, fairly close to the actual survival rates.

Once the responses are no longer taken at face value but transformed to interval data, all results become insignificant. As more questions are used to construct the response pattern, and as those respondents with the broadest intervals are omitted from the analysis the identification regions and CIs become more narrow. However, even in the most strict case of column E age and gender are still insignificant. Given the fact that ordinary analysis taking the subjective survival probabilities at face value leads to the conclusion that the effect of age and gender is significant, the effect of rounding is substantial.

It should be noted that the results for the 2006 HRS in table 4 differ slightly from those in Manski and Molinari (2010). As discussed in section 3.2, the 2006 HRS data used by Manski and Molinari (2010) differs slightly with respect to the data used for this analysis. This might cause the differences in the results of the analysis. Furthermore, as the CIs are calculated through a bootstrap procedure, differences will always arise. As Manski and Molinari (2010) do not report the number of repetitions used for the bootstrap, the accuracy cannot be determined. However, despite the differences, the conclusions drawn from the analysis are similar to those drawn by Manski and Molinari (2010).

5.2 Interval-data as dependent variable with probability weighting

As discussed in the literature review, insights from behavioral economics show that people evaluate probabilities in a non-linear way. The evaluation can be modeled through probability weighting. If subjective survival probabilities are evaluated by people in the same way as the standard probabilities as proposed by Kahneman and Tversky (1979), using probability weighting should improve the predictive power. The effect of more or less severe probability weighting through different values of c in equation 1 has been examined by varying the value of c between 0.5 and 1. The results however showed that as c increased towards 1, the parameter estimates, CIs and diagnostics converged to the estimates of the regressions without probability weighting. The remainder of the analysis will therefore report the results for $c = 0.5$, as this was the case that differed most from the case without probability weighting.

Due to probability weighting the interpretation of coefficients is no longer straightforward. To interpret the coefficients one should take the shape of the weighting function into account as shown in figure 4.

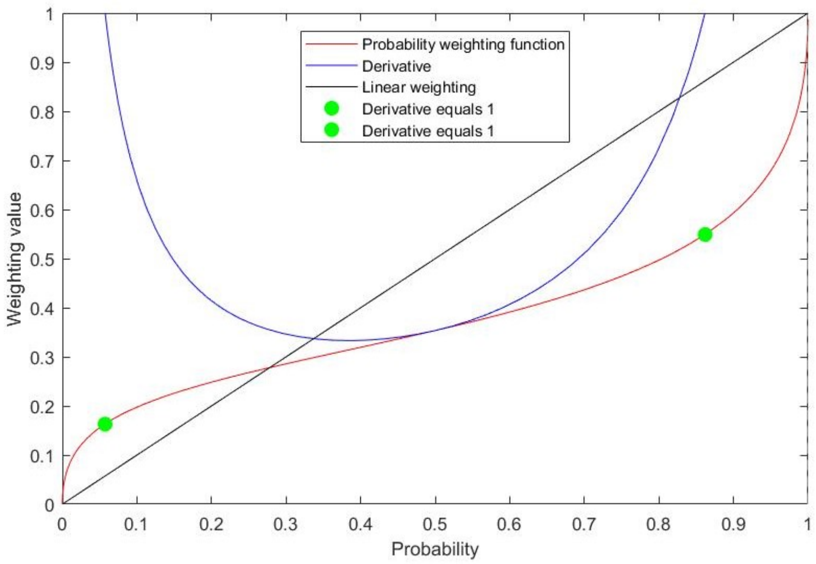


Figure 4: Probability weighting function

As discussed in the literature review, the inverse-S shaped weighting function has a derivative greater than 1 at the extremes. Given $c = 0.5$, increases in probabilities in the ranges 0-5 and 85-100 result in weighting increases of more than the increases in probabilities. In the range 5-85 changes in the probability lead to changes in the weighting of less than the change in probability. The lowest sensitivity is around 40. To illustrate, an increase from 0% to 1% leads to an increase in weight of approximately 8, an increase from 40% to 41% only leads to an increase in weight of 0.3 and an increase from 99% to 100% leads to an increase in weight of 18.

As the analysis of Manski and Molinari (2010) uses the subjective survival probability as dependent variable, nothing can be said about the predictive power. The regression results obtained after regressing the transformed subjective survival probability on age and gender have no real interpretation. Nevertheless, this analysis will be performed in section 5.2.1 to assess the effect of probability weighting on the parameter identification region and the CIs. In section 5.2.2 the bootstrap algorithm will be used to compute the empirical parameter distributions and CIs.

The analyses of the following sections focuses on the point estimation, column B and column E of table 4. Columns A, C and D are not considered. The reasoning for this is twofold. Column A is not considered as the aim of the research is not to analyze the effect of using only questions of a particular category. Columns C and D are not considered as there is no indication that particular focal point answers are more or less informative than other focal point answers. In order to assess the effect of possible focal point answers column E is considered and compared to column B. For the remaining analysis the regressions of column B and E will be referred to as the unrestricted and restricted sample respectively.

5.2.1 Theoretical estimates and confidence intervals

Table 5 shows the theoretical estimates of the parameters and their CIs. The estimates of the regressions without probability weighting are clearly identical to the results of the 2008 HRS in table 4 as they concern the same regressions.

The general results of the regressions do not change after applying probability weighting. For the point estimates age continues to have a positive, significant effect. Being male on the other hand still has a negative, significant effect and the constant is still positive and significant. The set estimates obtained using probability weighting are all insignificant. This implies that the only significant variable of the set estimates without probability weighting, the constant of the regression of the restricted sample, has become insignificant.

Interpreting the magnitude of the coefficients of the regressions with probability weighting is troublesome. If the transformed subjective survival probability increases with 1%, the actual subjective survival probability could increase with more or less than 1%, but it will always increase. The effect on the actual survival probability is the inverse effect of the effect described in the beginning of section 5.2. Changes in the subjective survival probability at the extremes cause changes in the actual subjective survival probability smaller than the original change. The point estimates with probability weighting should therefore be interpreted as follows. The weighted subjective survival probability increases on average with 0.3% for every year a respondent is older. Being male on the other hand decreases the weighted subjective survival probability by 3.8% on average. Given the fact that the constant of 35% and the average effect of age of 1.8% it should be noted that the predicted weighted subjective survival probability is always below 50% and above 20%. This implies that the predicted values are always in the range for which the effect on the actual subjective survival probability are greater than the effect on the weighted subjective survival probability. This corresponds to the coefficients of the regression with probability weighting being of smaller magnitude than the corresponding coefficients in the regression without probability weighting.

The width of the parameter identification regions and the CIs appear to have increased due to probability weighting. However, given that all results are insignificant, no conclusions can be drawn from this.

5.2.2 Empirical estimates of parameter bounds and confidence intervals

Table 6 shows the estimates of the regressions in table 5 as obtained through the bootstrap algorithm. As discussed in the methodology, the theoretical results of table 5 are conservative. The bounds of the parameter identification region are obtained through considering the most extreme combination of upper and lower bounds. If one assumes that the true response of a respondent comes from a uniform distribution with the bounds of the respondent as upper and lower bounds, as done through the bootstrap algorithm, the probability that the combination of lower and upper bounds as used in the theoretical calculation are near the true values, is very close to zero. To illustrate, assume that the probability that one's true answer is close to the bound as used in the theoretical calculation is 10%. The probability that the true responses of all respondents are close to the bounds used in the theoretical calculation is 0.1^N with N the number of respondents. For the unrestricted sample this would be 0.1^{5567} which is extremely close to 0.

Table 5: Comparison point estimation and unrestricted and restricted sample with and without probability weighting

	No Probability Weighting					Probability Weighting				
		Set estimates**					Set estimates**			
	Point estimates*	(B)		(E)		Point estimates*	(B)		(E)	
	MCAR imposed	LB	UB	LB	UB	MCAR imposed	LB	UB	LB	UB
Age	0.362	-0.348	1.009	-0.033	0.721	0.291	-0.580	1.078	-0.286	0.780
	(0.205, 0.519)	(-0.487, 1.163)		(-0.168, 0.867)		(0.148, 0.434)	(-0.701, 1.198)		(-0.397, 0.895)	
Male	-5.079	-13.642	4.251	-9.588	-0.060	-3.844	-14.004	7.557	-9.910	3.4617
	(-6.644, -3.514)	(-14.966, 5.578)		(-10.886, 1.225)		(-5.272, -2.246)	(-15.034, 8.713)		(-10.946, 4.569)	
Const.	46.635	9.005	86.026	26.099	68.967	34.463	-11.955	82.1370	4.8715	65.437
	(37.627, 55.643)	(-0.241, 93.889)		(17.448, 76.684)		(26.246, 42.681)	(-18.883, 88,876)		(-1.784, 71.844)	
N	5354	5567		5268		5354	5567		5268	

Note: *95% CIs based on normal approximation, **95% Imbens-Manski CIs based on the bootstrap with 1000 replications, (B) unrestricted sample (E) restricted sample.

Table 6: Bootstrap estimates of table 5

	No Probability Weighting					Probability Weighting				
		Set estimates**					Set estimates**			
	Point estimates*	(B)		(E)		Point estimates*	(B)		(E)	
	MCAR imposed	LB	UB	LB	UB	MCAR imposed	LB	UB	LB	UB
Age	0.362	0.263	0.403	0.166	0.326	0.291	0.322	0.367	0.198	0.291
	(0.205, 0.519)	(0.110, 0.555)		(0.045, 0.447)		(0.148, 0.434)	(0.168, 0.521)		(0.077, 0.412)	
Male	-5.079	-5.547	-3.886	-4.088	-2.353	-3.844	-5.047	-4.618	-3.667	-2.795
	(-6.644, -3.514)	(-7.057, -2.373)		(-5.287, -1.144)		(-5.272, -2.246)	(-6.587, -3.079)		(-4.876, -1.581)	
Const.	46.635	43.295	51.368	30.496	39.782	34.463	46.199	48.814	32.582	37.963
	(37.627, 55.643)	(34.547, 60.106)		(23.547, 46.729)		(26.246, 42.681)	(37.355, 57.656)		(25.646, 44.907)	
N	5354	5567		5268		5354	5567		5268	

Note: Note: *95% CIs based on normal approximation, **95% empirical CIs based on the bootstrap with 1000000 repetitions, (B) unrestricted sample (E) restricted sample.

Given this illustration it is of no surprise that the bootstrap estimates are much more narrow than the theoretical estimates. The conclusions drawn using the bootstrap estimates are similar to those drawn using point estimation. Given that the results were obtained through one million replications, it seems very plausible that the theoretical estimates are extremely conservative.

To ensure that the bootstrap estimates do not approach the theoretical estimates as the number of repetitions increase, table 7 reports the effect of increasing the number of repetitions for the regression of the unrestricted sample without probability weighting. The results for the other regressions are similar and can be obtained from the author on request.

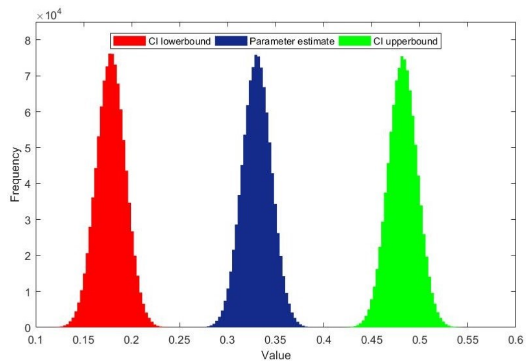
Table 7: Effect of number of repetitions on bounds for unrestricted sample without probability weighting

	Number of bootstrap repetitions									
	100		1,000		10,000		100,000		1,000,000	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
Par. Est. Age	0.293	0.372	0.286	0.376	0.267	0.391	0.263	0.395	0.263	0.403
CI Age*	(0.141, 0.524)		(0.134, 0.528)		(0.115, 0.543)		(0.110, 0.547)		(0.110, 0.555)	
Par. Est. Male	-5.043	-4.265	-5.127	-4.265	-5.401	-4.064	-5.401	-3.986	-5.547	-3.886
CI Male*	(-6.557, -2.753)		(-6.634, -2.748)		(-6.913, -2.554)		(-6.913, -2.481)		(-7.057, -2.373)	
Par. Est. Const.	45.037	49.621	44.945	50.139	43.971	51.165	43.956	51.165	43.295	51.368
CI Constant*	(36.308, 58.359)		(36.202, 58.854)		(35.238, 59.913)		(35.229, 59.960)		(34.547, 60.106)	

Note: UB=Upper Bound, LB = Lower Bound, *95% empirical CIs based on bootstrap

The width of the parameter identification regions approximately double when going from 100 to 1,000,000 repetitions while the width of the CIs increase by at most a quarter. These findings are similar for all of the different regressions of table 6. Relative to the theoretical estimates, the increases are however rather small. The width of the intervals of the parameter identification regions using 100 repetitions are approximately 5% of the width of the theoretical estimates. For 1,000,000 repetitions they are still only 10% of the width of the theoretical estimates. The width of the CIs increases from approximately 20% to approximately 25% of the width of the theoretical CIs. Given the fact that the CIs are not robust and still conservative (as discussed in the methodology) the increases in the widths of the CIs is remarkably small.

The bootstrap algorithm does not only yield the parameter and CI bounds, but also their empirical distribution. Figure 5 shows these empirical distributions of the parameter age for the regressions of the unrestricted and restricted sample without probability weighting. The results for the regressions with probability weighting are comparable.



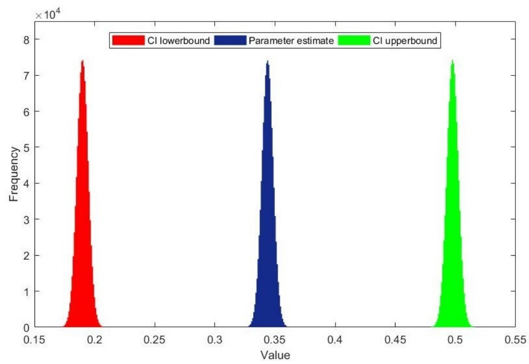


Figure 5: Empirical distribution parameter and CI of age for unrestricted sample (left) and restricted sample (right) without probability weighting

The distributions of the parameter estimate for age and its CIs are shaped like the normal distribution as the kurtoses are approximately three and the skewness is close to zero. Normality was not rejected for some of the empirical distributions based on the Jarque-Bera test. The effect of deleting the respondents with wide interval data (restricted sample) is evident. The means of the parameter distributions and CIs are not significantly different between the regression of the unrestricted and restricted sample at a 5% significance level using Student's t-tests. The variance of the distributions however do differ. Deleting the respondents with wide interval data decreases the variance of the distributions by as much as 90%. However, as discussed in the previous section, even though the variance decreases, the magnitude of the parameters and the conclusions drawn do not change.

5.3 Interval-data as independent variable with probability weighting

As discussed in the literature review, the main reason for using stated preferences in the form of probabilistic expectations is to improve the credibility and predictive strength of revealed preference analysis. The following sections will therefore focus on the predictive strength of the subjective survival probability. The empirical application under consideration is a simplified version of the analyses of Bloom et al. (2006) on the effect of the subjective survival probability on wealth. The only control factor which will be used is age as the basic algorithm developed by Cerquera et al. (2014) incorporates at most 3 regressors, including the constant. Other control factors such as race and martial status as also used by Bloom et al. (2006) can therefore not be incorporated leading to omitted variables. Furthermore, reserve causality is likely but the use of instrumental variable estimation for interval-data is beyond the scope of this paper. The setup of the analyses is similar to the setup of the analyses in the previous sections in terms of the various samples which will be used. Wealth will be regressed on age and the subjective survival probability. Point estimates and set estimation using the unrestricted and restricted samples will be considered. Probability weighting will be implemented on these 3 regressions to assess the effect it has on predictive power.

Section 5.3.1 discusses the theoretical estimates of the parameters and CIs. In section 5.3.2 the bootstrap algorithm will be implemented to compare the theoretical and empirical estimates of the parameters and CIs. Section 5.3.3 will analyze the effect of the number of bootstrap repetitions and examine the distributions of the empirical parameter estimates and empirical CIs. Lastly, section 5.3.4 will compare the various regressions based on predictive strength using the empirical distributions of the diagnostics.

5.3.1 Theoretical estimates

The estimation method of Cerquera et al. (2014) was implemented in R. The program did however not terminate and closer inspection showed that, due to the minima and maxima of the lower and upper bound being identical, the problem was unsolvable. The proof can be found in Appendix C. Manski and Tamer (2002) propose to alter equation (8) as follows:

$$\hat{H}_n[\beta] = \{b \in H[\beta] : f(x_i, v_{iL}, \beta) - \alpha \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq f(x_i, v_{iU}, \beta) + \alpha, i = 1, ..., N\} \quad (10)$$

However, to find a feasible solution in the analysis of this paper, α would have to equal the maximum or minimum value of $\hat{\eta}_N(x_i, v_{iL}, v_{iU})$. Doing so results in the identification region becoming unbounded. The remainder of the results will therefore focus on the empirical estimation method. Further research is required to compare the theoretical and empirical estimates.

5.3.2 Empirical estimates of parameter bounds and confidence intervals

Table 8 shows the estimation results obtained through the bootstrap algorithm. For all of the regressions the general results are comparable. The subjective survival probability is the only significant variable, having a positive effect on the wealth of the respondents.

For the regressions without probability weighting, the average effect of an increase in the subjective survival probability of 1% is approximately \$4,000. Even though the coefficient of the subjective survival probability in the regressions with probability weighting is slightly lower than the same coefficient in the regression without probability weighting, this does not imply that the effect is smaller. For respondents with an actual subjective survival probability in the 5-85 range the effect is smaller, ranging on average from an increase of approximately \$1,000 to at most \$3,500 for every one percent increase in the actual subjective survival probability. For respondents with subjective survival probabilities at the extremes, the effects of changes in the actual subjective survival probability can be of a much larger magnitude. An increase in the actual subjective survival probability of 1% could on average result in increases in wealth of \$3,500 to as much as \$60,000. The reason for these possibly large effects is that if one's actual subjective survival probability increases from 99% to 100 %, the weighted subjective survival probability increases with 18. This large increase in weight correspondents with the insights from behavioral economics that the 1% difference between 99% and absolute certainty results in very large differences in the perceptions of people. Despite these large differences in the marginal effects of the (weighted) subjective survival probability, the total effect is approximately equal for the regressions with and without probability weighting. The average difference in wealth between respondents with subjective survival probabilities of 0% and 100% is respectively \$400,000 and \$350,000 for the regressions without and with probability weighting. The difference in total effect of \$50,000 is compensated by the difference in constant of approximately equal size.

The difference between the regressions of the unrestricted and restricted sample, and thus the effect of deleting the respondents with the widest intervals, is small. Both the parameter bounds and CIs are slightly more narrow for the restricted sample, but this difference does not alter any conclusions. Furthermore, the parameter bounds and CIs are centered around the same values for both the restricted and unrestricted sample implying that the respondents of these specific response pattern categories do not distort the analysis. The difference between the regressions of the restricted and unrestricted sample will be further examined in section 5.3.3.

The positive effect of the subjective survival probability is in accordance with the life-cycle model as described in the literature review. Individuals with a higher subjective survival probability expect to have a longer lifespan. As the sample contains individuals below age 65, the majority (85.6%) have not yet retired (determined through question 5 of section J of the 2008 HRS). According to the life-cycle model respondents with a longer life expectancy should save more until they reach retirement age. Saving more until retirement age should result in having more wealth until retirement age. This implies that the effect of subjective survival probability on wealth should be positive, which is what is found in all regressions.

5.3.3 Effect of number of repetitions and empirical distribution of estimates

Table 9 reports the effect of increasing the number of repetitions of the bootstrap algorithm for the unrestricted sample without probability weighting. The results for the other regression are comparable and can be obtained from the author on request.

The results from table 9 are very similar to the results of table 7 in terms of convergence. The width of the bounds of the parameter identification regions approximately double as the number of repetitions increases from 100 to 1,000,000 while the width of the CIs increase by approximately 10%. This indicates that using 1,000,000 bootstrap repetitions is sufficient to determine the empirical CIs.

Table 8: Bootstrap estimates of the regression of age and subjective survival probability on wealth

	No Probability Weighting					Probability Weighting				
		Set estimates**					Set estimates**			
	Point estimates*	(B)		(E)		Point estimates*	(B)		(E)	
	MCAR imposed	LB	UB	LB	UB	MCAR imposed	LB	UB	LB	UB
SSP	4.082	3.908	4.353	4.047	4.327	3.405	3.174	4.683	3.644	4.899
	(3.094, 5.069)	(2.939, 5.321)		(3.020, 5.355)		(2.319, 4.492)	(1.956, 5.896)		(2.345, 6.209)	
Age	5.341	4.941	5.574	5.099	5.310	5.771	5.128	5.963	5.279	5.770
	(-0.445, 11.127)	(-0.666, 11.177)		(-0.763, 11.172)		(-0.029, 11.570)	(-0.493, 11.578)		(-0.590, 11.645)	
Const.	-187.138	-212.245	-171.212	-192.093	-173.343	-113.755	-166.311	-103.236	-159.626	-111.035
	(-523.451, 149.175)	(-538.727, 154.586)		(-533.153, 167.699)		(-449.844, 222.333)	(-493.113, 223.552)		(-500.964, 230.376)	
N	5354	5567		5268		5354	5567		5268	

Note: SSP=Subjective Survival Probability, UB=Upper Bound, LB=Lower Bound, *95% CIs based on normal approximation, **95% empirical CIs based on the bootstrap with 1000000 repetitions, (B) unrestricted sample (E) restricted sample.

Table 9: Effect of number of repetitions on bounds for the unrestricted sample without probability weighting

	Number of bootstrap repetitions									
	100		1,000		10,000		100,000		1,000,000	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
SSP	4.015	4.254	3.968	4.311	3.942	4.311	3.928	4.346	3.908	4.353
CI SSP*	(3.047, 5.223)		(3.000, 5.279)		(2.972, 5.279)		(2.959, 5.312)		(2.939, 5.321)	
Age	5.130	5.425	5.073	5.530	5.053	5.530	4.998	5.540	4.941	5.574
CI Age*	(-0.477, 11.030)		(-0.535, 11.133)		(-0.554, 11.133)		(-0.610, 11.144)		(-0.666, 11.177)	
Const.	-203.006	-181.527	-207.968	-179.648	-208.078	-176.332	-209.735	-174.138	-212.245	-171.212
CI Constant*	(-529.252, 144.471)		(-534.342, 146.339)		(-534.706, 149.587)		(-536.195, 151.720)		(-538.727, 154.586)	

Note: SSP=Subjective Survival Probability, UB=Upper Bound, LB = Lower Bound, *95% empirical CIs based on bootstrap

Figure 6 shows the empirical distribution of the parameter estimates and CIs of the subjective survival probability for the unrestricted and restricted sample without probability weighting.

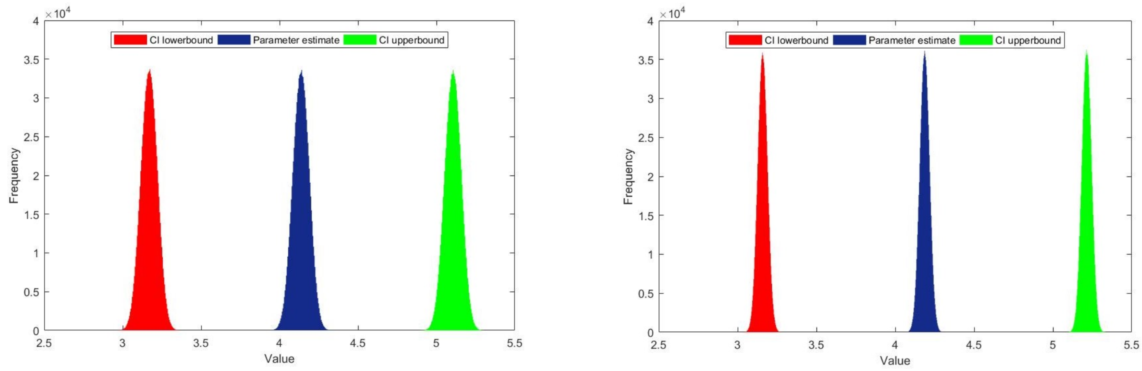


Figure 6: Empirical distribution parameter and distribution CI of subjective survival probability for the unrestricted sample (left) and restricted sample (right) without probability weighting

The general shape of the distributions is similar to those of figure 5. The main difference however is that the distributions are more centered. The effect of omitting the respondents with the widest interval data is also less strong. Given the fact that the variance only decreased slightly and the conclusions drawn, as discussed in the previous section, do not change, omitting the respondents with the widest intervals does not seem to improve the regression results. The results for the other variables are similar with the exception of course that 0 is in between the two empirical distributions of the CIs as the estimates for these variables are insignificant.

5.3.4 Empirical distribution of diagnostics

In order to decide what model is best several diagnostics can be used. The empirical distributions of these diagnostics were obtained through the bootstrap algorithm. The fit of the regressions are compared based on the R-squared statistic. Figure 7 shows the distributions of the R-squared statistics for the various regressions. Given the fact that the distributions of the unrestricted and restricted sample without probability weighting are more to the right than the distributions

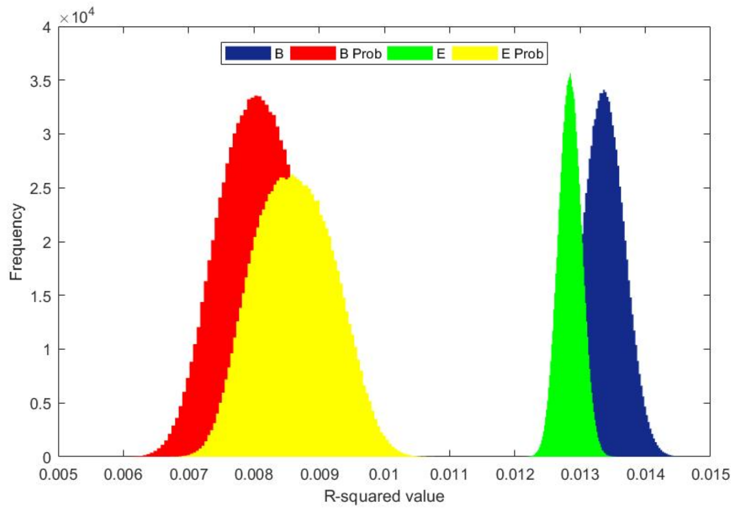


Figure 7: Empirical distribution R-squared

of the regressions in which probability weighting was implemented, it is evident that the fit of the regressions without probability weighting is better. Student's t-tests were used to test for equal means of the various regressions. Using a significance level of 5%, equality of the means was rejected for all different combinations. Omitting the respondents with the widest interval data improves the fit slightly, both for the regression with and without probability weighting.

Choosing a model based on R-squared only could lead to over-fitting. As the aim of many models using the subjective survival probability is to predict inter-temporal choices, diagnostics based on the forecasting ability should be used. One often used statistic for this is the Mean Squared Prediction Error (MSPE). Figure 8 shows the empirical distributions of the MSPE for the various regressions. Clearly, the best model should have the smallest MSPE. As can be seen from the figure the regressions of the unrestricted sample with and without probability weighting outperform the regressions of the restricted sample with and without probability weighting. This implies that those respondents which are omitted from the analysis for the restricted sample have on average lower prediction errors in the regressions of the unrestricted sample than those respondents included both in the regressions of the unrestricted and restricted sample. This implies that the wealth of respondents in the All NR, All 0 or 100 and All 0, 50 or 100 categories can be predicted better than the wealth of the respondents in the remaining categories. This indicates that the answers of the respondents in these categories are more informative than the answers of the respondents in the remaining categories, in contrast to what is found in the literature (Bruine de Bruin & Carman, 2012)

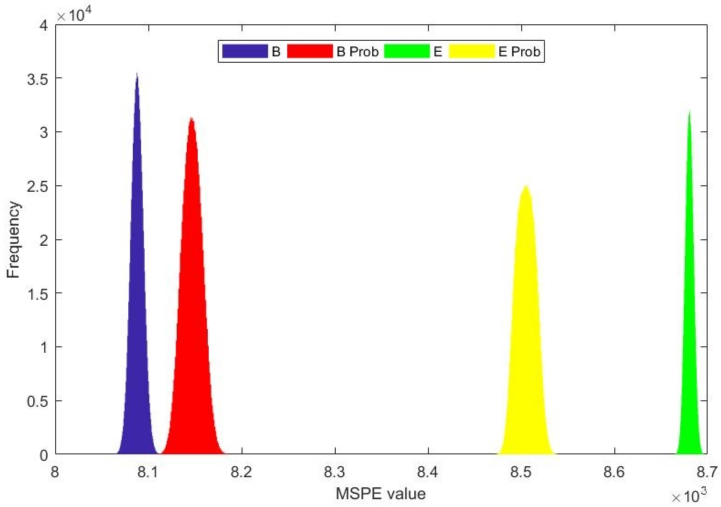


Figure 8: Empirical distribution MSPE

Using probability weighting increases the MSPE for the regression of the unrestricted sample but decreases the MSPE for the regression of the restricted sample.

Besides the fit and predictive strength of a model, several information criteria are often used. The main purpose of these is to select variables of a model. The R-squared cannot be used for this purpose as it always improves as the number of variables increases. The two most frequently used information criteria are the Akaike Information Criterion (AIC) and Bayesian Information Criterion(BIC). Both are based on the likelihood of models, but they include a different penalty term for extra variables. The BIC punishes more for adding an extra variable than the AIC. Selection based on AIC therefore potentially includes more variables than selection based on BIC. As the regressions under consideration in this paper all have the same amount of regressors, both the AIC and BIC lead to identical conclusions. Therefore only the AIC will be reported.

Due to the fact that the width of the empirical distributions is extremely small in comparison to the differences between the various regressions, plotting the distributions does not have any added value. The ranges for the distributions of the regression of the unrestricted sample are centered around 93,000 and those of the restricted sample are centered around 88,000. For the restricted sample the distribution with probability weighting is centered slightly more to the left than the distribution without probability weighting. For the restricted sample the distribution with probability weighting is located to the right of the distribution without probability weighting.

Based on the AIC the effect of probability weighting is ambiguous. For the unrestricted sample probability weighting worsens the distribution of the AIC whereas for the restricted sample probability weighting slightly improves the AIC. It is however clear that based on the AIC the regressions of restricted sample are better, as the AIC values for these regressions are significantly lower than those of the regressions of the unrestricted sample.

The empirical distributions of the diagnostics do not lead to clear conclusions. Probability weighting decreases the model fit for both the unrestricted and restricted sample but the effects on MSPE and AIC depend on the sample. Omitting the respondents of the categories All NR, All 0 or 100 and All 0, 50 or 100 improves the AIC, worsens the MSPE and has an ambiguous effect on the R-squared. Omitting respondents might however cause bias results. Given the ambiguous effects of deleting respondents, and the fact that including these respondents does not alter the conclusions drawn from the analysis, they should not be deleted.

It should be noted that the R-squared distributions indicate that even the model with the best fit only explains at most 1% of the variance of wealth. As the MSPE distribution values are very high compared to the squared level of wealth, the prediction of all models are very poor as well. The ambiguity concerning which model is best might thus be cased by the fact that all models are bad. This indicates that the subjective survival probability is a weak predictor of wealth. Given its weak predictive power and the issues it causes due to rounding, using it to predict wealth is not advisable. Standard OLS models incorporating for example education are much more credible and have higher forecasting ability. The acquired models should thus be compared to standard wealth models.

6 Conclusion

This research has focused on two aspects of inference using probabilistic expectations. First of all, it has shown that the theoretical estimates obtained for regression models with probabilistic expectation interval data are very conservative. Based on the assumption that the true value of the interval data comes from a uniform distribution, a bootstrap algorithm is proposed to estimate empirical parameter bounds. Using these empirical estimates, significant results were found both for interval data as dependent and independent variable. The bootstrap algorithm also enabled the computation of empirical distributions of diagnostics. These distributions can be used to compare various models. The second aspect this research focused on used the diagnostics obtained from the bootstrap algorithm to determine whether the use of probability weighting could improve the predictive power of probabilistic expectations. As illustration, subjective survival probabilities were used to forecast wealth. The analysis showed that in the particular case at hand, probability weighting did not improve the predictive power.

Several problems arose in the analysis which should be investigated. The theoretical estimates for the regressions in which interval data was the independent variable could not be computed. This made it impossible to assess whether and to what degree the empirical estimates acquired through the bootstrap algorithm were stricter than the theoretical estimates.

A second issue concerned the poor performance of all models considered to predict wealth. Even though a significant effect was found of the subjective survival probability on wealth, the diagnostics obtained through the bootstrap algorithm showed poor performance both in- and out-of-sample. Given the poor performance, no strong conclusions could be drawn with respect to the effect of both probability weighting and deleting respondents of certain response categories. To assess whether probability weighting does or does not improve the predictive power of probabilistic expectations, the analysis should be applied to a model in which probabilistic expectations are known to have strong predictive power.

Apart from the poor performance, reverse causality and omitted variables should be considered in further research. This should be done through including more control variables and applying instrumental variable estimation.

Furthermore, the analysis in this paper was based on the assumption that the true value of the probabilistic expectation of a respondent came from a uniform distribution. Determining the distribution is not possible given the current data. Probing questions should be used focusing on what the probability is that the probabilistic expectation is within a certain interval. Additionally, different distributions should be compared. Of special interest are distributions with a lot of weight in the center, such as a truncated normal distribution, and heavy-tailed distributions.

Lastly, probability weighting was implemented in its most basic form. The weighting function was assumed to be identical for all respondents and a fixed c was used. For the analysis in which the probabilistic expectations were one of the regressors, c values were used ranging from 0.5 to 1. This showed that in the specific case at hand, increasing c, resulting in less severe probability weighting, only caused the diagnostics of the regressions with probability weighting converging to the diagnostics without probability weighting. It should be investigated whether this occurs with other models as well, to determine whether probability weighting always reduces the explanatory power or not. To investigate this, one possibility would be to let c be a coefficient in the regressions and estimate the model through Non-linear Least Squares. If c value significantly different from 1 are found, probability weighting does improve the explanatory power. Furthermore, c could be made individual specific as proposed by Gonzalez and Wu (1999). This would however require obtaining individual specific c values through asking additional questions to respondents.

Despite the issues still remaining, the research has shown that by using the bootstrap algorithm, partially identified models can be compared. A wide range of possibilities to improve the explanatory power of probabilistic expectations can therefore be assessed. Probability weighting is only one of the possibilities. One could for example look into the effect of using other algorithms to obtain the intervals, the effects of focal point answers and probing questions.

7 Appendix

A Bootstrap algorithm for confidence intervals of theoretical identification regions

The bootstrap algorithm for computing the CI for partially identified models is based on Algorithm 4.2 of Beresteanu and Molinari (2008).

1. Generate a bootstrap sample of sample size n, $\{(y_{il}^*, y_{iu}^*, x_i^*) : i = 1, ..., n\}$ by drawing with replacement from the vector $\{(y_{il}, y_{iu}, x_i) : i = 1, ..., n\}$
2. Compute $\hat{\beta}_L^*$ and $\hat{\beta}_U^*$ based on the bootstrap sample
3. Repeat step 1 and step 2 b times to obtain the empirical distribution of $\hat{\beta}_L^*$ and $\hat{\beta}_U^*$
4. Construct a $(1-\alpha)\%$ two-sided CI of the true parameter value.
   - Calculate the one sided $(1-\alpha)\%$ lower bound of the CI of the empirical distribution of $\hat{\beta}_L^*$.
   - Calculate the the one sided $(1-\alpha)\%$ upper bound of the CI of the empirical distribution of $\hat{\beta}_U^*$

B Probing Questions Q28

Figure 9 shows the overview of the probing questions asked after question 28. The boxes show either the answers to the questions or the question a respondent is directed to, follow by the number of people who gave that answer or the number of people who were directed to that question. As an example, of all the people who were asked question 28, 227 answered 0. These 227 were all asked question 132. Question 132 had 3 possible answers; Don't know/refused, No chance and Approximate. 71 respondents answered Approximate and all of them were asked question 133.

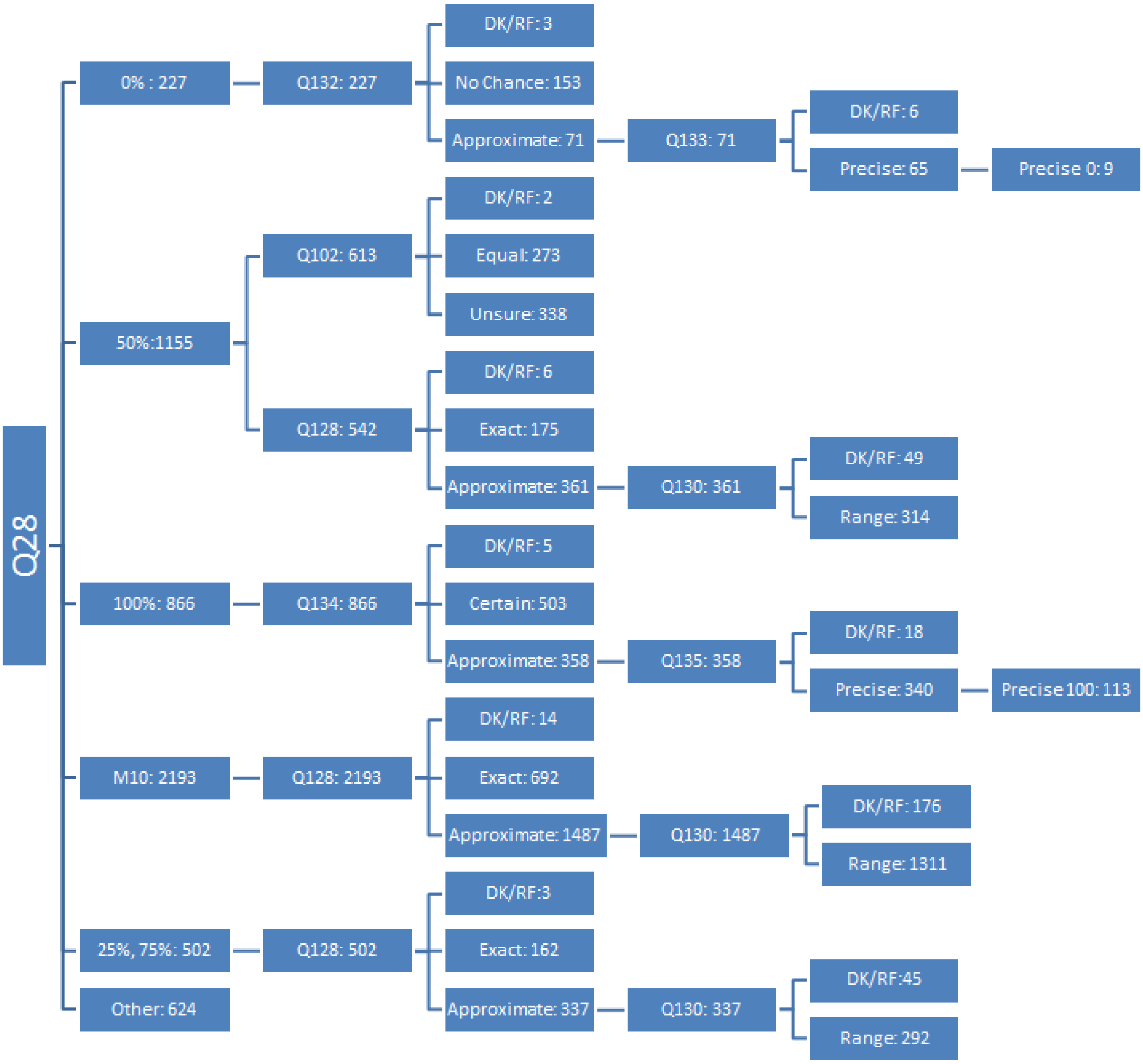


Figure 9: Probing for rounding following question 28 in 2008 HRS

The exact formulation of the probing questions was as follows:

- Question 132:When you say zero percent chance, do you mean that you see no chance at all you will live to 75 or beyond, or do you mean you see a small enough chance that zero is a good approximation?

- Question 133: (If you think there is a small chance that you will live to 75 or beyond,) please give your best estimate of what that chance is.
- Question 102: Do you think that it is about equally likely that you will die before 75 as it is that you will live to 75 or beyond, or are you just unsure about the chances?
- Question 128: When you said PERCENTAGE PER P028 percent just now, did you mean this as an exact number or were you rounding or approximating?
- Question 130: What range of numbers did you have in mind when you said PERCENTAGE PER P028 percent?
- Question 134: When you say 100 percent chance, do you mean that you are certain you will live to 75 or beyond, or do you mean you see a large enough chance that 100 is a good approximation?
- Question 135: (If you think there is a large chance that you live to 75 or beyond,) please give your best estimate of what that chance is.

The two boxes on the far right of the figure, Precise 0 and Precise 100, indicate that the respondent answered 0 and 100 respectively to question 28. Next the respondent stated that this was not an exact answer but an approximation. However, when asked for the exact probability the respondent again answered 0 or 100.

C Proof of the unfeasibility theoretical estimation with interval data as independent variable

The estimation of the theoretical bounds of the regressions with interval data as independent variable is based on eight pairs of restrictions. These restriction are obtained by using the minimums and maximums of age, the lower bound and the upper bound in the following restriction:

$$\beta_1 x_i + \beta_2 v_U + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq \beta_1 x_i + \beta_2 v_L + \beta_3 \tag{11}$$

The range of age is 25-64, and the ranges of both the upper and lower bound are 0-100. This leads to the following eight pairs of inequalities:

1. $25\beta_1 + 0\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 25\beta_1 + 0\beta_2 + \beta_3$
2. $25\beta_1 + 100\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 25\beta_1 + 0\beta_2 + \beta_3$
3. $25\beta_1 + 0\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 25\beta_1 + 100\beta_2 + \beta_3$
4. $64\beta_1 + 0\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 64\beta_1 + 0\beta_2 + \beta_3$
5. $25\beta_1 + 100\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 25\beta_1 + 100\beta_2 + \beta_3$
6. $64\beta_1 + 0\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 64\beta_1 + 100\beta_2 + \beta_3$
7. $64\beta_1 + 100\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 64\beta_1 + 0\beta_2 + \beta_3$
8. $64\beta_1 + 100\beta_2 + \beta_3 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) \leq 64\beta_1 + 100\beta_2 + \beta_3$

Given the identical bounds of the upper and lower bound, restrictions 1, 4 and 5 simplify to $0 \leq \hat{\eta}_N(x_i, v_{iL}, v_{iU}) - (25\beta_1 + \beta_3) \leq 0$ irregardless of the values of the coefficients. As $\hat{\eta}_N(x_i, v_{iL}, v_{iU})$ is never equal to $25\beta_1 + \beta_3$ for all respondents, these restrictions are never satisfied.

References

Allais, M. (1990). Allais paradox. In *Utility and probability* (pp. 3–9). London: Palgrave Macmillan UK. doi: 10.1007/978-1-349-20568-4_2

Arias, E., Heron, M., & Xu, J. (2016). United states life tables 2012. *National Vital Statistics Reports*, *63*(7), 1–63.

Beresteanu, A., & Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, *76*(4), 763–814.

Bloom, D. E., Canning, D., Moore, M., & Song, Y. (2006, November). *The effect of subjective survival probabilities on retirement and wealth in the united states* (Working Paper No. 12688). National Bureau of Economic Research. doi: 10.3386/w12688

Bruine de Bruin, W., & Carman, K. G. (2012). Measuring risk perceptions: what does the excessive use of 50% mean? *Medical Decision Making*, *32*(2), 232–236.

Camerer, C. F., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of risk and uncertainty*, *8*(2), 167–196.

Cerquera, D., Laisney, F., & Ullrich, H. (2014). *A note on regressions with interval data on a regressor* (Discussion Papers of DIW Berlin No. 1419). DIW Berlin, German Institute for Economic Research. Retrieved from `http://EconPapers.repec.org/RePEc:diw:diwwpp:dp1419`

Chernozhukov, V., Hong, H., & Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, *75*(5), 1243–1284.

Dominitz, J., & Manski, C. F. (1996). *Perceptions of economic insecurity: Evidence from the survey of economic expectations* (Working Paper No. 5690). National Bureau of Economic Research. doi: 10.3386/w5690

Dominitz, J., & Manski, C. F. (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association*, *92*(439), 855–867.

Elder, T. E. (2007). *Subjective survival probabilities in the health and retirement study: Systematic biases and predictive validity* (Michigan Retirement Research Center Research Paper No. WP 2007-159). Retrieved from `https://ssrn.com/abstract=1083823`

Fisher, I. (1930). *The theory of interest.* MacMillan, New York.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive psychology*, *38*(1), 129–166.

Gouret, F. (2017). What can we learn from the fifties? *Journal of Forecasting*. doi: 10.1002/for.2468

Horowitz, J. L., & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American statistical Association*, *95*(449), 77–84.

Howden, L., & Meyer, J. (2010). Age and sex composition: 2010. *United States Census Bureau*. Retrieved from `https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf`

Hurd, M. D., & McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human resources*, *30*, S268–S292.

Hurd, M. D., & McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, *112*(482), 966–985.

Hurd, M. D., Smith, J. P., & Zissimopoulos, J. M. (2004). The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, *19*(6), 761–775.

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845–1857.

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association*, *61*(315), 658–696.

Juster, F. T., & Suzman, R. (1995). An overview of the health and retirement study. *Journal of Human Resources*, *30*, S7–S56.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, *47*(2), 263–291.

Kleinjans, K. J., & van Soest, A. (2010). *Nonresponse and focal point answers to subjective probability questions* (IZA Discussion Paper No. 5272). Retrieved from `https://ssrn.com/abstract=1696896`

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

Manski, C. F. (2017). *Survey measurement of probabilistic macroeconomic expectations: Progress and promise* (Working Paper No. 23418). National Bureau of Economic Research. doi: 10.3386/w23418

Manski, C. F., & Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, *28*(2), 219–231.

Manski, C. F., & Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, *70*(2), 519–546.

McFadden, D., et al. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics.* New York: Academix Press.

Ruud, P. A., Schunk, D., & Winter, J. K. (2014). Uncertainty causes rounding: an experimental study. *Experimental Economics*, *17*(3), 391–413.

Salm, M. (2010). Subjective mortality expectations and consumption and saving behaviours among the elderly. *Canadian Journal of Economics/Revue canadienne d'économique*, *43*(3), 1040–1057.

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, *14*(2), 193–201.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, *77*(4), 1299–1315.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management science*, *42*(12), 1676–1690.