Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Jochem de Bresser

Between Goals and Expectations

Essays on Pensions and Retirement

PhD Thesis 2013-023

Between Goals and Expectations – Essays on Pensions and Retirement

Between Goals and Expectations – Essays on Pensions and Retirement

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op

donderdag 19 december 2013 om 14.15 uur

door

JOCHEM RUDOLF DE BRESSER

geboren op 26 februari 1986 te Utrecht.

PROMOTORES:	prof.dr. Arthur van Soest prof.dr. Frederic Vermeulen
OVERIGE COMMISSIELEDEN:	prof.dr. Rob Alessie prof.dr. Marcel Das prof.dr. Pierre-Carl Michaud dr. Martin Salm

Preface

There are many people without whom this thesis would either not exist, or would have been a lot less fun to write. Both my supervisors fall into both categories. Arthur's guiding questions pulled me out of the woods more than once, providing me with a new view on the problem at hand and hinting at possible solutions. His understated sense of humor gave color to our meetings. Frederic is a relentlessly optimistic coach, for whom a setback truly is only a victory in disguise. I continue to be impressed by his insight into the inner workings of the collective model. I really enjoyed talking with other members of the department, be it about work or play. Most importantly, I got to know Martin as a super nice guy who has the ability to fire incisive questions at rates far quicker than my ability to answer them. I would also like to thank Joachim, Liam, Luc, Marike and Thomas for the pleasant collaboration on some of the projects that make up this thesis.

Visiting Montreal was one of the highlights of my PhD period. I would like to thank Pierre-Carl and Raquel for making it possible. Your hospitality made that trip an unforgettable experience.

Though the thesis-related work was mostly enjoyable, the one perk of working at Tilburg University that made my day time and time again was having good friends around to have lunch, coffee and workouts with. First off, I want to thank Tim for being a cool office mate. We immediately found a comfortable balance between working in silence and talking about whatever we saw on- or offline. Aida, Amparo, Arian, Daniel, Gerard, Hettie, Jong-Yook,

Louis, Marc, Mitzi, Nathanael, Niels, Patrick, Peter, Rob, Roxanna, Sybren, Thomas and Tim, our lunches together acted as oases of fun and relaxation. The K-building staircase workouts I did with some of you helped me get in shape to conquer Tough Mudder and cool the mind after a day of pouring over my laptop. You know that the force of awesomeness glows strong in someone if that person is not afraid to look like a fool and sweat like a Greek monkey while bear-crawling up the stairs of the office building where you work. Guys, I am grateful to have been your colleague and I hope we continue to be friends!

Outside of the university, I would like to thank my fiancee, family and friends for being amazing people and supporting me whenever I need it. Ileana, I love you and I cannot wait to marry you next May and August! Maaike, Martien, David, Sanne and Marit, you are my foundation and as a team we stand strong. Whether the task is putting floors in place, painting a new apartment or making sense of life and the decisions that come with it, you are always the first people Ileana and I turn to. Opa and oma, it is great to visit you on lazy Sunday afternoons to discuss world politics, the Dutch education system, or just our most recent vacation or other daily affairs. I am immensely happy to still be good friends with the Gemert/Handel/Boekel-crew and hope to share many more legendary vacations and whiskey-related events with you. To the BitterBallen-Boys I can only admit that I know no better group to drink a round of Duvels and burn our tongues with. And last but not least, living in 's-Hertogenbosch was absolutely amazing thanks to the wonderful friends we have there. Nothing relaxes you after a difficult driving lesson like an evening of playing games with friends, “Shadows over Camelot, anyone?", or accidentally enjoying gay cinema. I believe a wise man once said: “I'll be back!". We intend to put those words into practice.

Finally, I would like to thank Rob Alessie, Marcel Das, Pierre-Carl Michaud and Martin Salm for being on my committee, reading this thesis and providing useful feedback. And of course Hendri for kindly sharing his LateX-template that combined the chapters into a neat booklet.

The research reported in this thesis was financed by The Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO). Data collection was supported financially by Netspar, Network for Studies on Pensions, Aging and Retirement. Furthermore, the thesis

benefitted from the constructive feedback given at many Netspar conferences and workshops where we were given the chance to present our research. The views expressed in the following chapters do not necessarily reflect those of these organizations.

Jochem de Bresser
October 2013

Contents

List of Tables

List of Figures

1 Introduction

This thesis consists of five chapters on various topics related to pensions and aging. The consequences of population aging and the reforms to cope with them, especially to ensure the long-term sustainability of pension systems, are subject to heated public debate. Such controversy is understandable, since pension reforms require individuals to adjust their plans and expectations concerning the use and availability of their income after retirement, an entitlement many believed to be certain. Expectations and uncertainty play important roles in the research described in the following chapters. The first two papers, chapters 2 and 3, describe expectations about pensions and their relationship with well-being and analyze the quality of the data. Chapter 4 focuses on mortality rather than pension expectations and investigates approaches to analyze beliefs held by survey respondents under minimal assumptions. The final two chapters do not concern expectations directly, but do relate to people's subjective ideas about retirement. Chapter 5 analyzes how much individuals want to spend after they retire and the extent to which they can expect to afford those expenditures. Finally, chapter 6 shows that participation in a survey about consumption during retirement led households to save less on average. Each individual essay starts with an introduction, so the remainder of this introductory chapter serves to link the various chapters together and to provide impatient readers with the main message of each paper.

1.1 Retirement expectations and satisfaction with retirement provisions

In chapter 2, we describe employees' expectations regarding their income after retirement relative to their income during working life for a representative sample from the Dutch population. We show how the expected replacement rate of income and the associated uncertainty relate to satisfaction with various aspects of people's pensions. In particular, we look at satisfaction with the age at which individuals expect to retire; with the expected income level; with the knowledge they have of their pensions; with their own pension provisions overall; and with the Dutch system of income provision after retirement. The relationship between expectations and pension satisfaction is interesting for policymakers who are concerned with maintaining support for the pension system in times of reform. The preferences of citizens can have a profound effect on welfare state policies (Brooks and Manza 2007, Cremer and Pestieau 2000). However, the evidence is mixed where it concerns the impact of expectations regarding people's personal pensions on satisfaction with the system, e.g. the extent to which satisfaction with the system is driven by self-interest (Lynch and Myrskylä 2009, O'Donnell and Tinios 2003). Moreover, pension satisfaction is closely related to general job satisfaction (Luchak and Gellatly 2002), which in turn is an important driver of satisfaction with life or happiness (Van Praag et al. 2003).

The expected replacement rate of income at retirement is positively associated with overall satisfaction with personal provisions. This relationship runs through satisfaction with the expected level of pension benefits. We use the fact that we observe the same individuals repeatedly over time, we have panel data on individuals, to show that upward revisions in the expected replacement rate lead to increases in satisfaction with the expected pension income and with personal provisions overall. However, we do not find robust evidence in support of an effect of expectations regarding personal pensions on satisfaction with the system as a whole, suggesting that support for the system is not driven by self-interest. We also find no association between subjective uncertainty and satisfaction. On a methodological level, this paper supports the validity of expectations data on a relatively abstract subject. The finding that expectations vary significantly only with those narrowly defined satisfaction scales for which

we would expect a priori to find an effect shows that the expectations data do not simply reflect varying degrees of general optimism. In that sense, the first chapter paves the way for a further analysis of the quality of expectations data.

1.2 Survey response in probabilistic questions and its impact on inference

Having established the validity of our data on pension expectations in chapter 2, chapter 3 looks more closely at the way respondents answer the questions that elicit those beliefs. When researchers want to know the expectations of survey respondents about a continuous variable, a variable that can take all values in a certain interval, they usually ask a number of questions about the probability that the variable of interest will be below certain thresholds. For the replacement rate of income after retirement, for example, the survey asked individuals about the probability that their replacement rate will be below 100%, 90%, 80%, 70%, 60% and 50% relative to their current real income. Such quantitative questions have the advantages that we can compare answers across respondents and that we can quantify the uncertainty that respondents experience (Dominitz 1998). However, the questions are difficult to answer for many respondents, which affects the quality of the resulting data. For instance, a fifth of the sets of probabilities in our sample violate the logical requirement that probabilities are weakly decreasing for lower thresholds.

In this essay, we formulate a joint model of the process by which respondents answer expectations questions of the type described above and the beliefs that are elicited that way. Our model is based on that presented byKleinjans and Van Soest (2013), but differs in that it analyzes expectations of a continuous rather than binary outcome. In terms of response behavior, we look at item non-response, non-informative focal answers, rounding and recall or reporting error. Item non-response simply means that some respondents do not answer the items on pension expectations, even though they do answer other parts of the same survey. The model incorporates that such non-responding individuals may differ in observed and unobserved ways from respondents who do answer all questions. Non-informative focal answers occur when survey respondents do provide probabilities, but those probabilities do not correspond to their underlying beliefs. We follow Bruine de Bruin et al. (2000) and assume that

such non-informative answers are always equal to 50 percent. Rounding is a type of measurement error that arises when respondents do not report their exact subjective expectations, but rather the nearest multiple of some integer. Rounding limits the informativeness of the data: a probability equal to 15% that is rounded to the closest multiple of five only tells us that the true expectation lies in the interval $[12.5; 17.5]$. Reporting or recall error is the final aspect of answering behavior that we model and it allows us to capture erratic reported probabilities, such as logically inconsistent responses. We assume expectations follow log-normal distributions, the parameters of which we model as a function of socio-economic covariates and unobserved differences between individuals.

We find that all aspects of reporting behavior are persistent. For instance, individuals who round crudely in one survey-wave tend to do so in other waves as well. For expectations, we find that the subjective uncertainty in the replacement rate varies less over time than the expected level. Rounding is common in our data: almost half of the reported probabilities are rounded to a multiple of 10. However, focal answers are rare: the 50/50 answers that we observe express true uncertainty rather than inability to answer the questions. Finally, we compare the estimated associations from the model of answering behavior and expectations with models of expectations that do not take reporting into account. The joint model yields stronger correlations that are more statistically significant, suggesting that it is important to take response behavior into account even if we only care about expectations.

1.3 Eliciting subjective survival curves: lessons from partial identification

This chapter continues the analysis of rounding in subjective probabilities reported in surveys. However, it takes a different perspective than the previous chapter. Instead of building a model that accommodates rounding, it asks what we can learn about expectations under minimal sets of assumptions, some of which allow for rounding. We shift focus from replacement rate expectations to subjective survival expectations that ask respondents about the likelihood that they will survive past age thresholds that range from age 75 to age 90. The starting point of the paper is the observation that mortality expectations are usually modeled using parametric models that characterize

objective survival well, such as the Gompertz and Weibull distributions (see, for example, Perozek 2008). We compare inference based on that approach with less restrictive alternatives. Linear and cubic spline interpolation allow one to pin down preferences exactly, but without the assumption that expectations follow a known parametric distribution (in econometrics parlance: they allow for the non-parametric point identification of expectations). Alternatively, if we allow for rounding and/or acknowledge that we do not know anything about expectations between the thresholds elicited in the survey, we are no longer able to describe exactly how long people expect to live. After all, we know the probability that the respondent attaches to living to age 70 or older and the same probability for age 75, but we do not know his subjective likelihood for surviving at least to age 73. We do know that the latter probability must lie in-between the former two, so that we can construct a region within which the subjective survival function is located. We investigate whether such regions contain useful information.

Our findings show that life expectancies calculated from fitted parametric distributions are similar to those calculated from non-parametric spline functions. Hence, given our relatively rich set of five age thresholds for which expectations are elicited, the choice of parametric form does not affect point identified expectations. Though the data are relatively rich, we cannot learn much about expectations if we are unwilling to make additional assumptions beyond the reported probabilities. The intervals we construct for subjective life expectancy based only on what is given in the data are 11 years wide on average. Models with interval-censored dependent variables indicate that this is too wide for useful inference: none of the associations between life expectancy and covariates that we observe in point identified linear models are confirmed by the bounds. Allowing for rounding only makes the intervals wider. However, if we simultaneously allow for rounding of the reported probabilities and smooth beliefs between the reported points on the subjective survival function, we can narrow down the resulting intervals to an average width of 3 years. Moreover, partially identified linear models show that those intervals-with-interpolation are sufficiently informative to confirm the relationship between self-reported health and life expectancy found in point identified linear models. One caveat is that the type of rounding matters: if we assume that each individual probability is rounded to the maximum extent possible, the intervals for life expectancy

are once again too wide to be informative. Finally, we use our partial identification framework to analyze the stylized fact that individuals, especially women, expect to die younger on average than actuarial life tables suggest (see Perozek 2008 and Kutlu and Kalwij 2012, for confirmations of this pattern using US and Dutch data respectively). The correspondence of expectations to actuarial forecasts is important, since economists commonly use the latter as substitutes for the former for reasons of availability (Peracchi and Perotti 2011). However, if people's expectations differ from life tables on average, the use of life tables leads to misspecified models. For our point estimates of life expectancy, we corroborate the result that women expect to live shorter than cohort life tables predict. However, our bounds show that this gap can be filled completely by allowing for rounding, even if we interpolate expectations between the reported probabilities.

1.4 Can the Dutch meet their own retirement expenditure goals?

Rather than investigating subjective expectations, this chapter looks at the personal goals people set in terms of their expenditures during retirement and their ability to meet those goals. Our starting point is a survey that asks respondents about the minimum level of consumption they would never want to fall below and about what they consider to be an adequate level of expenditures during retirement. Such individual-level consumption floors are a novelty: previous research on savings adequacy has either imposed an uniform consumption floor, usually a poverty line or a certain fraction of current income, or has derived optimal consumption and savings from a lifecycle model. The advantage of eliciting minimal and adequate expenditures directly from survey respondents is that it allows for more variation than do uniform expenditure levels: some people want to sail the world after they retire while others are happy to collect stamps. We link the survey on retirement expenditures to administrative data on the pension entitlements and wealth of the households of our survey respondents. Combining data from those two sources, we investigate at the level of the individual whether or not respondents can reasonably expect to meet their own expenditure goals. In addition, we compare entitlements

with an official poverty line and with a replacement rate of 70% of current income.

We find that needs vary widely in our sample: the average minimal level of expenditures is 1,500 euro per month and the standard deviation is 781 euro. Though this consumption floor is rather high, the poverty line in 2008 was 917 euros per month, most individuals are well prepared: based on pensions alone the median individual can expect to exceed their consumption floor by 25%. If we do take non-housing wealth into account, the gap increases to 37% and if we include all wealth the median individual can even afford 57% higher expenditures relative to their minimum. Almost a fifth of the sample will fall short of their consumption floor, but less than 5% is predicted to miss the poverty line.

Homeowners and highly educated households stand out as relatively rich, both in terms of pensions and (non-)housing wealth. The self-employed, on the other hand, are relatively poor, suggesting that they do not fully make up for their lack of occupational pensions by private savings. Alternatively, our data may miss the assets those households do accumulate in private pension accounts. Education and income are important covariates of minimal and adequate expenditures: highly educated and income-rich individuals report higher minimal and adequate consumption. For men individual and household income matter similarly, while the expenditure needs of women are related mostly to household income.

1.5 Can survey participation alter household financial behavior?

In the final chapter of this thesis, we look again at the questionnaire about retirement expenditures analyzed in chapter 5. However, this time we are not interested in the answers people give. Instead, we analyze the effect of survey participation on subsequent savings. Surveys can affect behavior, because they may remind respondents of certain aspects of the decisions they take that they would otherwise forget (for examples of such “limited attention", see DellaVigna 2009). For instance, Stango and Zinman (2011) show that individuals are less likely to incur the fees that banks charge when their current account balance turns negative after answering non-informative

survey questions about those fees. This happens even after questionnaires that do not mention those particular overdraft fees, but instead address spending controls in general. Similarly, a questionnaire on spending needs during retirement may make respondents more attentive to the need, or lack thereof, to accumulate additional savings to spend down after they stop working. In order to identify the causal impact of the survey on savings, we use the fact that CenterData only distributed the retirement expenditures survey to a randomly selected subsample of the LISSpanel. We exploit that variation in survey participation that lies outside the control of potential respondents to construct a valid comparison between households who did and did not partake in the survey. Moreover, we measure savings using tax records, rather than self-reported assets. Not only are those tax records a cleaner reflection of actual savings, they also rule out the possibility that the survey might affect the way people respond to survey questions rather than actual behavior.

We find that participation in the survey on retirement expenditures reduced savings during the year of the survey by 1,700 euros or 3.5 percent of disposable income on average. Such reduction in average savings is plausible in the particular institutional context of the Netherlands in 2008: universal public pensions and quasi-universal occupational pensions together provided households with extremely generous income replacement at retirement. The average after-tax replacement rate of income at retirement was close to 80% (Bovenberg and Meijdam 2001). Once attuned to such generous and mandatory pension schemes, it is not surprising that individuals feel comfortable to reduce their private savings.

As a falsification check we test whether there are differences in saving in the year before the survey, but we do not find any evidence in that direction. Moreover, the data do not suggest that the survey led to a reallocation of assets towards or away from risky assets, such as stocks or bonds. We do find strong evidence for heterogeneous effects: highly educated and older households reduce their savings a lot after filling out the survey, while young and poorly educated households marginally increase savings. This effect heterogeneity fits with the patterns in pension entitlements documented in chapter 5: highly educated and older households also have the most generous pension rights. Our findings have implications for the design of household panels: if we want to assemble enough data to draw strong statistical conclusions, we might

consider doing so by fielding surveys to a large sample infrequently rather than surveying a smaller panel intensively. Also, our study highlights the power of household panels as laboratories in which we can introduce random variation in information sets.

2 Retirement Expectations and Satisfaction with Retirement Provisions

This chapter is a reproduction of De Bresser and Van Soest (2013a), which is forthcoming in the Review of Income and Wealth.

2.1 Introduction

This paper analyzes the determinants of satisfaction with various dimensions of pension arrangements, emphasizing the role of subjective expectations regarding retirement income. The data come from a longitudinal sample of Dutch wage workers observed during five consecutive years. We consider satisfaction with the age at which workers expect to retire, with the level of the pension benefits they expect to receive, with the knowledge they have of their pension arrangements, with the overall nature of their pension plan, and with the Dutch pension system in general.

Pension satisfaction and its determinants is of substantial importance, since the preferences of citizens can have a profound effect on welfare state policies in many countries (Brooks and Manza 2007, Cremer and Pestieau 2000). Understanding the determinants of such preferences is therefore directly relevant for those who want to maintain support for pension systems in the current times of necessary reforms (O'Donnell and Tinios 2003). Moreover, pension satisfaction is closely related to general job satisfaction (Luchak and Gellatly 2002), which in its turn is an important driver of satisfaction with life or happiness (Van Praag et al. 2003).

In particular, we test whether the expected replacement rate of income at retirement and the associated uncertainty affect pension satisfaction. We expect that higher expected replacement rates lead to higher satisfaction with personal pension provisions, in particular satisfaction with the benefit level. It is less clear, however, if a higher replacement rate also leads to more satisfaction with the system as a whole. This would suggest that satisfaction with the pension system is partly driven by self-interest, and the existing evidence on this seems inconclusive (Lynch and Myrskylä 2009, O'Donnell and Tinios 2003).

Analyzing the predictive power of expectations for satisfaction scales is also of relevance by itself, since it provides insight into the validity of expectations data on a relatively difficult topic. Expectations about retirement are relevant, since they affect the saving behavior of pre-retirees (Bottazzi et al. 2006). Previous research indicates that subjective expectations correlate with background characteristics in sensible ways (Manski 2004), and the validity of expectations data has been established in this way mainly for conceptually straightforward examples such as individual mortality. We contribute to the literature by focusing on replacement rates. Moreover, the combination of panel data and several satisfaction scales allow us to go beyond the correlation of expectations with background characteristics, providing a stricter test for the validity of the expectations data.

We apply two different methods to construct subjective replacement rate distributions from the reported probabilities. The first, proposed in Dominitz and Manski (1997), fits an assumed underlying (log-normal) distribution for each observation by minimizing the squared difference between the probabilities implied by the assumed distribution and those reported in the data. Our second approach, adapted from Bellemare et al. (2012), uses spline interpolation to fit a subjective distribution that passes through the points corresponding to the probabilities reported by the respondents. This procedure is nonparametric, in the sense that it does not assume any parametric form of the underlying distribution.[1] Both methods allow calculating the median and standard deviation of the subjective distribution for each observation, which are then used as explanatory variables in models explaining the satisfaction scales.

[1]The only assumptions imposed by spline interpolation are continuity and smoothness of the distribution function. Hence, the procedure can be used to approximate a large class of subjective distributions.

Our results indicate that the median replacement rate of the respondent's subjective distribution affects satisfaction with various aspects of the pension arrangement significantly and with the expected sign. This finding is robust across parametric and non-parametric specifications of the subjective probability distributions. On a methodological level, the use of Fixed Effects (FE) estimation appears to mitigate the endogeneity of expectations with respect to unobserved heterogeneity. This is evident from Hausman tests comparing the coefficients on the expected replacement rate across Random Effects (RE) and FE models. The expected replacement rate enters almost all satisfaction regressions significantly when we estimate RE models, even those that concern satisfaction with the system as a whole instead of one's personal situation. In the FE models, on the other hand, only those scales related to overall satisfaction with personal provisions and satisfaction with expected pension benefits are affected by the median subjective replacement rate. We interpret this as evidence that there is indeed a part of the error term, say "general optimism," that is correlated with our measures of expectations. Once we remove all unobserved, time-constant, factors from the error term, all correlations but those that we would expect a-priori to be important lose their significance. Time varying optimism, or mood effects, are not a likely explanation of these results, because our satisfaction scales are not elicited in the same survey as the expectations.

On the other hand, our FE models for the complete sample (ages 25 and older) do not provide evidence that pension satisfaction is related to subjective replacement rate uncertainty. The results therefore suggest that the expected benefit level is the more salient concern in our sample, even though the insignificance of uncertainty might reflect attenuation bias stemming from measurement error. These patterns persist if we estimate models on the subsample of respondents that provide logically consistent probabilities and if we limit the sample to middle age respondents. We do find some evidence in RE models that more uncertain individuals tend to be less satisfied with their pension overall, if we control for aspect satisfaction. Hence our results suggest that pension satisfaction of all age groups is affected by the level of the expected pension income, but that uncertainty with respect to pension income is less important.

The structure of the paper is as follows. Section 2.2 provides a short summary of the related literature. Section 2.3 describes the institutional context of the Dutch pension system. Section Section 2.4 provides more details on our data. Section 2.5 introduces the econometric models used to relate satisfaction scales to expectations. Section 2.6 describes the subjective distributions of the replacement rates and Section 2.7 presents the empirical analysis of the effects of replacement rate expectations on pension satisfaction. Section 2.8 concludes.

2.2 Literature

The present paper is primarily concerned with the validity of subjective expectations elicited through probabilistic measures and with the causal impact of expectations on wellbeing. Interest in the direct measurement of expectations has increased considerably since the early 1990s, as expectations are of key interest in intertemporal economic models and measuring expectations helps to avoid making strong assumptions (Manski 2002, 2004).

The measurement of expectations in terms of probabilities has become widespread in economics. As noted by Dominitz (1998), the main advantages of probabilistic questions are ease of interpretation, interpersonal comparability and the ability to characterize uncertainty. Moreover, survey respondents are generally willing and able to think probabilistically and tend to do so using the full expanse of the 0-100 percent chance scale (Dominitz and Manski 1997, Hurd and McGarry 2002, Manski 2004). Dominitz and Manski (2006) measured expected old age social security benefits in the US using subjective probability questions and found large uncertainty and heterogeneity. They emphasized the additional information contained in probability questions compared to traditional questions on point forecasts.

While it is impossible to verify whether reported probabilities reflect the actual beliefs held by respondents, a lot of effort has been exerted to assess the internal consistency and plausibility of responses. On the whole, the evidence suggests that responses have such “face validity” when the questions concern well-defined events that are relevant to respondents’ lives (Manski 2004). For instance, Dominitz (1996) finds that individuals’ income expectations are

stable across successive waves of the HRS. Hurd and McGarry (2002) find that mortality expectations contain an element of expectation that subjective health indicators do not, because the death of a parent affects expectations but not measures of present physical health. Another branch of support for the validity of probabilistic expectations data derives from plausible correlation patterns between expectations and socio-demographic covariates. For instance, earnings expectations are found to be more uncertain among the self-employed than among wage workers (Dominitz 1998). Also, the median expected income one year in the future is lower for those who fear job loss, while reported uncertainty is greater (Dominitz 1998). Such intuitive correlation patterns are also found in data from the Netherlands; see Das and Donkers (1999).

The measurement of subjective wellbeing by means of satisfaction scales is commonplace in the applied literature. The reliability of such data in the context of general life satisfaction has been confirmed through tests of their stability over time (Krueger and Schkade 2008). Several studies have looked into the relationships among general satisfaction and satisfaction with aspects of life, suggesting that the latter is the product of complex interactions of the former (e.g. Van Praag et al. 2003). Similarly, we will analyze overall pension satisfaction in isolation and while controlling for interdependencies between satisfactions with various aspects of pensions. To the best of our knowledge, this paper presents the first effort that combines data on probabilistic expectations with satisfaction scales.

Some related studies look at the opinions and preferences for pension arrangements in different ways than using satisfaction questions. Luchak and Gellatly (2002), analyzing data from a large firm in Ontario, found a negative relation between pension accruals on job satisfaction, implying that those with a large (pension) incentive to stay on the same job are also less satisfied with that job. Van Groezen et al. (2009) analyze preferences for public, occupational, or private pensions using data from the Eurobarometer on 15 different countries and find that current pension provision has a larger explanatory power than personal characteristics. They emphasize the impact of citizens' preferences on welfare state policies. Lynch and Myrskylä (2009), on the other hand, find no relation between public pension levels and support for reform of the pension system among 45+ survey respondents in 11 European countries. Using unique data on satisfaction with various aspects of pension arrangements, we can

test more directly whether people who expect to benefit more are also more satisfied with their pension arrangements.

2.3 Institutional background

In the Netherlands, it is common to think of income during retirement in terms of four categories or pillars. The first pillar consists of public pensions that cover everybody who lived in the Netherlands between the ages of 15 and 65. This public pension (or AOW in Dutch), aims to provide retirees with a subsistence income during retirement. Its level is set in relation to the minimum wage and depends only on the number of years spent abroad during the accumulation period (payments are cut with 2 percent for each year spent abroad between age 15 and 65). The second pillar is that of occupational pensions that cover 90 percent of Dutch workers (Bovenberg and Meijdam 2001). The level of occupational pensions depends on the final or average wages of the individual worker throughout the accumulation phase. Though occupational pensions are mostly defined benefit, the possibility of incomplete adjustment for inflation introduces some uncertainty in payments. Together the first two pillars of the pension system replace on average 70 percent of gross final income (Bovenberg and Meijdam 2001). The third pillar offers saving vehicles aimed specifically at generating additional retirement income, such as life annuities. In contrast to the first two pillars, such third pillar pensions are voluntary and usually of the defined contribution type. The fourth pillar contains all other assets that individuals may decumulate to generate income during retirement, such as savings accounts and housing wealth.

The provision of information about pensions changed in the middle of our sample period. After some pension funds and financial institutions started providing standardized information about pensions to their members in 2007, the release of a yearly Universal Pension Overview (UPO) by pension funds and insurers became mandatory in 2008. UPOs give participants in second and third pillar plans yearly updates on their current entitlements and projected entitlements at age 65 conditional on continuation of the current employment situation.

Data 2.4

The data are taken from the Netspar Pension Monitor (NPM), a survey initiated and funded by Netspar and administered to participants of the CentERpanel, an ongoing online panel survey administrated by CentERdata at Tilburg University.[2] The CentERpanel covers the population in the Netherlands of ages 16 and older and is composed of over 2000 households in which one or more adults are invited to complete questionnaires at home every week over the Internet. Households are randomly selected and those without prior Internet access are given access and the necessary equipment by CentERdata. About 75% of all panel members respond to the questions in a given weekend. Attrition is low, making longitudinal research possible. Rich background information about the panel respondents is available from previous interviews.

The questionnaires of the NPM are distributed to all CentERpanel members of ages 25 and older. We use data from the period 2006 – 2010. The NPM consists of short monthly questionnaires including the questions on satisfaction with pension provisions and the pension system, and a longer annual survey including the questions on expected replacement rate. The monthly questionnaires were distributed to one third of the sample each month, so that every participant in the CentERpanel aged 25 or older got the questions on satisfaction once every three months. Since the annual data on replacement expectations were always collected in June, we used the monthly data on satisfaction obtained in May, June or July.[3] In this way replacement rate expectations and pension satisfaction are measured at approximately the same time. On the other hand, it should be emphasized that annual and monthly surveys were always administered in different weekends, so that satisfaction and replacement rate expectations were never measured in the same weekend. This prevents that mood effects could play a role as confounding factors (see below).

The satisfaction scales measure satisfaction with (aspects of) own pension provisions, as well as with the Dutch system of income provision for the elderly as a whole. Five questions were asked, using the same ten point answering scale from not at all satisfied (1) to completely satisfied (10). See the top panel

[2] See http://www.centerdata.nl/en/centerpanel.

[3] The timing was slightly different at the pilot stage of NPM in 2006.

of Table 2.5 in Appendix 2.A for the exact question wordings. In addition to overall satisfaction with personal pensions, the questions refer to satisfaction with the expected retirement age, the expected post-retirement benefit level, and the knowledge on one's personal pension provisions. The importance of these dimensions of pension arrangements is emphasized by, for example, Hyde et al. (2007). Furthermore, we include satisfaction with the system as a whole, which, in contrast to the other scales, does not refer to the individual's personal situation.

In addition to estimating models explaining each of the reported satisfaction levels, we also estimate a model explaining overall pension satisfaction from satisfaction with the aspects. The latter specification postulates that overall satisfaction is composed of satisfaction with various aspects of the phenomenon under consideration, as is common in the "domains of life"- literature (see Van Praag et al. 2003). It should be noted, however, that the latter regressions may be prone to endogeneity bias due to a mood effect at the time of the survey that affects different satisfaction levels measured during the same survey in the same direction.

The replacement rate questions were only given to respondents who indicated that their "most important activity" is wage labor, so that we cannot analyze the self-employed or those who are temporarily or permanently not working. Furthermore, our focus on wage workers implies that when we refer to pension income, this will include both the universal old age state pension (the first pillar; see Section 2.3) and the occupational pensions (the second pillar), as is emphasized in the questions; additional savings (the third and fourth pillar) are not included. The questions ask for the probability that the respondent's replacement rate will be below a series of thresholds, ranging from 50 to 100 percent. Two sets of such questions were asked, referring to the replacement rate at earliest and at latest possible retirement. Respondents were first asked at what age they expected to first have the possibility to retire, assuming they would stay with their current employer. This expected earliest retirement age was then inserted into the probability questions. Similarly, the probability questions for the replacement rate at latest retirement were preceded by the question whether their employer can dismiss the respondent upon reaching a certain age. If so, that age was elicited and inserted into the second set of probability questions. If not, the latest retirement age is replaced

by the earliest age plus five years. About 55 percent of the sample indicated that their present employer does not enforce a mandatory retirement age, in which case the rather arbitrary point of five years after the earliest retirement age was inserted. To avoid this problem, we only use the replacement rate expectations related to the earliest retirement option. The probability questions were phrased as follows:

> If you would retire at [earliest retirement age], please consider your net total pension income including public pension, relative to your present net wage or salary. What would you think is the probability that your net total pension income in the year after retirement will be worth in terms of purchasing power
>
> a. More than 100% of your present net wage?
>
> b. Less than 100% of your present net wage?
>
> ...
>
> g. Less than 50% of your present net wage?

Instead of the part in brackets, the respondents saw their own answer to the question on their earliest retirement age. Note that the answers to the first and second question should add up to 100% if the respondent's subjective distribution is continuous so that the probability that the replacement rate is exactly 100% equals zero. In the data, the answers to a. and b. add up to less than 100% in 38 percent of all cases. Since our analysis will use continuous distributions, we collapse the two questions into a single probability that the replacement rate is less than 100 percent (taking the average of 100 minus the answer to a. and the answer to b.).[4]

All replacement rate thresholds are presented on a single screen. As a result, respondents might misread the questions and interpret the thresholds as delimiters of bins and indicate, for instance, the subjective probability that their replacement rate will be between 90% and 100% (instead of smaller than 100). The reported probabilities suggest, however, that only very few respondents misinterpret the questions in this way: the fraction of respondents whose answers sum to less than 100 is only 0.022.

[4] We also estimated models using the answer to item b. only; this gave very similar results.

2.4.1 Descriptive statistics

Appendix 2.A presents definitions and descriptive statistics of the satisfaction scales and of all socioeconomic controls included in the regressions. Descriptive statistics are shown in Table 2.6 for the sample reporting wage labor as their most important activity. Relatively many respondents are employed in the industrial (16 percent), financial (16 percent) and healthcare (18 percent) sectors. About half of the respondents indicate that they have the option of gradual retirement. Almost half of the respondents report an expected earliest retirement age of 65 (the eligibility age for the public pension during the survey years). The majority (60 percent) of the sample are males, due to our selection of wage workers only. About 75 percent are living with a partner. By construction the age range is limited to 25 years and older, with an average age of 46. A large fraction (77 percent) own a house and the large majority of respondents are the head of their household. On average respondents have one child. The sample is relatively well educated: 44 percent have finished at least higher vocational training.

Table 2.7 in Appendix 2.A contains descriptive statistics for the satisfaction scales and expectations measures. We find that on average respondents rate their overall satisfaction with personal provisions with a 6 (out of 10). The aspect respondents are least satisfied with is the expected retirement age, with an average rating of 5.5. Both satisfaction with the expected post-retirement income and with insight into own provisions receive an average of 6.0. Compared to the personal provisions, respondents are slightly happier with the system as a whole, which receives an average grade of 6.2. The standard deviations of the satisfaction scales are around 2, so satisfaction varies considerably across the sample. Around 30 percent of the total variation in the scales occurs within individuals.

The reported subjective probabilities are used to estimate the median and the standard deviation of each respondent's subjective replacement rate distribution in each time period using a parametric method assuming log normality following Dominitz and Manski (1997), and using the non-parametric method of Bellemare et al. (2012). See Appendix 2.B for implementation details. The average median replacement rates obtained using the spline and parametric methods are both presented in Table 2.7 in Appendix 2.A. They are close to

each other and indicate that, on average, respondents expect a median replacement rate of 77-79 percent. There is considerable dispersion around this value: the standard deviation is 18 percentage points for both estimation methods. Dropping censored values of the expected replacement rates lowers the sample average only slightly to 75 percent. These averages are quite high but they are in line with generally overly optimistic expectations of the Dutch population, as documented by the Dutch authority that supervises financial markets (AFM 2010). Uncertainty in the sample is widespread as is evident in the average estimated standard deviation of 19-20 percent. Dispersion in uncertainty is almost twice as large for the measure based on log-normal expectations than for the spline estimates, due to the presence of some high uncertainty estimates for the former.

Possible selection issues that arise from either non-response or logically impossible answers to the probability questions are discussed in Appendix 2.E. As reported there, little evidence is found for selectivity on observable covariates or with respect to satisfaction. However, both the average level of the expected replacement rate and the average uncertainty are different in the subsample that reports logically inconsistent probabilities. Therefore we conduct robustness checks in which we limit the estimation sample to logically consistent responses.

Econometric models 2.5

We explain pension satisfaction from the estimated medians and standard deviations of the subjective replacement rate distributions and other factors using several panel data models. We prefer ordered logit models over linear models because they fit better with an ordinal concept of satisfaction. More important is the distinction between Random Effects (RE) and Fixed Effects (FE) models. The literature on life satisfaction emphasizes the importance of controlling for unobserved individual characteristics that have a large impact on various satisfaction measures (Ferrer-i-Carbonell and Frijters 2004). In light of the subjective nature of both expectations and satisfaction scales, FE models therefore seem most suitable. By controlling for any form of time-invariant unobserved heterogeneity, we account for unobserved personality traits, such

as optimism. Hence, to the extent that optimism is time-constant, our analysis is not affected by the potential endogeneity of the expectations with respect to personality types. Furthermore, the mood at the time of the survey does not confound most of our analysis since expectations and satisfaction were elicited in different surveys that fielded in different weeks. This is in line with Podsakoff et al. (2003) who note that separating measurements mitigates the effect of fleeting moods when dealing with subjective data. On the other hand, time-varying optimism may drive correlations between different satisfaction scales measured at the same time, also in FE models. This can affect the results of one of our models – the model explaining overall satisfaction from, among other factors, satisfaction with several aspects of the pension arrangement. We also explored using instrumental variables methods as an alternative identification strategy, exploiting exogenous variation in expectations across sectors of employment and due to the partial introduction of UPOs in 2007. However, we found that the instruments are too weakly correlated with expectations to allow for reliable inference.

We apply two different FE ordered logit estimators, proposed in Das and Van Soest (1999) and Baetschmann et al. (2011). The former divides the ordinal dependent variable into different binary variables that indicate whether or not the scale is above a certain threshold (for our 10-point scale there are 9 such thresholds). Then it estimates a binary FE logit model for each threshold and combines the resulting estimates in an efficient way (see Das and Van Soest 1999, for details).[5] The Blow-Up and Cluster (BUC) estimator proposed by Baetschmann et al. (2011) also estimates conditional logits on all possible dichotomizations of the dependent variable, but does not require two separate steps to obtain estimates. Instead, it estimates all dichotomizations jointly subject to the restriction that the parameters are equal across dichotomizations (see Baetschmann et al., 2011, for more details as well as Stata code). In the next section we only report results for the Das and Van Soest estimator, to save space. BUC estimates are always very similar and available upon request.

[5]We thank Paul Frijters for kindly sharing his GAUSS-code of the Das and Van Soest estimator.

Variation in replacement rate expectations 2.6

We first describe how both the medians and standard deviations of the subjective replacement rate distributions vary across socioeconomic groups, using kernel regressions and linear models. We perform kernel regressions on the full sample of expectations calculated by spline interpolation, since this methodology does not assume a certain form for expectations.

Figure 2.1 presents kernel regressions of the medians (upper panel) and standard deviations (lower panel) of the subjective replacement rate distributions on age and income. The top panel shows that the median declines with age up to the age of 40, after which it stabilizes. An explanation for this pattern may be that replacement rates are relative to current income, while the benefits paid through occupational plans usually depend on the average or the final salary. Younger respondents can expect to earn more in the future, implying that their replacement rates will be higher relative to their current earnings. The expected replacement rate does not change with income up to a net monthly income of 2000 euro, after which it declines from 78 to 73 percent. This decline may be due to the flat-rate public pension which does not depend on previous wages.

The lower panel of Figure 2.1 shows the intuitively plausible negative relationship between uncertainty and age. This can be due to an age effect on objective uncertainty or to an increase of information as the time period separating respondents from retirement gets shorter. Disentangling these pathways requires multivariate analysis; see below. Uncertainty also declines in income, suggesting that low income groups whose pensions depend to a larger extent on the state pensions, have become more uncertain due to the ongoing debate on state pensions during the time period covered by the survey.

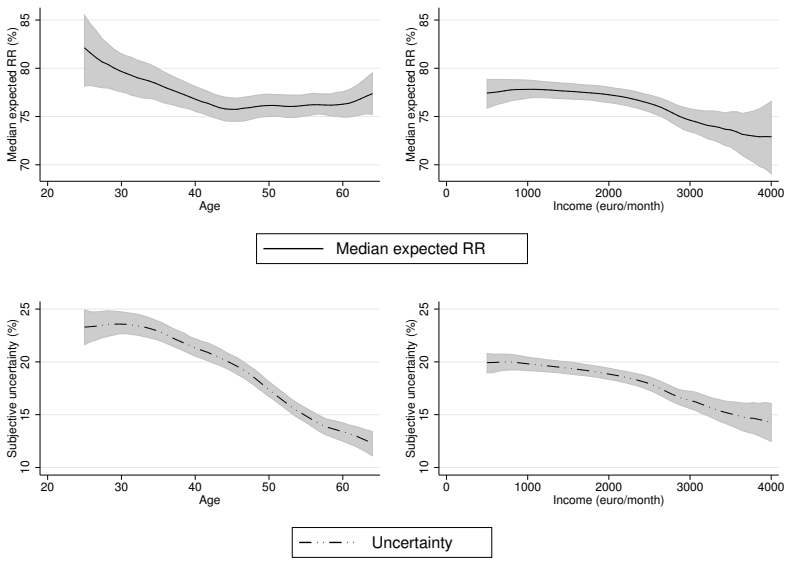


Figure 2.1: Kernel regressions of the median expected replacement rate (upper panel) and subjective uncertainty (lower panel) on age and income

2.6.1 Linear models

To gain more insight in the variation of replacement rate expectations across different socioeconomic groups, Table 2.1 presents estimation results from linear models explaining expectations constructed using the spline approach. We model the median (Table 2.1) and the standard deviation (Table 2.2) separately. We find intuitively plausible relationships between expectations and socioeconomic variables. Respondents who expect their earliest retirement option to occur no later than age 60 expect a replacement rate that is on average 5 percentage points lower than those who do not expect to be able to retire before age 65. On the other hand, an expected earliest retirement age above 65 is not associated with a higher expected replacement rate. As in the kernel regression presented above, log income is negatively associated with the median in the linear models with additional covariates, but the association is insignificant in the RE model. According to the FE estimates, the relation between income and the median expected replacement rate is significant and much larger in magnitude.

The expected replacement rate declines with age up to age 50 and then increases. This pattern is probably due to the definition of the replacement rate relative to current income, which is low relative to final or average earnings for younger workers. This makes it natural that younger respondents expect a higher replacement rate than their older peers. Respondents with high education level expect a 4.5 percentage points lower replacement rate at earliest retirement than respondents with the lowest education level. This may be because those who spent more time in full-time education entered the labor market later, giving them less time to build up a full pension. It may also be due to (relative) optimism of the poorly educated and pessimism of the higher educated. Alternatively, education may pick up some of the income effect since household income is probably measured imprecisely. We find a slightly lower expected replacement rate in the agricultural sector than in the manufacturing (the omitted category).

Table 2.2 presents estimates of linear models explaining the standard deviation of the expected replacement rate distribution. Uncertainty varies little with the expected retirement age: the only significant coefficient indicates that those who expect to retire between 60 and 64 are slightly less uncertain about their replacement rate than those who expect their earliest retirement to be at age 65. Respondents who think they will have access to gradual retirement are less uncertain than those without such an option. The interpretation of this difference is complicated by the fact that we do not observe whether or not respondents actually have access to gradual retirement. Hence gradual retirement may be associated with less uncertainty, whether through causality or self-selection among employees, or respondents without basic knowledge of their pensions may indicate that gradual retirement is not available for them. Age is negatively related to uncertainty, as would be expected since older respondents are closer to retirement. Better educated respondents report less subjective uncertainty, especially when the variation in education appears within respondents over time. The FE estimates show that women who find a partner become less uncertain about their replacement rate, while for men having a partner is not significant.

We find significant variation in subjective uncertainty across sectors: uncertainty at earliest retirement is about 2-3 percentage points lower in the (semi-) public sector compared to the industrial sector. It appears that working for

Table 2.1: Models of medians of subjective RR distributions

	Dependent variable: median RR			
	RE		FE	
Expected ret. age 50-60	-4.828***	(1.308)	-4.648**	(1.850)
Expected ret. age 61-64	-0.873	(0.861)	-1.335	(1.140)
Expected ret. age 66-70	-0.347	(1.365)	-1.382	(1.740)
log(net HH income)	-2.375	(1.678)	-10.81**	(5.363)
Parttime pension	0.408	(0.788)	-0.632	(1.097)
Age	-1.555***	(0.437)		
Age squared/100	1.537***	(0.489)		
Education middle	-1.884	(1.325)	20.16**	(9.733)
Education high	-4.579***	(1.429)	12.43	(10.47)
Male	4.142**	(1.932)		
HH. Head	1.152	(1.524)	-0.306	(3.930)
Number of children	0.143	(0.468)	-0.367	(1.570)
Partner	-0.266	(1.938)	-3.636	(6.753)
Partner*male	-0.0690	(2.373)	3.312	(8.652)
Homeowner	0.0555	(1.152)	-5.655*	(3.255)
Sector: agriculture	-4.960*	(2.938)	-18.50**	(8.852)
Sector: construction	-0.401	(2.502)	7.599	(8.749)
Sector: trade	0.722	(1.834)	-4.118	(7.337)
Sector: transport	-1.569	(2.701)	10.37	(8.510)
Sector: financial services	0.333	(1.609)	-2.289	(5.640)
Sector: education	-0.576	(1.790)	-3.306	(8.102)
Sector: healthcare	-0.0340	(1.708)	4.770	(6.528)
Sector: governance	-1.572	(1.759)	-3.190	(7.841)
Sector: other	2.902	(3.181)	7.171	(18.73)
Wave 2007	-1.225	(0.973)	-0.183	(1.127)
Wave 2008	-0.611	(1.060)	0.858	(1.248)
Wave 2009	1.454	(1.087)	3.487**	(1.353)
Wave 2010	2.615**	(1.169)	4.704***	(1.485)
Constant	132.2***	(14.51)	151.9***	(40.47)
Observations	2,360		2,360	
Number of respondents	1,042		1,042	

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 2.2: Models of standard deviations of subjective RR distributions

	Dependent variable: std. dev. RR			
	RE		FE	
Expected ret. age 50-60	-0.936	(0.778)	-1.965*	(1.045)
Expected ret. age 61-64	-1.179**	(0.508)	-1.203*	(0.644)
Expected ret. age 66-70	0.915	(0.802)	0.662	(0.983)
log(net HH income)	-1.387	(1.055)	-0.700	(3.031)
Parttime pension	-0.799*	(0.468)	-0.422	(0.620)
Age	0.125	(0.275)		
Age squared/100	-0.552*	(0.308)		
Education middle	-1.305	(0.848)	-11.65**	(5.501)
Education high	-2.502***	(0.912)	-10.89*	(5.919)
Male	-1.162	(1.233)		
HH. Head	0.655	(0.952)	-3.176	(2.221)
Number of children	0.0609	(0.296)	-0.223	(0.887)
Partner	-0.546	(1.221)	-6.887*	(3.816)
Partner*male	1.681	(1.502)	8.156*	(4.889)
Homeowner	-0.217	(0.725)	-0.286	(1.840)
Sector: agriculture	-3.069*	(1.855)	-0.0348	(5.002)
Sector: construction	-0.941	(1.588)	12.30**	(4.944)
Sector: trade	-1.236	(1.167)	4.016	(4.147)
Sector: transport	-1.517	(1.694)	8.131*	(4.809)
Sector: financial services	-1.614	(1.019)	2.634	(3.187)
Sector: education	-2.252**	(1.141)	1.347	(4.579)
Sector: healthcare	-3.284***	(1.084)	1.376	(3.689)
Sector: governance	-3.360***	(1.120)	6.022	(4.431)
Sector: other	-2.144	(2.033)	-5.398	(10.58)
Wave 2007	-1.679***	(0.560)	-2.171***	(0.637)
Wave 2008	-1.752***	(0.610)	-2.895***	(0.705)
Wave 2009	-1.306**	(0.628)	-2.917***	(0.765)
Wave 2010	-1.151*	(0.679)	-3.284***	(0.839)
Constant	41.02***	(9.111)	36.94	(22.87)
Observations	2,360		2,360	
Number of respondents	1,042		1,042	

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

public institutions is associated with less subjective uncertainty. This is in line with the relatively secure pension plans and stable careers that traditionally characterized the public sector during the period covered by our data.

2.7 Satisfaction with retirement provisions

We first sketch the bivariate relationship between satisfaction and expectations by means of kernel regressions. We only show the graphs for the nonparametric expectation measures; analogous figures using the parametric method show similar patterns and are available upon request. Figure 2.2 shows the results, with the median expected replacement rate in the left and the standard deviation in the right hand column. Different rows correspond to different satisfaction scales. The general picture is that satisfaction levels are positively associated with the median replacement rate, though most of the associations are not very strong. The exception is satisfaction with expected retirement income, for which the average score is around 5.2 for respondents who expect a replacement rate below 50 percent of their current wage and 6.5 for those who expect this rate to be more than 100 percent. Since income is not controlled for in Figure 2.2, this implies that satisfaction is related to the relative level of post-retirement income even though a high relative income may still be low in absolute terms. This pattern may reflect that, perhaps due to the affluence of most respondents in our sample, relative income matters considerably and current income forms the baseline against which post-retirement income is evaluated. The relationship between satisfaction and the expected replacement rate is slightly hump-shaped with a maximum around 80-90 percent, which is why we will also consider quadratic terms in the regression models.

The right hand panels of Figure 2.2 shows that there is a negative bivariate relationship between satisfaction and uncertainty, which is not very strong either. The strongest negative association with uncertainty is found for pension knowledge satisfaction, which seems intuitively plausible. For satisfaction with the age at which one can retire, only a very weak (and even non-monotonic) association is found.

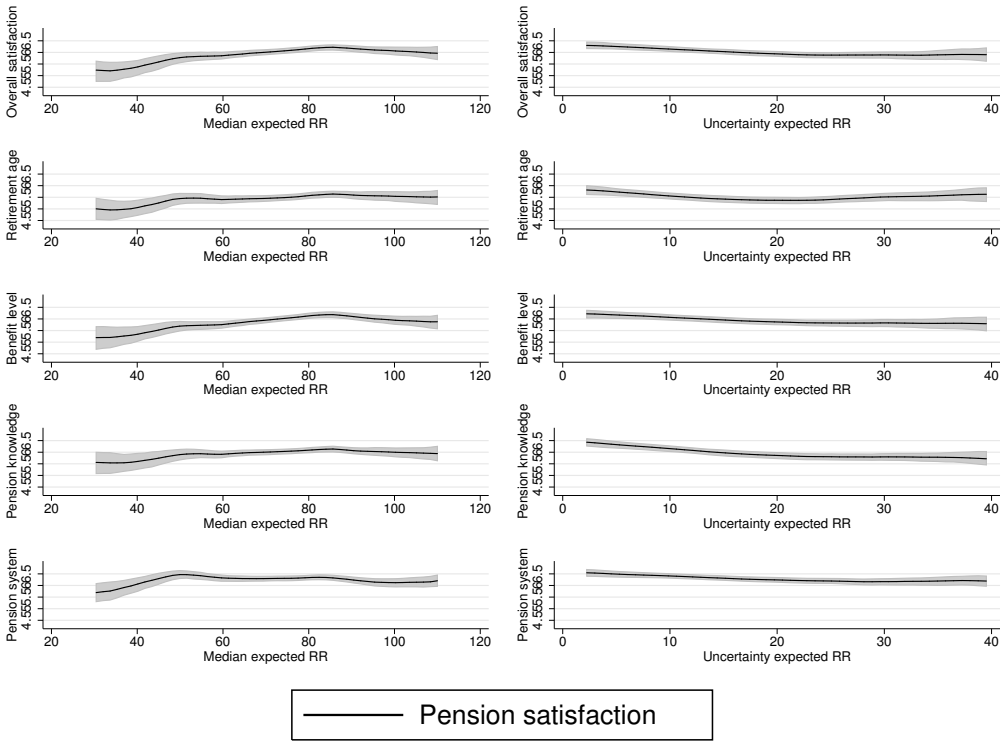


Figure 2.2: Kernel regressions of pension satisfaction on expectations: median expected replacement rate (left column) and uncertainty (right column)

2.7.1 Ordered logit models

Random effects (RE) and fixed effects (FE) ordered logit estimates are presented in Tables 2.3 and 2.4 respectively. We use the nonparametric estimates of the medians and standard deviations of the subjective replacement rate distributions; using the parametric estimates gives similar results.[6] In the RE models, we control for all covariates listed in Table 2.5 of Appendix 2.A. In the FE models, estimated using the Das and Van Soest estimator, we could not include so many controls, since this would limit the sample size severely (because each coefficient must be identified for each cutoff that is included in the estimation). Hence in the FE models, we only control for replacement rate expectations, income, expected retirement age, owning a house, and time effects. Importantly, all models control for (the log of) net monthly personal income. Keeping income constant, a higher replacement rate corresponds to a higher pension income. Hence, keeping income constant, we would expect a positive association between the replacement rate and pension satisfaction.

The RE models can be formally tested against FE through Hausman tests comparing the two sets of estimates. Considering the coefficient on the median subjective replacement rate, the RE null hypothesis is rejected in the models for satisfaction with pension benefits and for satisfaction with pension knowledge at a significance level of 1%. Though this implies that we need FE models for causal interpretation, we also present some results from RE ordered logit models for the sake of comparison and to see how heterogeneity in pension satisfaction is associated with time persistent characteristics. We prefer to estimate these correlations with random effects models rather than OLS, because of the panel structure of the data. Expectations are correlated across repeated observations of the same respondent: In the RE models, individual effects make up around 60% of the total unsystematic variance.

The first two models in Table 2.3 both explain overall satisfaction with personal pension provisions, but the second model also controls for aspect-satisfaction. These aspect satisfactions are all positive and significant. The RE estimates suggest that satisfaction with the benefit level is more important than the other two aspect satisfactions, but this is reversed in the FE estimates. The

[6] Parametric estimates are available upon request. All fixed effects results reported in this section were confirmed using the BUC estimator and for linear models (results available upon request).

difference suggests that unobserved individual effects driving satisfaction with benefits and overall satisfaction are particularly strongly correlated.

The median replacement rates at earliest retirement are strongly significant in four of the six RE models and always have the expected positive sign.[7] Higher expected replacement rates are associated with greater satisfaction with one's own pension provisions overall, but the significant coefficient in the second column suggests that this association is only partly captured by the significantly positive relations of the median replacement rate on satisfaction with the benefit level and the age at which one can retire. Although the final satisfaction scale refers specifically to satisfaction with the Dutch pension system, not taking into account one's personal situation, this measure also appears to be significantly positively related to the respondents' own median replacement rate. In this RE model, this positive association might reflect that respondents giving positive evaluations also tend to be optimistic. An alternative explanation could be that individuals' evaluations of the system as a whole are driven by self-interest. This is in line with the interpretation of O'Donnell and Tinios (2003), who find that the Greek pension system is evaluated better by those who benefit more. In the FE models we will be able to disentangle the various explanations.

As in the RE model, the median subjective replacement rate positively affects overall satisfaction with the personal provisions, mainly through satisfaction with the expected pension income - which is in this case the only aspect scale that is significantly affected by the median replacement rate. In contrast to the RE model, however, the FE estimates only provide limited support that satisfaction with the pension system as a whole is related to personal expectations. This result suggests that the RE result was due to correlation between optimism about replacement rates and a tendency to be positive about the pension system and does not reflect a causal (self-interest) effect.

In the RE models, the measure of uncertainty in the expected replacement rate is significant in only one case: more uncertainty is negatively associated

[7]Based on the kernel regressions in Figure 2.2, which reveal that the bivariate relationship between the expected replacement rate and satisfaction is slightly hump-shaped, we tested for quadratic relationships for both the median expected replacement rate and uncertainty. Since the quadratic terms were insignificant in all models, we dropped them from the final specifications.

Table 2.3: RE ordered logit models of pension satisfaction (expectations modeled using splines)

	Dependent variable: satisfaction with					
	Personal provisions					
	Overall	Overall	Ret. age	Benefits	Knowledge	The system
Satisfaction with ret. age		0.858***				
		(0.0447)				
Satisfaction with benefits		1.176***				
		(0.0639)				
Satisfaction with knowledge		0.660***				
		(0.0545)				
Median/10	0.148***	0.0668**	0.0656*	0.157***	0.0399	0.0680**
	(0.0372)	(0.0306)	(0.0358)	(0.0372)	(0.0348)	(0.0332)
Std. dev./10	-0.00313	-0.103**	0.0492	0.0652	0.0129	-0.0588
	(0.0686)	(0.0503)	(0.0598)	(0.0647)	(0.0588)	(0.0568)
Expected ret. age 50-60	0.612**	0.239	0.800***	0.660***	0.366	-0.168
	(0.246)	(0.189)	(0.228)	(0.242)	(0.227)	(0.223)
Expected ret. age 61-64	0.425***	0.0723	0.549***	0.355**	0.353**	0.0792
	(0.148)	(0.123)	(0.139)	(0.149)	(0.140)	(0.138)
Expected ret. age 66-70	0.0362	0.0781	0.00385	-0.298	0.0978	-0.130
	(0.235)	(0.207)	(0.221)	(0.242)	(0.218)	(0.213)
log(net HH income)	2.297***	0.609***	1.053***	2.356***	2.200***	1.064***
	(0.402)	(0.222)	(0.342)	(0.380)	(0.337)	(0.312)
Parttime pension	0.0486	-0.304***	0.307**	0.139	0.215	0.0415
	(0.143)	(0.111)	(0.137)	(0.141)	(0.131)	(0.129)
Age	-0.0757	0.0953*	-0.363***	0.00412	-0.0804	0.0312
	(0.106)	(0.0568)	(0.0956)	(0.104)	(0.0818)	(0.0795)
Age squared/100	0.139	-0.0992	0.454***	0.0288	0.138	0.00605
	(0.119)	(0.0633)	(0.106)	(0.115)	(0.0898)	(0.0876)
Education middle	0.0952	-0.00444	0.426	0.116	-0.176	0.581**
	(0.397)	(0.164)	(0.281)	(0.324)	(0.282)	(0.255)
Education high	0.289	-0.0983	0.619**	0.639*	-0.134	0.914***
	(0.372)	(0.178)	(0.267)	(0.337)	(0.292)	(0.275)
Male	0.0469	-0.0391	0.232	-0.141	0.133	0.111
	(0.527)	(0.237)	(0.387)	(0.376)	(0.343)	(0.335)
HH. Head	-0.567*	-0.293	-0.283	-0.559*	-0.426	-0.292
	(0.337)	(0.203)	(0.293)	(0.331)	(0.282)	(0.280)
Number of children	-0.0116	-0.0347	-0.0562	0.0228	0.121	-0.139
	(0.126)	(0.0581)	(0.0992)	(0.109)	(0.0991)	(0.0850)
Partner	0.00906	0.00221	0.132	-0.306	0.211	-0.109
	(0.495)	(0.249)	(0.398)	(0.401)	(0.367)	(0.356)
Partner*male	-0.0381	0.129	0.374	0.317	-0.291	0.0878
	(0.643)	(0.300)	(0.504)	(0.504)	(0.436)	(0.425)
Homeowner	0.460	0.243*	-0.0926	0.316	0.377*	0.192
	(0.292)	(0.147)	(0.263)	(0.257)	(0.210)	(0.200)
Observations	1,786	1,680	1,778	1,716	1,783	1,796
Number of respondents	835	796	842	808	833	842

***significant at 1%; **significant at 5%; *significant at 10%

Table 2.3: RE ordered logit models of pension satisfaction (expectations modeled using splines, continued)

	Dependent variable: satisfaction with					
	Personal provisions					
	Overall	Overall	Ret. age	Benefits	Knowledge	The system
Sector: agriculture	0.653	-0.493	0.999**	0.835*	1.339***	0.166
	(1.043)	(0.359)	(0.507)	(0.492)	(0.517)	(0.519)
Sector: construction	-0.126	0.319	0.310	-0.911*	-1.203**	-0.550
	(0.615)	(0.302)	(0.466)	(0.484)	(0.544)	(0.494)
Sector: trade	-0.379	-0.293	0.113	-0.396	-0.0308	0.0289
	(0.498)	(0.226)	(0.426)	(0.408)	(0.387)	(0.331)
Sector: transport	-0.172	1.020***	-0.526	-0.0761	0.518	-0.294
	(0.555)	(0.342)	(0.441)	(0.564)	(0.526)	(0.472)
Sector: financial services	0.258	0.00714	0.316	0.319	0.446	0.581**
	(0.390)	(0.200)	(0.305)	(0.384)	(0.328)	(0.284)
Sector: education	-0.0591	0.0547	0.0523	-0.00169	0.316	0.648**
	(0.460)	(0.218)	(0.366)	(0.402)	(0.378)	(0.329)
Sector: healthcare	0.482	0.148	0.00743	0.488	0.955**	0.458
	(0.505)	(0.213)	(0.364)	(0.379)	(0.372)	(0.308)
Sector: governance	0.980**	0.468**	0.503	0.280	0.713**	0.846***
	(0.414)	(0.217)	(0.309)	(0.394)	(0.335)	(0.310)
Sector: other	0.998	0.295	1.345*	0.449	0.331	0.781
	(0.920)	(0.461)	(0.702)	(0.863)	(0.610)	(0.695)
Wave 2007	-0.0844	-0.229	0.109	0.193	0.128	-0.117
	(0.154)	(0.146)	(0.147)	(0.154)	(0.148)	(0.146)
Wave 2008	-0.0925	-0.156	0.133	-0.0462	0.0231	-0.620***
	(0.173)	(0.161)	(0.163)	(0.171)	(0.160)	(0.163)
Wave 2009	-0.278	-0.154	0.0994	-0.204	-0.0782	-0.694***
	(0.182)	(0.159)	(0.166)	(0.173)	(0.163)	(0.164)
Wave 2010	-0.371*	-0.326*	0.0339	0.0817	-0.0399	-0.648***
	(0.201)	(0.172)	(0.182)	(0.193)	(0.181)	(0.180)
Fraction var. ind. Effects	0.672***	0.117***	0.645***	0.656***	0.616***	0.592***
	(0.0208)	(0.0401)	(0.0221)	(0.0225)	(0.0238)	(0.0263)
Observations	1,786	1,680	1,778	1,716	1,783	1,796
Number of respondents	835	796	842	808	833	842

***significant at 1%; **significant at 5%; *significant at 10%

Table 2.4: FE ordered logit models of pension satisfaction - expectations modeled using splines

	Dependent variable: satisfaction with					
	Personal provisions					
	Overall	Overall	Ret. age	Benefits	Knowledge	The system
Satisfaction with ret. age		0.788***				
		(0.102)				
Satisfaction with benefits		0.593***				
		(0.133)				
Satisfaction with knowledge		0.816***				
		(0.140)				
Median/10	0.118***	0.0279	-0.00207	0.104***	-0.00190	0.0705*
	(0.0356)	(0.0748)	(0.0332)	(0.0382)	(0.0383)	(0.0414)
Std. dev./10	0.0339	-0.172	0.107*	0.105	0.0399	0.0112
	(0.0660)	(0.149)	(0.0608)	(0.0726)	(0.0657)	(0.0678)
log(net HH income)	0.672	-1.840	0.632	0.770	0.994	-0.552
	(0.725)	(1.359)	(0.714)	(0.678)	(0.649)	(0.675)
Expected earliest age 50-60	0.480	0.190	0.327	0.163	-0.457*	-0.652**
	(0.257)	(0.418)	(0.227)	(0.286)	(0.269)	(0.263)
Expected earliest age 61-64	0.0713	-0.193	0.366**	0.0570	0.0967	0.0125
	(0.161)	(0.300)	(0.145)	(0.174)	(0.157)	(0.157)
Expected earliest age 66-70	0.080	-0.656	0.103	0.0946	0.208	0.185
	(0.244)	(0.411)	(0.249)	(0.237)	(0.283)	(0.221)
Homeowner	-0.302	1.009	-0.720**	-0.937***	-1.103***	-0.491
	(0.363)	(3.045)	(0.316)	(0.361)	(0.396)	(0.412)
Wave 2007	0.184	-0.155	0.230*	0.427***	0.436***	-0.0506
	(0.143)	(0.276)	(0.140)	(0.145)	(0.139)	(0.138)
Wave 2008	0.167	-0.0638	0.395**	0.235	0.404**	-0.516***
	(0.168)	(0.286)	(0.155)	(0.168)	(0.168)	(0.166)
Wave 2009	-0.044	-0.0453	0.605***	0.0382	0.195	-0.594***
	(0.184)	(0.354)	(0.168)	(0.181)	(0.172)	(0.172)
Wave 2010	-0.0226	-0.0393	0.453**	0.396*	0.390**	-0.428**
	(0.193)	(0.387)	(0.188)	(0.202)	(0.199)	(0.193)
Informative observations	1,016	941	1,001	990	1,003	1,042
Informative respondents	321	299	321	314	317	333
Observations	1,786	1,680	1,778	1,716	1,783	1,796
Number of respondents	835	796	842	808	833	842

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

with overall satisfaction with the personal provisions in the specification with controls for the aspect satisfaction levels.

The FE models, however, indicate that subjective uncertainty does not affect the satisfaction scales significantly, not even the scale measuring satisfaction with knowledge of one's pension rights. One interpretation of this result is that respondents may truly be indifferent to the uncertainty expressed through the subjective distributions, but another, perhaps more realistic, explanation which we cannot rule out is that subjective uncertainty is measured with considerable error so that the estimate suffers from attenuation bias. This consideration suggests using robustness checks in which we allow the parameters on the mean and standard deviation to differ across subsamples that are plausibly affected by measurement error to different degrees (see below).[8]

Household income is significantly positive in all RE models but not in any of the FE specifications. This suggests that, keeping replacement rate expectations and other factors constant, higher income groups are more satisfied with their pensions and the pension system, but these are not causal effects - a change in household income does not lead to more pension satisfaction in the same time period.

The RE models also show that a lower expected minimum age at which respondents can retire (earliest retirement age less than 65) is associated with higher satisfaction overall, with the retirement age, and with benefits. The effects largely disappear, however, in the FE estimates. Perhaps there is not enough genuine variation in the expected retirement age (other than reporting errors) to get reliable estimates of the causal effect.

According to the FE estimates, home ownership has a significantly negative effect on satisfaction with the retirement age, benefits, and pension knowledge. This could be related to the fact that mortgage interest payments generally increase financial needs and respondents anticipate that this will be the same after they have retired. Many mortgages in the Netherlands are interest-only. Interest payments are tax deductible for 30 years so that the after tax burden may increase after retirement. The FE estimates of the time effects suggest that satisfaction with the pension system has fallen over time, perhaps because of

[8]Accounting for fixed effects often increases attenuation bias but may also help to reduce the influence of measurement error: If a lack of understanding of the questions leads to uninformative answers that follow the same pattern in different waves, FE estimates will not be affected by the such time persistent measurement errors.

the negative publicity about the financial sector in general and pension funds in particular in the time period considered. On the other hand, satisfaction with the three aspects of personal pension provisions seems to have increased after 2006, although the time trends are not very clear.

Finally, the sector dummies in the RE models indicate that civil servants ("governance") tend to be more satisfied with their personal provisions as well as the system as a whole than most others (keeping expected replacement rates and other variables constant). Construction sector workers are particularly unhappy with their benefit levels and the knowledge about their pensions.

2.7.2 Robustness checks

In order to evaluate the influence of measurement error on our results, Table 2.8 of Appendix 2.C reports estimates of FE ordered logit models based on the subsample of respondents that report monotonic probability sequences. As mentioned above, the intuition is that violation of the monotonicity requirement of the cumulative distribution function signals poor understanding of probability and hence noisier reported probabilities. Limiting the sample to monotonic responses does not affect our conclusions. We still find a positive and significant effect of the expected replacement rate at earliest retirement on overall pension satisfaction, which runs through satisfaction with expected post-retirement income. Moreover, for the monotonic subsample we again find no significant effect of subjective uncertainty on satisfaction.

Another potential concern is that retirement might not be perceived as an urgent topic to think about by younger respondents. In light of the findings in Manski (2004), there is reason to doubt the value of retirement expectations data for the younger cohorts if they do not see their retirement income as relevant to their lives. Moreover, the sources of uncertainty about retirement income may be different for younger and older cohorts, since future wages are less certain at younger ages. Both arguments underline the importance of analyzing cohort-based subsamples. Appendix 2.D, Table 2.9, displays estimates of the effect of the expected replacement rate and uncertainty on the various satisfaction scales for the subsamples that are over 40 and 50 years of age. Moreover, we also include different panels for subsamples according to logical consistency of the responses. The findings described above are corroborated

for both age-based subsamples, regardless of further restrictions based on consistent answers. The expected replacement rate affects overall satisfaction with one's pension provisions positively, and this effect runs (partly) through satisfaction with the expected retirement income. As explained above, these subsamples of older respondents probably provide a cleaner test for the effect of replacement rate uncertainty on welfare. However, we do not find robust evidence for an effect of subjective uncertainty on satisfaction even for older respondents. Hence, the effect of pension uncertainty on welfare appears to be less pronounced than that of expecting a low replacement rate.

As a final test of the robustness of our results we redid the analysis using a different range for the replacement rate. Note that the survey questions from which we estimate expectations do not ask for bounds on the replacement rates (see Section 2.4). Hence we need to impose a maximum and minimum replacement rate in order to carry out spline interpolation. All results described above are based on the relatively wide bounds of 20 and 170 percent. However, sensitivity analysis imposing bounds equal to 30 and 120 shows that none of our results is sensitive to changing the bounds. Moreover, all findings are robust to using the interquartile range as an alternative measure of uncertainty. Finally, we investigated the role of skewness in expectations by including the 10th and 90th percentiles of the replacement rate distributions as additional regressors in the fixed effects models. However, we found no evidence of a relationship between these higher order moments and satisfaction. Detailed results are available upon request from the authors.

Conclusion 2.8

Public attitudes towards pension provisions and a country's pension system play an important role in the political debate on pensions and retirement and the willingness to accept the necessary reforms. In this paper we have analyzed the determinants of satisfaction of Dutch employees with own pension provisions overall, with three aspects of own pension provisions (retirement age, retirement income, and insight into own entitlements), and with the Dutch pension system, emphasizing the role of the employees' expectations of the retirement income replacement rate. To this end we have constructed indexes of the level

(represented by the median of the subjective distribution) and the uncertainty (represented by the standard deviation) of future retirement replacement rates from survey data on subjective probabilities that these replacement rates will be below certain thresholds. We used these indexes, together with other factors, to explain satisfaction scores. We used a longitudinal data set comprised of Dutch wage workers spanning the five year period 2006-2010. To account for the possibility that personality traits such as optimism may drive the subjective variables on both sides of the regression equation, we focused on Fixed Effects (FE) panel data models. Time-varying optimism is unlikely to drive our results, since expectations and satisfaction were elicited in different questionnaires.

Our results indicate that the level of the expected replacement rate has a substantial positive effect on overall satisfaction with personal provisions. This plausible relationship is robust to non-parametric estimation of the probability distributions characterizing expectations. Moreover, we find that this effect on overall satisfaction runs through satisfaction with expected post-retirement income. In RE models we also find a positive association between the level of the expected replacement rate and satisfaction with other aspects of pension provisions and with the Dutch pension system as a whole, but these become insignificant in FE models and therefore probably do not reflect causal effects. Similarly, the RE models reveal that higher income groups and workers who expect to be able to retire before age 65 tend to be more satisfied with their pensions, but the FE results suggest that income changes or changes in the expected earliest age at which retirement is possible do not have a causal effect on pension satisfaction.

We do not find much evidence that subjective risk is related to pension satisfaction. Only if we restrict the sample to individuals over 40 years of age, a marginally significant negative effect is found in the FE model, suggesting that uncertainty matters more for individuals who are closer to retirement (despite the fact that subjective risk declines strongly with age). An alternative explanation would be that the measurement of expectations is noisier for younger respondents, since they might perceive retirement as far away and not particularly relevant to their current situation. However, the finding that lower expected replacement rate levels reduce satisfaction over the entire age range 25-65 suggests that also for younger respondents, the data contain information on actual expectations.

Despite the fact that subjective expectations play an important role in intertemporal economic models, elicitation of expectations by means of probabilistic measures only really took off at the end of the 1990s. Since then, the face validity of expectations data has been established for many conceptually simple topics, such as individual mortality and next year's wages. A second contribution of our study is that it specifically assesses the validity of measures of expectations regarding retirement.

Acknowledgements 2.9

This research was funded by NWO and Netspar. The authors are grateful to Viola Angelini, two anonymous referees and participants at the 2011 Netspar International Pension Workshop in Turin for useful comments.

2.A Definitions of variables and descriptive statistics

Table 2.5: Variable definitions.

Variable	Description
Overall satisfaction	"All in all, how satisfied are you with you pension?" - 10 point scale
Age satisfaction	"How satisfied are you with the age at which you expect to retire/have retired?" - 10 point scale
Amount satisfaction	"How satisfied are you with the height of your (expected) income after retirement?" - 10 point scale
Knowledge satisfaction	"How satisfied are you with the knowledge you currently have about your pension/had about your pension before retirement?" - 10 point scale
System satisfaction	"All in all, how satisfied are you with the current system of pensions and welfare in the Netherlands? This concerns the Dutch system, not your own situation." - 10 point scale
Ln(net income)	Log of respondent's net monthly income
Agriculture, water, energy	Respondent employed in agriculture, water purification or energy sector
Construction	Respondent employed in construction sector
Trade	Respondent employed in trade sector
Transport	Respondent employed in transport sector
Financial and business services	Respondent employed in financial and business services sector
Government: education	Respondent employed in educational sector
Government: healthcare	Respondent employed in healthcare sector
Government: governance	Respondent employed in governance sector
Other: food and interests	Respondent employed in food and interests sector
Part-time pension	Respondent has the option of part-time retirement
Expected retirement age 50-60	Expected earliest retirement before age 60
Expected retirement age 61-64	Expected earliest retirement age 61-64
Expected retirement age 66-70	Expected earliest retirement age older than 65
Male	Respondent is male
Partner	Respondent lives with partner
Male*partner	Interaction term of being male and living with partner
Age	Age of the respondent
Number of children	Number of children in the respondent's household
Education: middle	Finished secondary school/lower vocational training
Education: high	Higher vocational training/university
Household head	Respondent is self-reported head of the household
Homeowner	Respondent is (co-)owner of his house

Table 2.6: Descriptive statistics

	N (obs)	n (res)	Mean	Std. Dev. Overall	Std. Dev. Within
SE-variables: income					
Ln(net income)	3,048	1,323	7.42	0.39	0.07
SE-variables: sector					
Industry[a]	3,210	1,415	0.16	0.36	0.06
Agriculture, water, energy	3,210	1,415	0.03	0.16	0.04
Construction	3,210	1,415	0.04	0.21	0.04
Trade	3,210	1,415	0.10	0.30	0.06
Transport	3,210	1,415	0.04	0.19	0.05
Financial and business services	3,210	1,415	0.16	0.37	0.07
Government: education	3,210	1,415	0.14	0.35	0.05
Government: healthcare	3,210	1,415	0.18	0.39	0.07
Government: governance	3,210	1,415	0.12	0.33	0.05
Other: food & interest	3,210	1,415	0.03	0.16	0.03
Pension characteristics					
Part-time pension	2,960	1,243	0.50	0.50	0.30
Expected ret. age 50-60	2,971	1,261	0.11	0.32	0.19
Expected ret. age 61-64	2,971	1,261	0.36	0.48	0.30
Expected ret. age 65[a]	2,971	1,261	0.44	0.50	0.32
Expected ret. age 66-70	2,971	1,261	0.09	0.29	0.19
Demographic variables					
Male	3,210	1,415	0.60	0.49	0.00
Partner: yes	3,210	1,415	0.75	0.43	0.08
Partner*male	3,210	1,415	0.47	0.50	0.06
Age	3,210	1,415	45.65	10.21	1.10
Homeowner	3,210	1,415	0.77	0.42	0.10
Household head	3,210	1,415	0.74	0.44	0.08
Number of children	3,210	1,415	0.95	1.13	0.20
Education					
Education low[a]	3,210	1,415	0.24	0.43	0.03
Education middle	3,210	1,415	0.32	0.47	0.06
Education high	3,210	1,415	0.44	0.50	0.05
Time					
May 2006[a]	3,210	1,415	0.21	0.40	0.34
June 2007	3,210	1,415	0.26	0.44	0.37
June 2008	3,210	1,415	0.21	0.41	0.33
June 2009	3,210	1,415	0.17	0.37	0.33
June 2010	3,210	1,415	0.15	0.36	0.3

[a] Baseline for categorical variables with more than 2 categories.

Table 2.7: Descriptive statistics of the satisfaction scales and measures of retirement expectations.

				Std. Dev.	
	N (obs)	n (res)	Mean	Overall	% within
Satisfaction scales[a]					
With personal pension provisions	2,274	1,067	5.97	1.81	29
With expected retirement age	2,266	1,073	5.52	2.16	30
With expected retirement income	2,188	1,036	5.88	1.88	30
With knowledge of personal provisions	2,265	1,061	5.95	1.94	31
With the retirement system in general	2,292	1,071	6.24	1.74	33
Expectations: log-normal[b]					
Median	2,593	1,128	78.78	18.01	61
Standard deviation	2,593	1,128	20.26	19.83	60
Expectations: splines[b]					
Median	2,482	1,102	77.09	18.55	61
Standard deviation	2,482	1,102	18.82	11.73	54

[a] We analyze satisfaction scales as dependent variables.
[b] Expectations are included among the independent variables.

Subjective distributions of replacement rates 2.B

This appendix presents two approaches to derive subjective probability distributions from the observed answers to the probabilistic questions. As explained in Section 2.3, we use the subjective probabilities that the replacement rate will be less than six numbers (100, 90, ..., 50%). These probabilities are interpreted as points on the subjective cumulative probability distribution function (cdf) of the replacement rate for the respondent at the given point in time. These probabilities are taken at face value – we do not incorporate measurement or recall error.[9]. We restrict the support of the replacement rate distributions to the interval $(20, 170)$, to limit the maximum uncertainty. This is necessary for cases where the replacement rate is reported to be equally likely to exceed 100 percent or less than 50 percent. Imposing these bounds gives two additional probabilities ($Pr(RR \leq 20) = 0$ and $Pr(RR \leq 170 = 1)$ treated in the same way as the reported subjective probabilities. The values 20% and 170% are chosen so that the bounds do not affect the estimated medians or standard deviations of more informed respondents.[10] Robustness checks show that results are not sensitive to varying the support of the subjective distributions.

Parametric approach 2.B.1

The first approach, proposed by Dominitz and Manski (1997), assumes that the reported probabilities follow from some parametric underlying distribution. Given the distribution and the replacement rate thresholds R_k specified in the questionnaire, the parameters θ_{it} of the distribution can be estimated by fitting the probabilities implied by the distribution, $F\left(R_k; \theta_{it}\right)$, to those reported in the data. Assuming that subjective distributions are lognormal, we can write

[9]We conducted robustness checks for subsamples in which measurement error is likely to be less important (for internally consistent sequences of responses only, or for age cohorts that are closest to retirement). These gave very similar results

[10]We checked this through the correlation between the estimates with and without upper and lower bounds, estimated using nonlinear least squares, for those observations that report no zeros or ones and do not violate monotonicity. For the first four waves of data we find a correlation between medians estimated with and without bounds of 0.98, N = 304. For the standard deviations we find a correlation of 0.88, N = 293.

$F\left(R_k;\theta_{it}\right)$ as:

$$F\left(R_k;\theta_{it}\right) = \Phi\left(\frac{\ln\left[R_k\right] - \mu_{it}}{\sigma_{it}}\right) \tag{2.1}$$

where $\Phi(.)$ is the standard normal cdf and μ_{it} and σ_{it} are respondent/year-specific parameters to be estimated.

The objective function defining the best possible fit chosen by Dominitz and Manski is the sum of the squared differences between implied and reported probabilities, so their approach can be referred to as Nonlinear Least Squares (NLS). Hence for each i and t, we choose the pair $\left(\mu_{it},\sigma_{it}\right)$ that solves the least squares problem:

$$\min_{\mu_{it},\sigma_{it}} \sum_{k=1}^{8} \left[F_{itk} - F\left(R_k;\mu_{it},\sigma_{it}\right)\right]^2 \tag{2.2}$$

The least squares solution is a degenerate log-normal distribution with $\sigma_{it} = 0$ whenever at least seven out of eight probabilities are equal to zero or one (since the first and the last probability are defined to equal zero and one respectively, this occurs whenever at least five out of six intermediate probabilities take the values zero or one). Such distributions are completely concentrated at some specific replacement rate, with standard deviation zero. Hence this approach interprets some individual-year combinations as having no subjective uncertainty regarding the replacement rate. Once the parameters of the underlying log normal distribution are estimated, it is easy to obtain the median and standard deviation of the subjective replacement rate (or other characteristics of the distribution), since these are functions of the estimated parameters.

An advantage of the nonlinear least squares approach for our data is that it can cope naturally with non-monotonic reported probabilities. Naturally the fitted distribution does not fit the reported probabilities very closely for such observations.

Non-parametric approach 2.B.2

A drawback of the NLS approach is that it requires the assumption of a parametric distribution function generating the responses, although the choice for specific subjective distributions can be based on empirical evidence concerning objective distributions Dominitz and Manski (1997). Since subjective and objective distributions do not necessarily coincide, however, it is desirable to compare the results of the parametric nonlinear least squares approach with results that do not rely on parametric assumptions. In particular, we focus on the relationship between the first two moments of the subjective distributions and pension satisfaction. The relevant question is not whether the assumed log normality of the distribution corresponds exactly to the expectations, but rather whether assuming log normality leads to misleading estimates of the effects of replacement rate expectations on pension satisfaction. To address this, we compare the parametric NLS technique with the nonparametric approach introduced by Bellemare et al. (2012), based upon spline interpolation. The spline approach does not require any assumptions on the subjective distributions except that they are continuous and satisfy some mild regularity conditions. See Bellemare et al. (2012) for details.

2.C FE ordered logit models estimated on monotonic subsample

Table 2.8: FE ordered logit models for the internally consistent subsample.

	Dependent variables: satisfaction with					
	Personal provisions					
	Overall[a]	Overall[b]	Ret. age	Benefits	Knowledge	The system
	Expectations modeled using splines					
Median/10	0.137***	-0.0337	0.00177	0.117***	-0.00918	0.0675
	(0.0399)	(0.117)	(0.0387)	(0.0445)	(0.0458)	(0.0474)
Std. dev./10	-0.00849	-0.109	0.0931	0.133	0.0417	0.0748
	(0.0884)	(0.214)	(0.0869)	(0.0941)	(0.0868)	(0.0947)
Informative observations	802	671	770	790	830	830
Informative respondents	261	216	251	258	270	272
Observations	1,526	1,433	1,522	1,463	1,525	1,534
Number of respondents	761	724	767	735	760	766
	Expectations modeled using log-normal distributions					
Median/10	0.149***	-0.0559	0.0187	0.177***	-0.0149	0.131**
	(0.0515)	(0.133)	(0.0493)	(0.0539)	(0.0560)	(0.0590)
Std. dev./10	-0.0143	-0.241	0.0235	0.112	0.0630	-0.0374
	(0.0695)	(0.160)	(0.0692)	(0.0771)	(0.0636)	(0.0648)
Informative observations	781	725	750	768	813	815
Informatie respondents	254	236	246	252	265	269
Observations	1,503	1,411	1,500	1,440	1,501	1,513
Number of respondents	753	717	760	727	752	759

[a] This specification does not control for satisfaction with aspects of one's pension.
[b] This specification does control for satisfaction with aspects of one's pension.
All coefficients and standard errors (in parentheses) are multiplied by 100.
All covariates from Table 6 are also included in these specifications.
***significant at 1%; **significant at 5%; *significant at 10%

Robustness checks: estimates on subsamples defined by age-group 2.D

Table 2.9: Robustness checks: sample limited to older respondents

	Dependent variable: satisfaction with					
	Personal provisions					
	Overall[a]	Overall[b]	Ret. age	Benefits	Knowledge	The system
	40 years and older - complete sample					
Median	0.0227***	0.0171*	0.00642	0.0156**	0.00314	0.00682
	(0.00665)	(0.00951)	(0.00599)	(0.00691)	(0.00621)	(0.00655)
Std. dev.	-0.00665	-0.0172	-0.00343	0.00816	0.000423	-0.00173
	(0.0105)	(0.0122)	(0.0104)	(0.0111)	(0.0111)	(0.0107)
Number of observations	1,254	1,194	1,251	1,210	1,250	1,263
Number of respondents	581	562	583	567	581	585
	40 years and older - internally consistent response					
Median	0.0255***	0.0155	0.00842	0.0205***	0.00109	0.00424
	(0.00861)	(0.0120)	(0.00700)	(0.00738)	(0.00656)	(0.00732)
Std. dev.	-0.00285	-0.0266	0.00256	0.0173	0.00467	0.00644
	(0.0144)	(0.0181)	(0.0142)	(0.0148)	(0.0151)	(0.0159)
Number of observations	1,068	1,014	1,066	1,027	1,063	1,073
Number of respondents	533	514	533	519	531	534
	50 years and older - complete sample					
Median	0.0199**	0.00835	-0.000319	0.0247**	0.00848	0.0107
	(0.00958)	(0.0113)	(0.00937)	(0.0106)	(0.00874)	(0.0102)
Std. dev.	0.00495	-0.00342	0.00608	0.00122	-0.00499	-0.00675
	(0.0128)	(0.0170)	(0.0121)	(0.0140)	(0.0139)	(0.0116)
Number of observations	743	722	745	725	743	749
Number of respondents	346	334	347	336	346	349
	50 years and older - internally consistent response					
Median	0.0230**	-0.00338	0.00973	0.0374***	0.0124	0.00691
	(0.0116)	(0.0161)	(0.0108)	(0.0114)	(0.00955)	(0.0112)
Std. dev.	-0.0101	-0.00420	0.000223	-0.00208	-0.0118	0.00204
	(0.0188)	(0.0281)	(0.0174)	(0.0190)	(0.0228)	(0.0197)
Number of observations	625	607	626	609	625	628
Number of respondents	310	299	309	301	310	311

[a] In this specification we do not control for aspect satisfaction.
[b] In this specification we control for aspect satisfaction.
Standard errors in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%

2.E Tests for selectivity from non-response to expectations questions

As explained in Appendix 2.B.2, the spline approach to constructing the subjective distribution functions is only applied to observations that contain at most 2 non-monotonic probabilities. Hence, we drop severely non-monotonic sequences from the analysis as far as it draws on splines. Doing so opens the door to endogenous sample selection, which occurs if unobserved factors that affect the likelihood of reporting a non-monotonic sequence are related to the unobserved factors driving pension satisfaction. To get a better idea of this selection process, the top panel of Table 2.10 reports the means of several covariates across the different subsamples that report monotonic probabilities, non-monotonic probabilities or no probabilities at all. For most variables the averages are similar across the categories of sample selection, suggesting that these subsamples are similar in terms of socioeconomic characteristics. The only exceptions are that the subsample that does not report a complete sequence of probabilities contains a higher fraction of females and a lower fraction of household heads compared to the other subsamples; and that respondents who report non-monotonic probabilities are more likely to have low education than the other groups.

Summary statistics of the medians and standard deviations of the estimated subjective distributions are reported in the lower panels of Table 2.10. We provide descriptives for monotonic and non-monotonic answers separately in order to verify whether large differences exist that need to be taken into account. For the estimated medians, we find that in the monotonic subsample the average median expected replacement rate is 76-77 percent for the earliest retirement option. Furthermore, there are large differences between the averages in the monotonic subsample compared to the non-monotonic subsample, especially for the NLS estimates. Therefore it is potentially important to estimate the effect of the median of the subjective distribution on pension satisfaction separately for those subsamples. With regard to uncertainty, we find evidence of substantial subjective uncertainty in the data, with the average estimated standard deviation equal to 15-17 percentage points at earliest retirement. Moreover, we find that average uncertainty is considerably larger in the non-monotonic subsample. Based on these findings, we will conduct robustness

Table 2.10: Descriptive statistics: sample selection.

	Complete				Incomplete	
	Monotonic		Not monotonic			
	N	Mean	N	Mean	N	Mean
Ln(net income)	2,033	7.44	636	7.37	379	7.35
Male	2,115	0.62	659	0.6	436	0.51
Partner	2,115	0.75	659	0.77	436	0.75
Age	2,115	45	659	46	436	47
Household head	2,115	0.75	659	0.75	436	0.68
Education low	2,115	0.21	659	0.33	436	0.26
Education middle	2,115	0.31	659	0.36	436	0.33
Education high	2,115	0.48	659	0.31	436	0.41
Medians						
Log-normal	2,108	76.01	639	87.22		
Spline interpolation	2,115	76.78	367	78.92		
Std. dev.						
Log-normal	2,079	15.01	514	41.49		
Spline interpolation	2,115	17.38	367	27.1		

checks for all estimated equations in which we allow the effect of expectations to differ between answer sequences that are internally consistent and those that are not.

Combining non-response and internally inconsistent answers, about 30 percent of potential respondents do not provide a usable response to the expectations questions. Such rate of non-response and internally inconsistent responses raises worries of selection issues, which would occur if those who select themselves out of the sample have different average satisfaction levels conditional on expectations compared to those who do answer the items. There is one important difference between the sample selection problem described above and that commonly discussed in the literature, namely that selection applies to independent rather than dependent variables. In other words: we do observe the satisfaction scales for most of the respondents that do not report (useful) probabilities. Therefore, we can compare the average satisfaction of those who do report probabilities to those who do not (conditional on observed covariates). If expectations affect wellbeing and selectivity in expectations is present, we would expect that dummies indicating non-response would be significant predictors of pension satisfaction. Table 2.11 reports estimates from

Random Effects (RE) ordered logit models that regress the satisfaction scales on dummies indicating the two forms of sample selection that we observe. The top half of the table uses indicators for selection based the questions related to earliest retirement, while the bottom half investigates selection for the latest retirement questions. Within each half, the top panel does not include controls whereas the bottom one does.[11] We report both the parameter estimates on the selection dummies and tests of their joint significance. From Table 2.11 we conclude that even though average satisfaction differs significantly across the selection subsamples for some scales, these differences largely disappear once we control for socioeconomic covariates of pension satisfaction. The dummies indicating sample selection are jointly insignificant at the 5 percent level for all satisfaction scales once we control for socioeconomic characteristics. Hence sample selection based on non-response to the subjective probability questions does not appear to be an important problem in our equations explaining the satisfaction scales.

[11] All controls from Tables 1 and 2 are included, except for those describing expectations.

Table 2.11: RE ordered logit models of satisfaction - selectivity through non-monotonic/incomplete response.

	Dependent variables: satisfaction with					
	Personal provisions					
	Overall[a]	Overall[b]	Ret. age	Benefits	Knowledge	The system
No controls						
Non-response	-0.0922	-0.123	0.547***	0.00739	0.0555	-0.00616
	(0.210)	(0.161)	(0.200)	(0.213)	(0.200)	(0.190)
Inconsistent response	-0.280**	-0.0851	-0.112	-0.369***	-0.143	-0.245*
	(0.138)	(0.116)	(0.134)	(0.137)	(0.132)	(0.130)
Joint test (chi2, 2df)	4.14	0.96	9.08**	7.44**	1.35	3.61
N	2,274	2,133	2,266	2,188	2,265	2,292
Controls						
Non-response	-0.386	-0.162	-0.240	-0.118	0.00789	-0.481*
	(0.289)	(0.240)	(0.288)	(0.292)	(0.271)	(0.263)
Inconsistent response	-0.267*	-0.0848	-0.0442	-0.334**	-0.106	-0.166
	(0.140)	(0.121)	(0.138)	(0.141)	(0.134)	(0.131)
Joint test (chi2, 2df)	4.84*	0.84	0.74	5.66*	0.64	4.43
N	2,078	1,952	2,069	1,999	2,072	2,094

[a] In this model we do not control for satisfaction with aspects of one's pension.
[b] In this model we do control for the aspect satisfaction scales.
Standard errors in parentheses.
In the lower panel the controls from Table 3 are included in all specifications.
***significant at 1%; **significant at 5%; *significant at 10%

3 Survey Response in Probabilistic Questions and Its Impact on Inference

This chapter is a reproduction of De Bresser and Van Soest (2013b), which is forthcoming in Journal of Economic Behavior and Organization.

3.1 Introduction

The idea of forward-looking agents who take account of the future consequences of their decisions has long been commonplace among economists. The behavior of such agents is driven by their preferences and expectations, neither of which are usually directly observed. Before the 1990s almost all researchers pinned down expectations by assuming rationality, so that expectations coincide with actual stochastic processes that can be estimated. From the early 1990s economists have started to ask survey respondents about their expectations, allowing rationality assumptions to be verified and relaxed. Since then questions on subjective expectations have been included in several large socio-economic surveys like the Michigan Survey of Consumers, the Health and Retirement Study and the Survey of Economic Expectations, see Dominitz and Manski (2006). To simplify interpretation and interpersonal comparisons, expectations are often measured by means of items that ask respondents to think in terms of probabilities. The quality of this kind of data has been questioned; as noted by Manski (2004), the mere fact that surveys come up with relatively high response rates on probabilistic items does not imply that the responses indeed reflect respondents' true beliefs.

In this paper, we analyze the validity of survey items used to measure expectations by simultaneously modeling expectations and various aspects of survey response. We draw on a representative, longitudinal sample of Dutch employees aged 25 – 64 who were asked about their expected replacement rate of income at retirement. Our measure of the replacement rate compares real expected pension income with real current income. We adjust the model of Kleinjans and Van Soest (2013) for probabilities of binary events to a situation in which several probabilistic items are used to fit subjective log-normal distributions of a continuous outcome along the lines of Dominitz and Manski (1997). Location and dispersion measures are allowed to depend upon observed and unobserved respondent characteristics. In our data, we have questions on the probability that the retirement replacement rate is less than 50%, 60%, 70%, 80%, 90% and 100%, for five waves of panel data (2006 – 2010). We model non-response, non-informative answers and rounding, issues which have been identified as potentially important aspects of the answering process of subjective probability questions (Bruine de Bruin et al. 2000, Dominitz and Manski 1997). By comparing with models ignoring these aspects, we also analyze the impact of such processes on the conclusions regarding the location and dispersion of the fitted distributions.

Retirement income in the Netherlands is organized along three pillars: pay-as-you-go public pensions, fully funded occupational pensions and private savings (Bovenberg and Meijdam 2001). The first two together provide 80 percent of income during retirement for the average employee. The first pillar pay-as-you-go public pension provides retirees with a subsistence level income; it neither reflects individual contributions nor previously earned income. The second pillar consists of fully funded occupational pensions that cover about 90 percent of Dutch employees. These industry- or company-specific provisions are usually defined benefit schemes, though defined contribution arrangements are becoming more common.[1] In contrast to the public pension, the level of occupational benefits is based on previous wages through average or final wage arrangements.

The third pillar of retirement income consists of private savings, but due to the comprehensive level of the first two pillars, private arrangements are

[1]Compensation for inflation (indexation) is not guaranteed so that only nominal benefits are guaranteed.

a relatively minor part of pension income for most employees. The first and second pillar pensions are probably easier to predict than defined contribution pensions or third pillar pension savings (often invested in the stock market), making the Dutch pension system especially suitable for research on retirement expectations. In our survey, pension income is defined to include the first two pillars, discarding potential income from the third pillar.

Our results indicate that there is statistically significant and quantitatively important unobserved heterogeneity at the level of the individual in all parts of the model. Especially, there is a small group of respondents that consistently give non-informative focal point (fifty-fifty) answers. Moreover, the individual effects of the various aspects of survey response are correlated among each other and with the unobserved heterogeneity terms in the parameters of the location and dispersion of the subjective distributions. Though we also find some statistically significant sequence effects (effects common to the series of six questions answered by a given respondent in one specific wave), these effects are small compared to the individual effects that are persistent over time. We find rounding is important, with almost 50 percent of the answers rounded to a multiple of 10. Non-informative fifty-fifty answers, on the other hand, make up only 1.3 percent of the reported probabilities.

A full specification including covariates indicates that non-response, rounding, and reporting error vary significantly with socio-economic respondent characteristics, as well as the location and dispersion of the underlying subjective probability distributions. Finally, the magnitude and significance of the relationships between the location and dispersion of the subjective distributions and the covariates is larger in the joint model of the answering process and expectations than in a linear random effects model that ignores rounding, non-response, and focal answers. This suggests it is useful to take the answering process into account when analyzing the association between subjective expectations and respondent characteristics.

The structure of this paper is as follows. Section 3.2 summarizes related literature, both with regard to subjective (pension) expectations and the econometric model applied. Section 3.3 introduces the data and provides descriptive statistics. Section 3.4 describes the model in detail. Estimation results are presented in section 3.5. Section 3.6 concludes.

3.2 Literature

This paper fits in with the literature on probabilistic expectations in general and pension expectations in particular. Expectations play an important role in economic models of intertemporal choice. Rational, forward-looking agents not only consider the present consequences of their actions, but also (their perception of) their impact on future utility. For instance, Delavande and Rohwedder (2011) show that individuals who are uncertain about their future level of social security benefits tend to hold a smaller share of their portfolio in stocks, reducing the riskiness of their retirement provisions. Direct measurement of expectations is especially important to separately identify expectations and preferences – different combinations of preferences and expectations may be behaviorally equivalent, but may yield different policy implications (Manski 2002).

Spurred by a growing body of evidence in favor of survey respondents' willingness and ability to answer questions in a probabilistic format, measuring expectations directly through probabilistic questions gained momentum during the late 1990s and early 2000s and has become the standard elicitation method (see Manski 2004, and Hurd 2009, for recent overviews). Compared to earlier strategies, the advantages of probabilistic expectations are ease of interpretation, interpersonal comparability, and the ability to characterize subjective uncertainty (Dominitz 1998). Moreover, subjective expectations data appear not to be affected by cognitive biases such as cognitive dissonance or biased recall (Zafar 2011).

Subjective probabilities allow straightforward analysis of the distribution of a binary outcome: we need only one probability. A popular approach to analyze the subjective distributions of a continuous outcome variable has been to elicit a number of subjective probabilities identifying points on the subjective cumulative distribution function, and then fit a parametric distribution using nonlinear least squares (see Dominitz and Manski 1997). Once the parameters of the distributions are estimated, one can analyze various moments of the distributions, such as the median, standard deviation or interquartile range (Dominitz 1998, Dominitz and Manski 1997, 2006).

Because retirement is a major life-event, interest in expectations regarding retirement has been especially strong. Early research comparing expectations

to administrative data found evidence for widespread misinformation and missing information among workers about basic features of their pension plan (such as early retirement opportunities and whether the system was defined benefit or defined contribution, Mitchell 1988). Subsequent research has confirmed this finding (see Gustman and Steinmeier 2005, for the US and AFM 2010, using Dutch data). Dominitz and Manski (2006) applied the approach of fitting parametric distribution functions to represent retirement expectations. They found considerable heterogeneity in beliefs about the long-term sustainability of the public pension system in the US, with younger people being more concerned. Moreover, Americans are uncertain about future benefits conditional on existence of the system, even the middle-aged respondents.

Van Santen et al. (2012) use the same data we will analyze (though fewer waves). They observe that almost a third of the sample answer the replacement rate expectations questions in ways that violate the laws of probability (probabilities add up to more than 100% or violate monotonicity of the cumulative distribution function). Using Heckman selection models, they then analyze the effect of removing logically inconsistent observations on their subsequent analysis of expectations. They conclude that doing so induces endogenous sample selection, leading to underestimation of the medians of the subjective distributions and overestimation of subjective uncertainty, especially among the poorly educated. As will be explained in section 3.4, our approach which explicitly models the errors induced by the answering process offers an alternative interpretation of the inconsistent observations. Instead of discarding them as completely uninformative, we regard such responses as potentially informative, but relatively noisy.

We base our model on Kleinjans and Van Soest (2013).[2] They model the response to a single subjective probability question in a panel data setting, taking into account the issues of selection, rounding and non-informative answers. They also motivate their model from the psychological literature on how respondents answer subjective probabilities; see the references in their Section 2. We change their basic setup in three ways. Firstly, our objective is

[2]See also Hudomiet et al. (2011), who present a structural model of stock market expectations allowing for rounding and focal points. They assume all reported probabilities are rounded to a multiple of 10 while we allow for multiple forms of rounding. Furthermore, they consider a single question and do not model item non-response, focal answers or internally inconsistent response sequences.

not to model the answers to a single probabilistic question, but rather to model moments of the distribution functions characterizing expectations. To this end we approximate subjective expectations by log-normal distributions in the spirit of Dominitz and Manski (1997). Secondly, since we analyze sequences of six subjective probabilities, we take into account that some processes occur at the level of a sequence (e.g. sample selection) while others apply to the individual answers in a sequence (e.g. rounding). The structure of our data enables us to distinguish between two levels of unobserved heterogeneity, which we interpret as individual effects (which remain constant across all questions answered by a given individual in all survey waves) and sequence effects (which are fixed only across the questions of a given respondent in one specific wave). Thirdly, we offer two alternative interpretations of logically inconsistent probabilities. In our main model specification we interpret them as potentially informative yet noisy: perturbed by an idiosyncratic recall or reporting error. The other interpretation that we explore is that of another form of selection out of the sample, implying that internally inconsistent sequences of probabilities are on the whole uninformative about subjective expectations. We investigate this alternative model to see whether response behavior differs fundamentally between the consistent and inconsistent subsamples. However, we do not report it in the main text because we find that its added complexity does not improve model fit.

3.3 Data

3.3.1 Dataset and phrasing of the questions

The data are taken from the yearly surveys of the Netspar Pension Monitor ("Pensioenbarometer"). This survey is administered to participants of the CentERpanel,[3] an ongoing online panel survey administrated by CentERdata at Tilburg University. The CentERpanel is representative for the population in the Netherlands of ages 16 and older and is composed of over 2000 households in which one or more adults are invited to complete questionnaires at home every week over the Internet. Households are randomly selected and those

[3] See http://www.centerdata.nl/en/centerpanel.

without Internet access are provided with access and the necessary equipment by CentERdata. About 75% of all panel members respond to the questions in a given weekend. Attrition is low, making longitudinal research possible. Rich background information about the panel respondents is available from previous interviews.

The yearly questionnaires of the Pension Monitor are distributed to all CentERpanel members of ages 25 and older – younger respondents were assumed not to have thought seriously about retirement yet. We draw on the five waves from 2006 until 2010. The surveys were done in June, except in 2006, the pilot phase, when it was done one month earlier. The replacement rate questions are asked only to respondents who indicate that their "most important activity" is wage labor, so we cannot analyze expectations of the self-employed. The focus on employees implies that when we refer to pension income, this includes the old age state pension in the first pillar and the occupational pensions of the second pillar, as is also emphasized in the questions (see below).

The items of interest ask respondents about the probabilities that their replacement rate at earliest retirement will be below a series of thresholds, ranging from 50 to 100 percent. Before eliciting the probabilities, respondents are asked at what age they expect to first have the possibility to retire (assuming they stay with their current employer). This age is then inserted into the probability questions, which are phrased as follows:

> If you would retire at [earliest retirement age], please consider your net total pension income including public pension, relative to your present net wage or salary. What would you think is the probability that your net total pension income in the year after retirement will be worth in terms of purchasing power
>
> a. More than 100% of your present net wage?
>
> b. Less than 100% of your present net wage?
>
> c. Less than 90% of your present net wage?
>
> ...
>
> g. Less than 50% of your present net wage?

In order to prevent complications with mixed discrete-continuous subjective distributions, we only use the answers to questions b through g. These can be challenging for the respondents. The links between pension income and final or average wages require the respondents to know or estimate their wage profile and integrate this with the rules of their occupational pension. The difficulty of this task might deteriorate data quality and lead to rounding, non-response or focal answers.

3.3.2 Descriptive statistics

Table 3.1: Variable definitions

Variable	Description
HH. income <€1,150	Net monthly household income less than €1,150
HH. income €1,150-€1,800	Net monthly household income €1,150-€1,800
HH. income €1,800-€2,600	Net monthly household income €1,800-€2,600
HH. income >€2,600	Net monthly household income above €2,600
Construction	Respondent employed in construction sector
Trade	Respondent employed in trade sector
Transport	Respondent employed in transport sector
Financial and business services	Respondent employed in financial and business services sector
Education	Respondent employed in educational sector
Healthcare	Respondent employed in healthcare sector
Governance	Respondent employed in governance sector
Other	Respondent not employed in any of the above sectors
Part-time pension	Respondent has the option of part-time retirement
Prob. purchasing power increase	Probability that the purchasing power of household is higher one year from now
Prob. purchasing power decrease	Probability that the purchasing power of household is lower one year from now
Expected retirement age 50-60	Respondent expects retirement before age 60
Expected retirement age 61-64	Respondent expects retirement between ages 61-64
Expected retirement age 65	Respondent expects retirement at age 65
Expected retirement age 66-70	Respondent expects retirement after age 65
Male	Respondent is male
Age	Respondent's age
Partner	Respondent lives with partner
Male*partner	Interaction term of being male and living with partner
Children	Number of children in the respondent's household
Education: middle	Finished secondary school/lower vocational training
Education: high	Higher vocational training/university
Household head	Respondent is self-reported head of the household
Homeowner	Respondent is (co-)owner of his house

We limit our sample to respondents between the ages of 25 and 64 who report wage labor as their primary activity and indicate an expected age of earliest

retirement between 50 and 70 years. The definitions of the independent variables used in our models are given in Table 3.1. Table 3.2 provides the usual summary statistics of dependent and independent variables. For 93 percent of the observations we observe a complete sequence of probabilities. Furthermore, non-response appears to occur homogeneously across the sample: the sample averages of the socio-economic characteristics are very similar whether or not we limit our sample to those who answer all replacement rate questions. Similarly to Van Santen et al. (2012), we find that 22 percent of respondents report probabilities that violate monotonicity of the cumulative distribution function.[4] The average expected retirement age is 63.6 years and 44 percent of respondents expect to be able to retire when they are 65, which is the current eligibility age for the Dutch public pension. When we estimate the model we combine the income groups with household income below 1,150 euro and in the 1,150-1,800 range, because only two percent of observations fall in the lowest income category. (We use dummies for income categories rather than a continuous measure for income because the categorical measure in the data has fewer missing values.)

Summary statistics for the reported probabilities are reported at the top of Table 3.2. The averages obey monotonicity of the cumulative distribution function. Moreover, there is considerable variation in expectations, both between respondents and across different waves for a given respondent. Non-response is described in more detail in Table 3.3, showing that 2,733 out of the 2,952 sequences are complete. Of the remaining 219 incomplete sequences (7.4 percent of the total number of observations), 154 are missing completely and only 7 contain one to three missing responses. This pattern of missing data suggests that non-response mostly occurs at the level of the question sequence, rather than at the level of the individual answers. This is why we will not model the distinction between partial and complete nonresponse and treat partially missing sequences as if they were missing completely.

Table 3.4 shows that reporting a probability equal to 50 percent does not tend to occur at sequence-level. Out of the 2,733 complete sequences, 60

[4]Van Santen et al. (2012) report that 33% of their sample violate the laws of probability. Most of the discrepancy is explained by the fact that they take into account both adding up errors and violations of monotonicity while we consider violations of monotonicity only. Adding-up errors make up 9% of their data; the incidence of monotonicity violations in their data is similar to ours.

Table 3.2: Descriptive statistics

			Mean		Std. Dev.	
	# Obs.	# Resp.	Overall	Response[a]	Overall	Within
Pr(RR<100)	2,749	1,165	68.36	–	39.25	23.89
Pr(RR<90)	2,747	1,162	60.59	–	37.44	23.15
Pr(RR<80)	2,750	1,167	53.08	–	34.39	21.79
Pr(RR<70)	2,759	1,173	39.66	–	31.91	20.43
Pr(RR<60)	2,742	1,162	24.37	–	27.53	17.47
Pr(RR<50)	2,739	1,164	16.13	–	24.70	16.39
HH. income <€1150	2,952	1,256	0.02	0.02	0.15	0.06
HH. income €1150-€1800	2,952	1,256	0.20	0.20	0.40	0.15
HH. income €1800-€2600	2,952	1,256	0.29	0.29	0.45	0.20
HH. income >€2600	2,952	1,256	0.49	0.48	0.50	0.16
Sector: industry	2,952	1,256	0.16	0.16	0.37	0.06
Sector: agriculture, water, energy	2,952	1,256	0.03	0.03	0.16	0.05
Sector: construction	2,952	1,256	0.04	0.04	0.20	0.04
Sector: trade	2,952	1,256	0.10	0.10	0.30	0.05
Sectpr: transport	2,952	1,256	0.03	0.03	0.18	0.04
Sector: financial and business services	2,952	1,256	0.16	0.16	0.37	0.07
Sector: education	2,952	1,256	0.14	0.14	0.35	0.05
Sector: healthcare	2,952	1,256	0.18	0.18	0.39	0.06
Sector: governance	2,952	1,256	0.12	0.12	0.33	0.06
Other: food & interest	2,952	1,256	0.03	0.03	0.17	0.03
Partime pension	2,905	1,227	0.50	0.50	0.50	0.29
Prob. PP increase during coming year	2,872	1,214	2.48	2.51	2.82	1.64
Prob. PP decrease during coming year	2,872	1,214	4.03	4.08	3.44	2.17
Male	2,952	1,256	0.60	0.61	0.49	0.00
Male*partner	2,952	1,256	0.48	0.48	0.50	0.06
Age	2,952	1,256	45.34	45.46	9.86	1.12
Partner	2,952	1,256	0.75	0.76	0.43	0.08
Homeowner	2,952	1,256	0.77	0.77	0.42	0.10
Household head	2,952	1,256	0.75	0.75	0.44	0.09
Number of children	2,952	1,256	0.97	0.97	1.14	0.20
Education low	2,952	1,256	0.24	0.24	0.43	0.03
Education middle	2,952	1,256	0.32	0.32	0.47	0.06
Education high	2,952	1,256	0.44	0.44	0.50	0.05
Expected ret. age	2,952	1,256	63.55	63.57	2.35	1.36
Exp. retirement age 50-60	2,952	1,256	0.11	0.11	0.32	0.19
Exp. retirement age 60-64	2,952	1,256	0.36	0.36	0.48	0.30
Exp. retirement age 65	2,952	1,256	0.44	0.44	0.50	0.32
Exp. retirement age 66-70	2,952	1,256	0.09	0.09	0.28	0.19

[a] Sample size: 2,733 observations.

Table 3.3: Item non-response by question sequence

	Overall		By wave (%)				
# Missings	N	%	2006	2007	2008	2009	2010
0	2,733	92.58	95.09	87.72	95.36	94.75	92.16
1	3	0.10	0.16	0.12	0.00	0.00	0.21
2	1	0.03	0.00	0.12	0.00	0.00	0.00
3	3	0.10	0.16	0.00	0.00	0.00	0.41
4	2	0.07	0.16	0.00	0.00	0.00	0.21
5	56	1.90	1.15	1.74	2.13	1.69	3.09
6	154	5.22	3.27	10.3	2.51	3.56	3.92
Total	2,952	100	100	100	100	100	100

percent contain no 50/50s and 38 percent report a 50/50 answer once or twice. We would expect that giving focal point answers does not change from one question to the next. Hence, this suggests that the 50/50 answers are informative, though perhaps rounded, probabilities rather than uninformative focal point answers.

Table 3.4: Number of 50% answers per question sequence

	Overall		By wave (%)				
# 50/50s	N	%	2006	2007	2008	2009	2010
0	1,630	59.64	61.27	61.24	60.04	57.23	57.27
1	938	34.32	31.84	32.81	34.48	36.63	37.14
2	103	3.77	3.96	3.39	3.85	4.16	3.58
3	17	0.62	1.03	0.71	0.20	0.40	0.67
4	7	0.26	0.34	0.42	0.00	0.40	0.00
5	6	0.22	0.17	0.28	0.41	0.20	0.00
6	32	1.17	1.38	1.13	1.01	0.99	1.34
Total	2,733	100	100	100	100	100	100

More information on the importance of focal answers is given by the fraction of 50/50s across the different replacement rate thresholds. Table 3.5 shows that subjective probabilities equal to 50 percent are quite common for the questions whether the replacement rate is lower than 70 or 80 percent, which seems plausible since these thresholds are often close to the median of the subjective distributions. Respondents do not often reply with 50 percent for thresholds in the tails of the distribution – here answers equal to 0 and 100 percent are

more common. This implies that most respondents are rather certain that their replacement rate will lie in the 50 – 100 range. The pattern of 50/50s across replacement rates is consistent with rounding and genuine uncertainty rather than focal answers due to a lack of ability to answer the questions.

Table 3.5: Frequencies of 50% answers across replacement rate cutoffs

	Cutoff: Probability RR <K						
% equal to	K = 100	K = 90	K = 80	K = 70	K = 60	K = 50	All thresholds
0	12.66	12.08	11.97	17.39	29.61	41.80	19.37
50	5.82	5.82	11.82	14.17	8.78	6.48	8.16
100	41.40	23.06	15.19	7.28	3.48	2.93	14.40
Sum	59.88	40.96	38.98	38.84	41.87	51.21	41.93
N	2,733	2,733	2,733	2,733	2,733	2,733	16,398

Histograms of the reported probabilities for the six thresholds are presented in Figure 3.1. They all show clear evidence of peaks around multiples of 10, suggesting that rounding is important. Moreover, the overall pattern shows that the fraction of 50/50 answers is not disproportionally large compared to the numbers of 0 and 100 percent answers. This supports the notion that the observed 50s are caused by crude rounding rather than by focal answers.

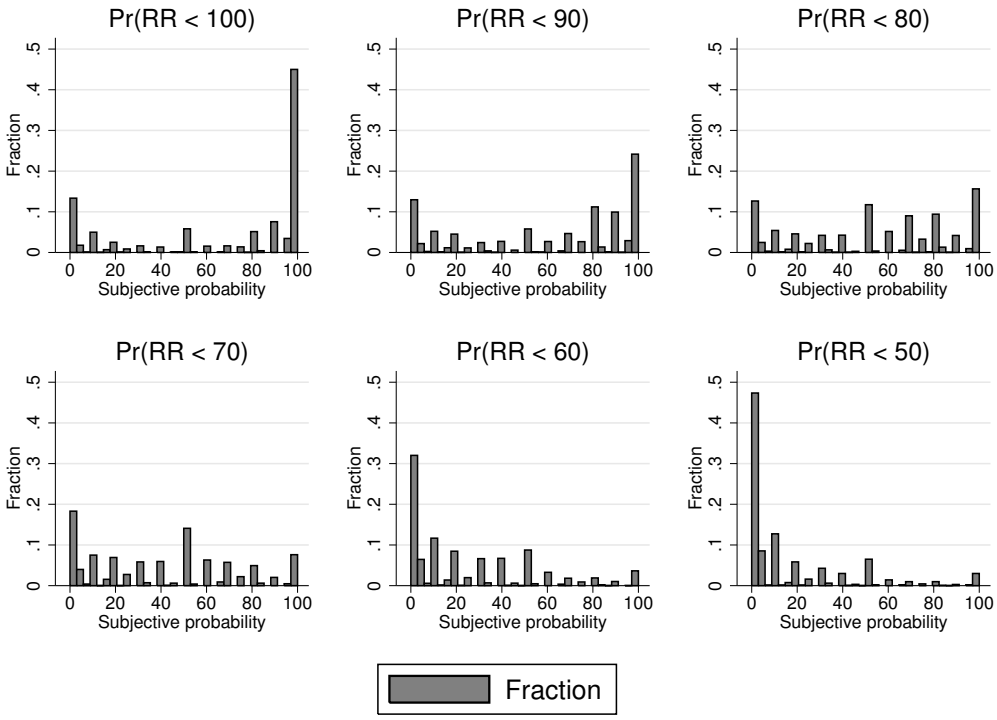


Figure 3.1: Histograms of reported probabilities by threshold

Econometric model 3.4

Our model captures beliefs through parametric probability distributions along the lines of Dominitz and Manski (1997), accounting for rounding, focal points, and non-response as in Kleinjans and Van Soest (2013). The dependent variables are the respondent-wave specific sequences of six reported subjective probabilities that the replacement rate of income at retirement compared to current earnings will be below the thresholds 100, 90, 80, 70, 60 and 50 percent. We assume the true subjective distribution of respondent i in year t is log-normal with parameters μ_{it} and σ_{it}. A parametric assumption is necessary to keep the econometric model parsimonious and the log-normal distribution has the advantage that it has two parameters that have a clear interpretation: μ_{it} captures the level and σ_{it} the dispersion of the log replacement rate.[5]

[5]Another motivation is that actual incomes are often found to be approximately log-normally distributed; see, e.g., Soltow and Van Zanden (1998) for the Netherlands. If the logs of pre- and post-retirement income are jointly normal, then the log replacement rate is also normally distributed. But of course expectations need not be rational so subjective distributions could still be different. Using the same data as we use, De Bresser and Van Soest (2013a) compare the

The parameterization of expectations as log-normal leads to the following probabilities (as percentages):

$$P_{itk}^{true} = \Phi\left(\frac{\ln\left[RR_k\right] - \mu_{it}}{\sigma_{it}}\right) \times 100 \tag{3.1}$$

Here $\Phi(.)$ denotes the standard normal distribution function and RR_k represents the k-th threshold replacement rate specified in the questionnaire: $RR_1 = 100, \ldots, RR_6 = 50$. Reported probabilities can be affected by reporting errors. These are incorporated in a latent variable P_{itk}^{*} given by:

$$\begin{aligned} P_{itk}^{*} &= P_{itk}^{true} + \varepsilon_{itk}^{P} \\ \varepsilon_{itk}^{P} &\sim N\left(0, \sigma_{\varepsilon^P}^2\right) \end{aligned} \tag{3.2}$$

The error term ε_{itk}^{P} is similar to the implicit error in the non-linear least squares approach of Dominitz and Manski (1997). In contrast to Dominitz and Manski (1997) we do not estimate μ_{it} and σ_{it} separately for each observation but model them (see below). Our model specification ensures that the true probabilities are between 0 and 100 and are strictly decreasing with k. The reporting errors, however, make it possible that observed probability sequences are sometimes non-monotonic (in line with the data). Reporting errors ε_{itk}^{P} are assumed to be independent across and within sequences (ε_{itk}^{P} is independent of ε_{jsr}^{P} for all $j \neq i$, $s \neq t$ and $r \neq k$). Since we expect that the magnitude of these errors varies across socio-economic groups, we allow ε_{itk}^{P} to be heteroscedastic. Hence we also include an equation allowing for multiplicative heteroscedasticity:

$$\ln\left[\sigma_{\varepsilon^P}\right] = x_{it}'\beta_1 \tag{3.3}$$

Where β_1 is a parameter vector to be estimated and x_{it} is a vector of observed socio-demographic variables. This feature of the model accommodates the observation that non-monotonic sequences are more prevalent among certain groups, such as less educated respondents.

medians and standard deviations of estimated log-normal distributions with non-parametric estimates; they find large correlations between the two sets of estimates and robust correlations with other variables of interest (demographics, pension satisfaction).

We model the parameters μ_{it} and σ_{it} in Equations 3.1 and 3.2 as follows:

$$\mu_{it} = x'_{it}\beta_2 + \nu^{\mu}_{it} \tag{3.4}$$

$$\ln\left[\sigma_{it}\right] = x'_{it}\beta_3 + \nu^{\sigma}_{it} \tag{3.5}$$

Here μ_{it} and $\ln\left[\sigma_{it}\right]$ are decomposed into two parts, reflecting observed heterogeneity ($x'_{it}\beta$) and unobserved heterogeneity (ν_{it}). We model unobserved heterogeneity as the sum of two error components, one that is fixed over time and across questions for a given respondent (an individual effect α^h_i), and one that is fixed for a given sequence of six answers from a given respondent in a given wave, but may vary across waves for the same respondent (a sequence effect ζ^h_{it}):

$$\nu^h_{it} = \zeta^h_{it} + \alpha^h_i \tag{3.6}$$
$$h = \mu, \sigma$$

The sequence effect ζ^h_{it} can be interpreted as the influence of the mood a respondent is in when answering the questions. For instance, a positive mood might raise the median expected replacement rate or reduce the subjective uncertainty. Alternatively, it may also reflect genuine transitory changes in expectations. The individual effect α^h_i may reflect variation in actual replacement rates but can also be interpreted in terms of personality types: an optimistic person might consistently have a higher expected replacement rate or experience less uncertainty. Distributional assumptions for the components of ν^h_{it} will be given below.

An implication of this decomposition of ν^h_{it} is that we allow the random parts of the location and of the dispersion of the subjective distributions to be correlated across repeated observations for a given respondent. Also, we allow this correlation to be stronger across answers to different questions posed in the same survey wave.

The reported probabilities, in percentage points, are denoted P_{itk} and may either be an integer between 0 and 100 or missing. As shown in Figure 3.1, most respondents give answers that are multiples of 10 or other multiples of 5, suggesting that rounding plays a role. Following Kleinjans and Van Soest (2013), we allow for rounding to the nearest multiple of 1, 5, 10, 25 or 50 and

model the extent of rounding as an ordinal process, with rounding to multiples of 1 and 50 as the least and most extreme amounts of rounding:

$R_{itk} = 1$: the probability is rounded to a multiple of 1

$R_{itk} = 2$: the probability is rounded to a multiple of 5

$R_{itk} = 3$: the probability is rounded to a multiple of 10

$R_{itk} = 4$: the probability is rounded to a multiple of 25

$R_{itk} = 5$: the probability is rounded to a multiple of 50

We allow the amount of rounding to vary across answers within a sequence (and across sequences). For each question we use the following ordered response model:

$$\begin{aligned} R^*_{itk} &= x'_{it}\beta_4 + \zeta^R_{it} + \alpha^R_i + \varepsilon^R_{itk} \\ R_{itk} &= j \text{ if } m_{j-1} < R^*_{itk} \leq m_j \text{ with } j = 1, ..., 5 \\ \varepsilon^R_{itk} &\sim iid\ N(0,1) \end{aligned} \tag{3.7}$$

As in equations 3.5 and 3.6, we distinguish between two levels of clustering in the error term by including sequence effects as well as individual effects, in addition to the idiosyncratic term ε^R_{itk}.

We allow for two other behavioral processes associated with survey response to probabilistic questions: item non-response and non-informative focal point answers. We define focal point answers as giving a number that is completely unrelated to the true subjective probability and assume that the number in this case is always 50 percent. As motivated in the previous section, we model the uninformative 50/50s on a question-by-question basis.[6] Since the decision whether to give informative answers or not is binary, we use a logit model:

$$\begin{aligned} y^{F*}_{itk} &= x'_{it}\beta_5 + \zeta^F_{it} + \alpha^F_i + \varepsilon^F_{itk} \\ y^F_{itk} &= 1 \text{ if } y^{F*}_{itk} > 0,\ y^F_{itk} = 0 \text{ otherwise} \\ \varepsilon^F_{itk} &\sim iid \text{ logistic} \end{aligned} \tag{3.8}$$

[6]Though we model focal point answers on a question-by-question basis, the concept of non-informative answers arising from epistemic uncertainty suggests that individual effects are likely to be very important in this equation. We expect a lack of understanding of the kind of questions that we analyze to be quite persistent over time, especially since the sample consists of working-age individuals who finished their education.

With $y^F_{itk} = 1$ if respondent i provides a non-informative 50/50 answer to question k in period t.

Finally, as explained in section 3.3.2, we model non-response as a single decision for the sequence as a whole: respondent i in year t answers all six or none of the probabilistic questions. Again, we use a logit model:[7]

$$\begin{aligned} y^{S*}_{it} &= x'_{it}\beta_6 + \alpha^S_i + \varepsilon^S_{it} \\ y^S_{it} &= 1 \text{ if } y^{S*}_{it} > 0,\ y^S_{it} = 0 \text{ otherwise} \\ \varepsilon^S_{it} &\sim iid \text{ logistic} \end{aligned} \tag{3.9}$$

Here $y^S_{it} = 1$ if respondent i did not answer the sequence in period t. Note that in the selection equation sequence effects are subsumed in the error term ε^S_{it}. Distributional assumptions on the individual effects are given below.

Figure 3.2 summarizes the model of the behavioral process of answering sequences of probabilistic questions. The model is sequential, with respondents first deciding whether or not to provide answers in a particular wave. If they do respond, they decide for each question separately whether to provide an informative or a focal 50/50 answer. In case of an informative answer respondents report a certain probability P_{itk}, possibly affected by reporting error and rounding.

Since nonresponse is observed, the selection equation could be estimated separately and is identified. Error terms are assumed to be independent of each other (and correlation enters through the individual effects), hence the identification problem in cross-section selection models does not arise and exclusion restrictions are not needed. The identification of rounding and focal 50/50s is based on the relative frequencies of probabilities that result from different rounding rules. For instance, the fact that we observe few probabilities that can only result from rounding to multiples of 1 (like 21% or 22%, but not 20% or 25%) suggests that few respondents round to multiples of 1. Similarly, we observe relatively many probabilities that can result from rounding to multiples of 1, 5 or 10, such as 10% or 20%, but fewer probabilities that can be rounded to multiples of 1 or 5, such as 5% or 15%, showing that rounding to multiples of 10% is more prevalent than rounding to multiples

[7] We prefer logit to probit here, since it is easier to extend to multinomial non-response; see Appendix 3.B.

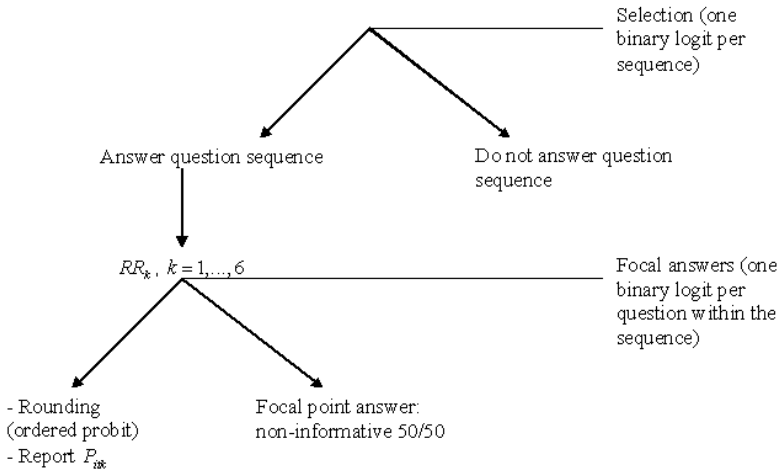


Figure 3.2: Structure of the model for response behavior

of 5%. Focal 50/50s can be inferred from the frequency of 50%-answers relative to that of 0 and 100%: The fraction of observed 50/50s that cannot be explained by rounding to multiples of 50%, 25%, 10%, etc. is attributed to non-informative answers. See also Kleinjans and Van Soest (2013) for a more extensive discussion of identification.

To complete the model specification, we need distributional assumptions on the individual effects α_i^h $(h = \mu, \sigma, R, F, S)$ and the sequence effects ζ_{it}^q $(q = \mu, \sigma, R, F)$. First, sequence effects are assumed to be independent of individual effects and of all idiosyncratic shocks. Second, we assume α_i^h (ζ_{it}^q) follow a five- (four-) dimensional normal distribution with mean zero and covariance matrix Σ_α (Σ_ζ). To ensure positive-definiteness of these covariance matrices, we decompose them into Cholesky factors Λ_r $(r = \alpha, \zeta)$ and estimate the elements of Λ_r along with the other parameters of the model ($\Sigma_r = \Lambda_r \Lambda_r'$, with Λ_r a positive semi-definite lower diagonal matrix).

We specify individual and sequence effects as random rather than fixed, because our aim is to describe variation in survey response and expectations without necessarily estimating causal effects. The random effects-assumption is suitable for that purpose. Moreover, Table 3.2 shows that many covariates are stable over time, so fixed effects estimation would not be feasible. In principle

the model could be extended with quasi-fixed effects, including the individual means of selected time-varying variables as additional covariates.

Estimation proceeds by Maximum Simulated Likelihood (see, e.g., Train 2003). We take draws of α_i^h and ζ_{it}^q using Halton sequences and calculate the likelihood contribution of each respondent conditional on those values of the individual and sequence effects. These conditional likelihood contributions are combinations of univariate normal and logistic probabilities, which are easy to calculate. Then we average across the distributions of α_i^h and ζ_{it}^q to obtain simulated likelihood contributions, take the log and sum over all individuals in the sample. We present more details on the likelihood contributions Appendix 3.A.

Appendix 3.B describes an alternative model specification that interprets logically inconsistent answer sequences as completely uninformative with respect to expectations and rounding and corrects for this selection step, in the spirit of Van Santen et al. (2012). We do not present this model in the main text because we found that its added complexity is not justified by a better overall fit: the alternative model does not capture the average subjective probability in the sample as well, and also is not able to reproduce the sample fractions of 0 and 100 probabilities (cf. Table 3.6 in the text and Table 3.14 in the appendix). We do, however, report details on this model in Appendix 3.B, to investigate whether or not answering behavior differs fundamentally between monotonic and non-monotonic subsamples.

Results 3.5

This section presents estimation results. First, we investigate model fit for models with and without covariates and present our main results on the prevalence of rounding and focal answers. Second, the importance of both types of unobserved heterogeneity is assessed, again comparing models with and without covariates. Third, we report results from a full model including covariates, and describe how answering behavior and expectations vary with socio-demographic variables. Finally, we compare with estimates from linear random effects models.

3.5.1 Model fit

The top panel of Table 3.6 reports summary statistics of the reported probabilities in the estimation sample, providing the benchmark against which to evaluate model predictions. For instance, the fraction of item non-response in the data is 7.4 percent and the average reported probability, pooled across all thresholds, is 43.6 percent. The remaining panels present corresponding statistics simulated using various estimated specifications of the model. The second and third panel use estimates from constant-only models that either allow (panel 2) or do not allow (panel 3) for the possibility of focal answers. The model used in the fourth panel also omits focal answers, but includes socio-economic covariates in all equations. Notice that only one of our estimated specifications allows for non-informative reported probabilities. This is due to the numerical difficulty of estimating the model with both a sufficient number of simulation draws and the possibility of focal answers: the estimate of the intercept of the focal equation tends to minus infinity and the estimated variance to positive infinity. Hence we report estimates from models with and without focal answers in the remainder of this section, and illustrate that focal answers are not an important concern for the data at hand.

For each specification, Table 3.6 presents both reported and latent probabilities. The latter are not affected by rounding, focal answers or item non-response (but censoring at 0 and 100 is incorporated). We simulate the average as well as the proportion equal to 0, 25, 50, 75 and 100 percent. Because the latent probabilities are censored by zero and one hundred but not rounded, the probability of them being exactly equal to any specific value other than 0 or 100 is zero. This is why the table does not present simulated probabilities at 25, 50, 75, multiples of 1, etc. for the latent probabilities. Model fit is assessed by comparing the simulated quantities to their observed counterparts.

Looking first at the overall averages shown in the leftmost column, we find that the estimated models all reproduce the sample mean rather well. All specifications reproduce the fraction of item non-response up to less than a single percentage point. Furthermore, the estimates imply averages of the latent indices and reported probabilities that are close to the average reported probability in the data (the differences are less than 1.5 percentage points). Note that the average of the latent index is very similar to that of the simulated

Table 3.6: Model fit: observed vs. simulated samples

	Simulated observed/latent probabilities								
		Percentage equal to					Percentage multiples of		
	Mean	P = 0	P = 25	P = 50	P = 75	P = 100	1	1; 5	1; 5; 10
				i. Sample statistics					
Selection	7.4								
Probabilities	43.6	20.9	1.8	8.8	1.8	15.6	4.3	8.4	38.5
				ii. Simulations: constant only; focal answers[a]					
Selection	7.7								
Probabilities	43.3	24.8	1.7	12.0	1.3	15.1	3.7	7.0	34.3
Latent probs.	43.5	19.5	–	–	–	11.1	–	–	–
				iii. Simulations: constant only; no focal answers[b]					
Selection	7.9								
Probabilities	42.7	26.2	1.7	10.9	1.4	15.0	3.5	6.8	34.5
Latent probs.	42.9	19.6	–	–	–	10.5	–	–	–
				iv. Simulations: covariates; no focal answers[c]					
Selection	7.5								
Probabilities	44.9	23.5	1.7	10.9	1.3	16.8	4.1	7.3	34.4
Latent probs.	44.3	17.9	–	–	–	13.0	–	–	–

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
Fraction of non-response, average probabilities and proportions equal to 0, 25, 50, 75 and 100 all reported in percentages.
Simulated latent probabilities P^*_{itk} are censored to lie between 0 and 100, but are not rounded.
Simulations are based on 1,000 samples

reported probability, so rounding and focal answers do not affect the predicted average probability. Not allowing for focal answers leads to slightly lower averages that are further from the sample statistics, but adding covariates removes that discrepancy.

Censoring at zero and one hundred reproduces the proportions equal to 0 and 100 rather well (18 – 20 percent compared to 20.9 percent at zero and 11 – 13 percent compared to 15.6 percent at one hundred). This indicates that a large share of the observed bunching at zero and one hundred is due to censoring. Comparing the second and third panels shows that allowing for focal answers hardly changes the simulated incidence of 50/50s: the model produces 10.9 percent fifties if we do not allow for focal answers and 12.0 if we do, both exceeding the 8.8 percent observed in the data. Adding covariates does not bring the proportion of 50/50s closer to that in the data.

The right panel of Table 3.6 presents the observed and simulated sample proportions of probabilities falling into the various rounding categories. For example, 4.3 percent of the probabilities in the data can only be a result of rounding to a multiple of 1, while 8.4 percent is rounded to a multiple of either 1 or 5. We find that the models reproduce these statistics quite closely, especially once we include covariates. On the other hand, the incidence of answers in the third category (multiples of 1, 5 and 10), is under-predicted by about 4 percentage points.

The evaluation of model fit can be formalized using the tests proposed by Andrews (1988). We report the results of such tests in Appendix 3.C (Table 3.17). The tests support the conclusion that the model successfully reproduces many features of the data. However, the model-implied proportions of zeros and multiples of 1, 5 and 10 deviate enough from the data to reject the model for broad partitions of the dependent variable.

In order to illustrate the model-fit graphically, Figure 3.3 presents histograms of "reported" probabilities observed in the data (left panel) and from 1,000 simulated samples using the model with covariates but without focal answers (right panel). The model successfully reproduces the peaked nature of the data with bunching at multiples of 10.

The model-implied incidence of rounding and focal answers is described in Table 3.7. As expected from the histograms we find that crude rounding is quite important: only about six or seven percent of reported probabilities are

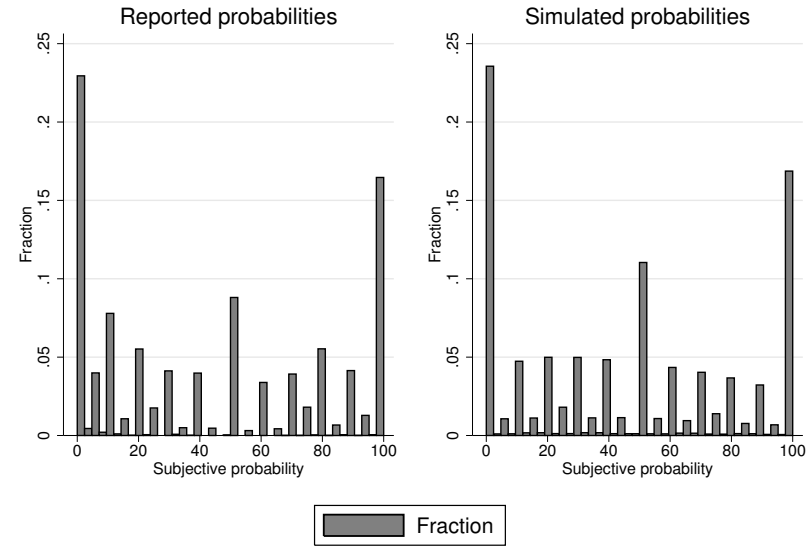


Figure 3.3: Histograms of reported and simulated probabilities, all thresholds pooled

rounded to a multiple of 1. The single most important category of rounding is to a multiple of 10: close to 50 percent of informative responses fall in this category. Only 3.5 percent rounds to a multiple of 25. The estimated probability of non-response is around 7.5 percent, which accords with the data (compare with Table 3.6). Moreover, focal 50/50s are not very prevalent – the simulated probability is only 1.3 percent. This suggests that the vast majority of 50 percent answers are informative rounded answers. When we do not allow for focal 50/50s the rounding probabilities remain similar, with a slightly larger tendency to round to multiples of 50.

As can be seen in the right panel of Table 3.7, including covariates in all equations does not affect the simulated proportions much. Rounding to multiples of 1 and 5 is estimated to be somewhat more prevalent and rounding to multiples of 50 somewhat less likely in the full specification, but the differences are only 2 – 3 percentage points.

Taken together, these results suggest that rounding and focal answers hardly affect the mean reported probability, that censoring explains most of the bunching at 0 and 100, but that rounding is needed to explain the bunching at intermediate values. Moreover, focal 50/50s do not play a large role. The latter result is in line with Kleinjans and Van Soest (2013) who find that focal 50/50s

Table 3.7: Simulated probabilities for rounding, non-response and focal answers

	Simulated probabilities		
	No covariates		Covariates
	Focal answers[a]	No focal answers[b]	No focal answers[c]
Rounding: multiples of			
1	6.4	5.8	7.3
5	23.6	22.4	25.0
10	49.3	48.9	48.2
25	3.6	3.8	3.1
50	17.1	19.2	16.4
Non-response			
Prob. non-response	7.7	7.9	7.5
Focal 50/50			
Prob. focal 50/50	1.3	0.0	0.0

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
Simulated probabilities reported in percent.
Rounding probabilities are conditional on giving an informative answer.
Probability of giving a focal answer is conditional on providing an answer.
Simulations based on 1,000 samples.

are rare for three of the four subjective probability questions they analyze. The one question for which they do find a substantial role for focal 50/50s has a large fraction of observed 50% answers, much larger than in our data (see Figure 3.3).

Simulations based on the model described in Appendix 3.B show that these results on the prevalence of rounding and focal answers remain unchanged if we interpret logically inconsistent answers as a form of non-response (see Table 3.15). Allowing for the possibility of non-informative answers complicates estimation severely and leads to convergence problems, so we will focus on models that do not allow for the possibility of focal answers in order to interpret the correlation patterns of unobserved heterogeneity described in the next subsection.

Unobserved heterogeneity 3.5.2

Table 3.8 presents the estimated variances of the individual effects for the three specifications, as well as the fractions of unsystematic variance in each equation captured by the individual effects. We find significant unobserved heterogeneity at the level of the individual in all equations. In the item response equation, individual effects account for only 8 – 24 percent of the unsystematic variance. Responding with a non-informative 50/50, on the other hand, is driven almost exclusively by unobserved heterogeneity (99% of error variance comes from the individual effects). Combined with the finding that the incidence of focal answers is only 1.3 percent, this indicates that there is a small group of respondents persistently giving focal answers. Individual effects are also significant in the rounding equation accounting for 54 – 59 percent of the unsystematic variance. Similarly, for the two parameters of the subjective distributions we find significant individual-level unobserved heterogeneity accounting for the majority of unsystematic variance (49 – 67 percent for μ and 84 – 92 percent for $\ln[\sigma]$).

Table 3.9 shows the estimated correlations among the individual effects. Individuals who are prone to non-response are also more likely to use a crude level of rounding but are less likely to give focal point answers. This finding supports the idea that focal answers are different from crude rounding. Interestingly, unobserved heterogeneity in the parameters of the subjective distributions

Table 3.8: Estimated variances of individual effects

	Variance of individual effects					
	No covariates				Covariates	
	Estimate[a]	% error var.	Estimate[b]	% error var.	Estimate[c]	% error var.
Selection	0.265	0.0746*	1.040***	0.240***	0.539***	0.141***
	(0.171)	(0.0445)	(0.310)	(0.0545)	(0.197)	(0.0441)
Focal 50/50s	540.5***	0.990***	–	–	–	–
	(194.3)	(0.00313)				
Rounding	1.235***	0.542***	1.525***	0.586***	1.398***	0.573***
	(0.0689)	(0.0137)	(0.0925)	(0.015)	(0.0928)	(0.0162)
Mu	0.0530***	0.669***	0.0500***	0.493***	0.0408***	0.555***
	(0.00193)	(0.0118)	(0.00254)	(0.0123)	(0.00154)	(0.0122)
Sigma	0.477***	0.922***	0.811***	0.835***	0.481***	0.903***
	(0.0436)	(0.0241)	(0.0567)	(0.0298)	(0.0386)	(0.0231)

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

is related to that in the equations of the answer process. Respondents who tend to give focal answers expect a higher replacement rate and are less certain. Rounding, on the other hand, is associated with a higher expected replacement rate and more certainty. The other estimates of the correlations are sensitive to the specification of the model and to the number of Halton draws used when simulating the likelihood. We also find mixed evidence for a relationship between the individual effects of the location and dispersion measures of the subjective distributions, though the magnitude of the correlation is small for all specifications.

Table 3.10 presents estimates of the variance of the sequence effects. We find no evidence for such effects in the equation for focal answers. Though their variance is statistically significant, sequence effects also play a minor role compared to individual effects in the rounding process (accounting for 2 – 3 percent of the unsystematic variance). For the two parameters of the subjective distributions sequence effects are highly significant and quantitatively more important: they account for 33 – 51% of error variance for μ and 8 – 17% for σ.

The estimated correlations among the sequence effects are shown in Table 3.11. They are always significant and generally larger in size than the correlations among the individual effects in Table 3.9. Their signs sometimes differ

Table 3.9: Correlations among individual effects

	Correlation matrices of the individual effects				
	Selection	Focal	Rounding	Mu	Sigma
	i. Constant only; focal answers[a]				
Selection	1				
Focal	-0.826***	1			
Rounding	0.443***	-0.340***	1		
Mu	-0.533***	0.494***	-0.0368	1	
Sigma	-0.763***	0.828***	-0.455***	-0.00311	1
	ii. Constant only; no focal answers[b]				
Selection	1				
Focal	–	–			
Rounding	0.302***	–	1		
Mu	-0.139***	–	0.110***	1	
Sigma	-0.814***	–	-0.262***	-0.158***	1
	iii. Covariates; no focal answers[c]				
Selection	1	–			
Focal	–	–			
Rounding	-0.0366	–	1		
Mu	0.984***	–	0.101***	1	
Sigma	0.204***	–	-0.113***	0.0775*	1

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 3.10: Estimated variances of sequence effects

	Variance of sequence effects					
	No covariates				Covariates	
	Estimate[a]	% error var.	Estimate[b]	% error var.	Estimate[c]	% error var.
Focal 50/50s	2.048	0.00375	–	–	–	–
	(2.403)	(0.00358)				
Rounding	0.0444***	0.0195***	0.0769***	0.0295***	0.0421***	0.0172***
	(0.0114)	(0.00486)	(0.0164)	(0.00615)	(0.0116)	(0.00465)
Mu	0.0262***	0.331***	0.0514***	0.507***	0.0327***	0.445***
	(0.00122)	(0.0118)	(0.00177)	(0.0123)	(0.00122)	(0.0122)
Sigma	0.0404***	0.0781***	0.161***	0.165***	0.0517***	0.0969***
	(0.0132)	(0.0241)	(0.0316)	(0.0298)	(0.0129)	(0.0231)

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

from those in Table 3.9. For example, while individuals with a tendency to round more tend to have higher medians and less uncertainty, more rounding in a given sequence is associated with lower median and more uncertainty in that sequence – the opposite pattern. Also, the mood effects of mu and sigma are strongly negatively correlated. Note, however, that the importance of these findings is limited by the small magnitudes of the (variances of the) sequence effects (Table 3.10).

Table 3.11: Correlations among sequence effects

	Correlation matrices of the sequence effects			
	Focal	Rounding	Mu	Sigma
	i. Constant only; focal answers[a]			
Focal	1			
Rounding	-0.992***	1		
Mu	0.602***	-0.641***	1	
Sigma	-0.908***	0.892***	-0.813***	1
	ii. Constant only; no focal answers[b]			
Focal	–			
Rounding	–	1		
Mu	–	-0.737***	1	
Sigma	–	0.838***	-0.985***	1
	iii. Covariates; no focal answers[c]			
Focal	–			
Rounding	–	1		
Mu	–	-0.882***	1	
Sigma	–	0.311***	-0.720***	1

[a] Model estimated using 50 Halton draws.
[b] Model estimated using 500 Halton draws.
[c] Model estimated using 100 Halton draws.
*significant at 10%; **significant at 5%; ***significant at 1%

In summary, both reporting behavior and expectations show considerable persistence at the level of the individual. Sequence (or mood) effects are often statistically significant but are quantitatively not very important except for the median of the subjective distribution. The answering process and the subjective uncertainty appear to be quite persistent over time. Importantly, aspects of the answering process are correlated with beliefs, suggesting that a model that does not account for the answering process may yield biased results.

Covariates 3.5.3

Table 3.12 presents the correlations of response behavior and expectations with the covariates described in Table 3.2. Non-response hardly varies with respondent characteristics, in line with the small differences in the means for the complete sample and for respondents only in Table 3.2. Males are somewhat more likely to answer the probability questions than otherwise similar females. The wave dummies indicate significant variation in non-response across time, with non-response being more prevalent in the waves of 2007 and 2010.

With regard to rounding, we find that most coefficients have the expected sign. The middle education group, however, rounds significantly more than the lowest category. Single males round more than single females. Having a partner is associated with more precise answers for men, but not for women. Again we observe significant differences across waves. Rounding was strongest in 2008, which was the year of the financial crisis. This suggests that a sudden increase in uncertainty in the economic environment may lead to changes in data quality due to cruder rounding.

The third column shows how the variance of the reporting errors varies with the covariates (heteroscedasticity). As was mentioned in section 3.4, we introduce heteroscedasticity to account for the observation that erratic response behavior, such as logically inconsistent probability sequences, is more prevalent in some socio-economic groups, such as those with lower education. We find that the size of reporting errors varies significantly with background characteristics in an intuitively plausible way. For example, higher education is associated with less noisy answers. The estimated coefficients on the age terms imply that reporting error increases with age.

The fourth column reports estimates of the equation explaining μ_{it}, the log of the median of the subjective distribution (equation 3.4 in section 3.4). Respondents who expect they can retire before age 65 expect a lower replacement rate than those who do not expect to be able to retire early. The difference of about 6 percent is rather small, however (much less than what would be actuarially neutral). The age pattern is highly non-linear, with a strong decline up to age 50 followed by a marked increase. This non-linearity can be interpreted as reflecting expected medium-term adjustments of the pension system to accommodate the aging population. Higher income groups expect

Table 3.12: Estimates from joint model of survey response and expectations

	Selection	Rounding	Heterosc.	Mu	Sigma
Expected ret. age 50-60				-0.0594***	0.206***
				(0.0104)	(0.0656)
Expected ret. age 61-64				-0.00381	-0.0476
				(0.00609)	(0.0458)
Expected ret. age 66-70				0.0182*	0.0563
				(0.00982)	(0.0610)
HH. income €1800-2600	-0.379*	-0.0195	-0.0221	-0.0706***	-0.272***
	(0.223)	(0.0722)	(0.0268)	(0.0109)	(0.0605)
HH. income >€2600	-0.0475	-0.172**	0.0174	-0.0746***	-0.342***
	(0.230)	(0.0836)	(0.0291)	(0.0113)	(0.0693)
Prob. pp. increase				-0.00545***	0.00192
				(0.00112)	(0.00760)
Prob. pp. decrease				-0.0146***	0.0217***
				(0.000956)	(0.00641)
Part-time pension				0.0356***	-0.0432
				(0.00646)	(0.0398)
Age	0.0330	-0.0214	-0.00569	-0.0150***	0.146***
	(0.0725)	(0.0256)	(0.00879)	(0.00317)	(0.0230)
Age sqrd/100	-0.0551	0.0419	0.0266***	0.0149***	-0.209***
	(0.0816)	(0.0288)	(0.00993)	(0.00350)	(0.0267)
Education middle	-0.236	0.320***	-0.0484**	-0.0581***	-0.00720
	(0.208)	(0.0913)	(0.0243)	(0.0113)	(0.0722)
Education high	-0.274	0.143*	-0.106***	-0.117***	-0.254***
	(0.209)	(0.0857)	(0.0238)	(0.0101)	(0.0720)
Male	-0.679**	0.308***	-0.0352	-0.0153	-0.0711
	(0.303)	(0.116)	(0.0342)	(0.0138)	(0.0848)
HH. head	0.169	-0.00956	0.138***	-0.0138	0.120*
	(0.255)	(0.101)	(0.0325)	(0.0106)	(0.0708)
Number of children				0.00342	-0.0902***
				(0.00296)	(0.0250)
Partner	0.0361	0.230*	0.103**	0.0256*	0.203**
	(0.318)	(0.130)	(0.0410)	(0.0151)	(0.0931)
Partner*male	-0.121	-0.476***	-0.0949**	-0.0757***	-0.114
	(0.384)	(0.153)	(0.0455)	(0.0169)	(0.106)
Homeowner				0.0430***	0.156***
				(0.00764)	(0.0528)
Wave 2007	1.072***	0.0204	0.207***	-0.00633	0.0120
	(0.224)	(0.0554)	(0.0281)	(0.00991)	(0.0554)
Wave 2008	-0.0315	0.163***	0.0515	-0.0214**	-0.171***
	(0.289)	(0.0624)	(0.0321)	(0.00968)	(0.0584)
Wave 2009	0.136	-0.159***	0.00927	-0.0370***	0.0175
	(0.279)	(0.0603)	(0.0303)	(0.00961)	(0.0569)
Wave 2010	0.552**	-0.107*	-0.0817**	-0.0142	-0.120**
	(0.264)	(0.0637)	(0.0323)	(0.0103)	(0.0604)
Constant	-2.955*		2.842***	4.964***	-3.368***
	(1.597)		(0.191)	(0.0770)	(0.503)
Number of obs.			2,931		
Number of respondents			1,246		
Log likelihood			-40,588.96		

Model estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 3.12: Estimates from joint model of survey response and expectations (continued)

	Mu	Sigma
Sector: construction	0.0670***	-0.0677
	(0.0192)	(0.118)
Sector: trade/transport	0.00671	0.0130
	(0.0170)	(0.0828)
Sector: financial services	0.0213	-0.0722
	(0.0153)	(0.0823)
Sector: education	-0.0920***	-0.245***
	(0.015)	(0.0831)
Sector: healthcare	0.00777	-0.337***
	(0.0150)	(0.0766)
Sector: governance	0.0163	-0.244***
	(0.0144)	(0.0804)
Sector: other	-0.0277	-0.470***
	(0.0177)	(0.105)
Number of obs.	2,931	
Number of respondents	1,246	
Log likelihood	-40,588.96	

Model estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

lower replacement rates, perhaps due to the public pension which does not vary with income. This may also explain why the higher educated expect lower replacement rates. Our results on age and education are in line with those of Van Santen et al. (2012), but we find significant income effects where they do not; we return to this below. Having the possibility of part-time retirement is associated with a higher expected replacement rate.

The final column of Table 3.12 presents estimates of the sigma-equation, which drives the standard deviation of the log-replacement rate (equation 3.5 in section 3.4).[8] We find that retirement before the age of 60 is associated with significantly higher replacement rate uncertainty than retirement at age 65. The highest income group reports lower uncertainty on average. The age pattern is in line with intuition: uncertainty decreases non-linearly with age from age 35 onward. Uncertainty is lower among the highly educated than in the lowest education category. Van Santen et al. (2012) find the opposite pattern: more highly educated respondents are more uncertain. This difference could be due to differences between the covariates used (they include experience while we do not, while we include sector of employment), or to differences between the models (they regard logically inconsistent answers as a form of sample selection while we interpret them as noisy). Furthermore, as can be seen in the second part of Table 3.12, subjective uncertainty is lower in the education, health care and governance sectors compared to the industrial sector.

For the model as a whole, the McFadden R-squared is 0.025, indicating that the improvement in the log-likelihood by adding the covariates to the constant-only specification is rather small. Unobserved heterogeneity is much more important for the fit of the model: the McFadden R-squared comparing constant-only models with and without unobserved heterogeneity is 0.134.

[8]The dispersion of the *level* of the expected replacement rate depends on both the mu and sigma parameters, so that the coefficients in the sigma-equation do not correspond directly to effects on replacement rate uncertainty (and the size of coefficients cannot be compared with those in Van Santen et al. (2012)). However, the sign and significance of these coefficients are informative of the correlation between respondent characteristics and uncertainty keeping the scale of the distribution constant. We also calculated the average marginal effects of covariates on replacement rate uncertainty, taking into account simultaneous changes in mu and sigma. The signs of these effects are always in accordance with Table 3.12, though some standard errors were much larger. (Details are available upon request.)

Comparison with linear RE models 3.5.4

To address the question whether incorporating response behavior in the model affects the conclusions concerning variation in subjective distributions, we compare the results in Table 3.12 with those of linear Random Effects (RE) models. The latter are presented in Table 3.13. They are obtained by first fitting observation-specific log-normal distributions and then regressing the parameters of these distributions on the same covariates that were used in the joint model. We find different results, both in terms of statistical significance and magnitude. For instance, in the equation explaining the location parameter mu, linear RE models suggest that the expected replacement rate does not vary with income, similar to the results of Van Santen et al. (2012), while Table 3.12 shows that the highest income group expects a significantly lower replacement rate. Also, the magnitude of the coefficient on high education in Table 3.13 is only about half the size of the estimate for the full model and the dummy on the middle education category becomes insignificant. Furthermore, the linear RE model fails to uncover any systematic variation in the expected replacement rate across sectors of employment.

Smaller and less significant coefficients in the linear RE model than in the full model that incorporates response behavior are also found for the equation for the uncertainty parameter sigma. According to the linear RE model age and employment in the health care sector are the only socio-economic variables that are related to uncertainty in a significant way, while the full model also gives significant relations of uncertainty with the expected retirement age, income and education. Hence, compared to linear RE models, incorporating response behavior does affect our conclusions regarding expectations. These insights can be formalized through Hausman tests of the null-hypothesis that the estimators applied in the two models have same probability limit. We find that this null-hypothesis is soundly rejected for both mu and sigma ($\chi^2(29) = 609$ for the mu-equation and 608 for the sigma-equation, against a critical value of 49.6 at $\alpha = 0.01$).

Table 3.13: Linear RE models of subjective distributions

	Mu	Sigma
Expected ret. age 50-60	-0.0607***	0.162
	(0.0187)	(0.134)
Expected ret. age 61-64	-0.0175*	0.0543
	(0.0104)	(0.0949)
Expected ret. age 66-70	0.00765	0.0833
	(0.0183)	(0.234)
HH. income €1800-2600	-0.0160	-0.126
	(0.0151)	(0.154)
HH. income >€2600	-0.0172	-0.138
	(0.0170)	(0.169)
Prob. pp. increase	-0.00222	-0.0129
	(0.00179)	(0.0197)
Prob. pp. decrease	-0.0106***	0.00379
	(0.00151)	(0.0166)
Part-time pension	0.00307	-0.126
	(0.00954)	(0.0957)
Age	-0.0186***	0.185***
	(0.00578)	(0.0542)
Age sqrd/100	0.0209***	-0.253***
	(0.00645)	(0.0628)
Education middle	-0.0190	-0.208
	(0.0155)	(0.139)
Education high	-0.0694***	-0.125
	(0.0166)	(0.139)
Male	0.0431*	-0.141
	(0.0241)	(0.203)
HH. head	-0.00172	-0.154
	(0.0188)	(0.186)
Number of children	0.00550	-0.0917*
	(0.00555)	(0.0523)
Partner	0.0190	-0.186
	(0.0249)	(0.278)
Partner*male	-0.0153	0.350
	(0.0295)	(0.274)
Homeowner	-0.00543	0.130
	(0.0147)	(0.133)
Robust Hausman test	602.9	601.3
Number of obs.	2,686	
Number of respondents	1,143	

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 3.13: Linear RE models of subjective distributions (continued)

	Mu	Sigma
Wave 2007	-0.0224*	-0.0112
	(0.0128)	(0.0597)
Wave 2008	0.00338	-0.0274
	(0.0131)	(0.0676)
Wave 2009	0.00891	0.00430
	(0.0133)	(0.0693)
Wave 2010	0.0188	-1.690***
	(0.0148)	(0.239)
Sector: construction	-0.000527	0.132
	(0.0285)	(0.211)
Sector: trade/transport	0.00874	0.00822
	(0.0198)	(0.176)
Sector: financial services	0.0137	0.109
	(0.0189)	(0.151)
Sector: education	-0.0203	-0.244
	(0.0213)	(0.198)
Sector: healthcare	-0.00188	-0.392**
	(0.0205)	(0.193)
Sector: governance	-0.0185	-0.0811
	(0.0224)	(0.181)
Sector: other	-0.0103	-0.453
	(0.0279)	(0.293)
Constant	4.814***	-4.155***
	(0.127)	(1.168)
Number of obs.	2,686	
Number of respondents	1,143	

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

3.6 Conclusion

As potential determinants of well-being and behavior, subjective expectations continue to attract the attention of applied economists. Over the past two decades, the consensus has emerged that measurement of such expectations should proceed by questions phrased in terms of the probability that some event will materialize (see Manski 2004, for an overview). For continuous outcome variables, this implies that several probabilities are elicited to characterize the subjective cumulative distribution function.

Though this elicitation method has advantages over other methods, the process by which survey respondents answer such questions, with random reporting errors, rounding, focal answers, or selective non-response, is not well understood. In this paper we present an econometric model that captures the process by which respondents answer probabilistic questions and simultaneously explains the location and dispersion of the subjective distributions of a continuous outcome from socio-economic covariates. We exploit the panel data features of our data and model unobserved heterogeneity on two levels, namely that of the individual respondent and of the respondent/survey wave-combination.

We apply the model to expectations of the retirement income replacement rate for a representative sample of Dutch employees. Our results indicate that non-informative focal answers do not occur often: the simulated probability of a focal answer is about 1 percent. However, reported probabilities are affected by substantial rounding: close to half of the reported probabilities are rounded to a multiple of 10 percent. These conclusions with regard to rounding and focal answers are similar whether we interpret logically impossible responses as noisy or as completely uninformative.

Our model allows for unobserved heterogeneity at the level of the respondent and at the level of the question sequence in a given wave. The magnitude of the individual effects suggests that respondents stick to a certain answering strategy across survey waves: rounding, non-response and focal answers all are persistent at the level of the survey respondent. Moreover, respondents who tend not to answer the questions also tend to round more crudely when they do answer. In addition, we find small but significant transitory shocks to rounding at the level of the survey wave which are strongly correlated with

subjective expectations: respondents who round roughly in a certain sequence of probabilities have a lower median replacement rate and more uncertainty compared to those who round less.

Both the location and dispersion of the replacement rate distributions also contain quantitatively important permanent components, with individual effects accounting for 50 – 60 and 84 – 92 percent of the unsystematic variance respectively. Moreover, while the expected replacement rate and uncertainty are not strongly correlated at the level of the individual, they are strongly negatively correlated at the level of the question sequence (correlation coefficient 0.7 – 0.98). Respondents who receive a negative transitory shock to their expected replacement rate also tend to report high uncertainty in the same wave.

Comparing the full model with linear RE models, we find that the size and significance of estimated relationships between expectations and covariates are larger for the full model. This applies both to the equation explaining the location of the subjective distributions and that explaining subjective uncertainty. Moreover, formal tests confirm that these differences are statistically significant. Hence, the analysis suggests that even if one does not aim to understand the process of answering probabilistic questions, it is worthwhile to take it into account in order to draw conclusions about variation in expectations across socio-economic groups.

Acknowledgements 3.7

We would like to thank Luc Bissonnette, an anonymous referee and participants at the 2011 Netspar workshop in Amsterdam and at the 2011 ESEM conference in Oslo for helpful comments on an earlier draft of the paper.

3.A Likelihood Contributions

First define the vectors of individual effects $\bar{\alpha}_i = \left(\alpha_i^{\mu}, \alpha_i^{\sigma}, \alpha_i^{R}, \alpha_i^{F}, \alpha_i^{S}\right)'$ and sequence effects $\bar{\zeta}_{it} = \left(\zeta_{it}^{\mu}, \zeta_{it}^{\sigma}, \zeta_{it}^{R}, \zeta_{it}^{F}\right)'$. We calculate the conditional likelihood contribution of individual i, conditional on both types of unobserved effects and covariates:

$$L_i^c = L_i^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right) = \prod_{t=1}^{T_i} L_{it}^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right)$$

T_i denotes the number of waves during which an individual responded to the questionnaire, the panel is unbalanced so this number varies across respondents. If the respondent does not answer the sequence of probability questions in wave t, L_{it}^c equals the probability of non-response at the sequence-level:

$$L_{it}^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right) = \Pr\left(y_{it}^S = 1 | x_{it}, \alpha_i^S\right) \text{ if } y_{it}^S = 1$$

If the respondent does answer the questions, we use independence of choices conditional on individual and mood effects and express the conditional likelihood contribution of observation i, t as:

$$L_{it}^c = \Pr\left(y_{it}^S = 0 | x_{it}, \alpha_i^S\right) \prod_{k=1}^{6} L_{itk}^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right)$$

with the following question-specific probabilities:[9]

$$\begin{aligned} L_{itk}^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right) &= \Pr\left(y_{itk}^F = 0 | x_{it}, \alpha_i^F, \zeta_{it}^F\right) \times \Pr\left(R_{itk} = 1 | x_{it}, \alpha_i^R, \zeta_{it}^R\right) \\ &\quad \times \Pr\left(P_{itk} - 0.5 < P_{itk}^* \leq P_{itk} + 0.5 | x_{it}, \alpha_i^P, \zeta_{it}^P\right) \\ &\quad \text{if } P_{itk} \in \{1, ..., 100\} \text{ and } P_{itk} \notin \{0, 5, 10, ..., 95, 100\} \end{aligned}$$

$$\begin{aligned} L_{itk}^c\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right) &= \Pr\left(y_{itk}^F = 0 | x_{it}, \alpha_i^F, \zeta_{it}^F\right) \\ &\quad \times \left[\Pr\left(R_{itk} = 1 | x_{it}, \alpha_i^R, \zeta_{it}^R\right)\right. \\ &\quad \times \Pr\left(P_{itk} - 0.5 < P_{itk}^* \leq P_{itk} + 0.5 | x_{it}, \alpha_i^P, \zeta_{it}^P\right) \end{aligned}$$

[9] In order to save space we substitute $\alpha_i^P = \left(\alpha_i^{\mu}, \alpha_i^{\sigma}\right)'$ and $\zeta_{it}^P = \left(\zeta_{it}^{\mu}, \zeta_{it}^{\sigma}\right)'$.

$$
\begin{aligned}
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-2.5<P_{itk}^*\leq P_{itk}+2.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&\text{if } P_{itk}\in\{5,15,35,45,55,65,85,95\}
\end{aligned}
$$

$$
\begin{aligned}
L_{itk}^c\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right)=&\Pr\left(y_{itk}^F=0|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-0.5<P_{itk}^*\leq P_{itk}+0.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-2.5<P_{itk}^*\leq P_{itk}+2.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=3|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-5<P_{itk}^*\leq P_{itk}+5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&\text{if } P_{itk}\in\{10,20,30,40,60,70,80,90\}
\end{aligned}
$$

$$
\begin{aligned}
L_{itk}^c\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right)=&\Pr\left(y_{itk}^F=0|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-0.5<P_{itk}^*\leq P_{itk}+0.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-2.5<P_{itk}^*\leq P_{itk}+2.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=4|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(P_{itk}-12.5<P_{itk}^*\leq P_{itk}+12.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&\text{if } P_{itk}\in\{25,75\}
\end{aligned}
$$

$$
\begin{aligned}
L_{itk}^c\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right)=&\Pr\left(y_{itk}^F=0|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*\leq 0.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*\leq 2.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=3|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*\leq 5|x_{it},\alpha_i^P,\zeta_{it}^P\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\Pr\left(R_{itk}=4|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*\leq 12.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=5|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*\leq 25|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&\text{if } P_{itk}=0
\end{aligned}
$$

$$
\begin{aligned}
L_{itk}^c\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right)=&\Pr\left(y_{itk}^F=0|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*>99.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*>97.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=3|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*>95|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=4|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*>87.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=5|x_{it},\alpha_i^R,\zeta_{it}^R\right)\times\Pr\left(P_{itk}^*>75|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&\text{if } P_{itk}=100
\end{aligned}
$$

$$
\begin{aligned}
L_{itk}^c\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right)=&\Pr\left(y_{itk}^F=0|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(49.5<P_{itk}^*\leq 50.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=2|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(47.5<P_{itk}^*\leq 52.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=3|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(45<P_{itk}^*\leq 55|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=4|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(37.5<P_{itk}^*\leq 62.5|x_{it},\alpha_i^P,\zeta_{it}^P\right)\\
&+\Pr\left(R_{itk}=5|x_{it},\alpha_i^R,\zeta_{it}^R\right)\\
&\times\Pr\left(25<P_{itk}^*\leq 75|x_{it},\alpha_i^P,\zeta_{it}^P\right)\Big]\\
&+\Pr\left(y_{itk}^F=1|x_{it},\alpha_i^F,\zeta_{it}^F\right)\\
&\text{if } P_{itk}=50
\end{aligned}
$$

The unconditional likelihood contribution is the expected value of L_{it}^c over the distributions of $\bar{\alpha}_i$ and $\bar{\zeta}_{it}$. We have that $\bar{\alpha}_i=\Lambda_\alpha u_\alpha$ and $\bar{\zeta}_{it}=\Lambda_\zeta u_\zeta$, where

u_α (u_ζ) is a vector of 5 (4) independent random variables that follow the standard normal distribution and Λ_r ($r = \alpha, \zeta$) are the Cholesky factors of the covariance matrices of the individual and sequence effects (these covariance matrices are called Σ_r in the body of the text). Using this notation we obtain the following unconditional likelihood contribution for individual i:

$$L_i = \int_{\mathbb{R}^5} \int_{\mathbb{R}^4} L_i^c \left(x_{it}; \Lambda_\alpha u_\alpha, \Lambda_\zeta u_\zeta \right) \varphi \left(u_{\zeta_1} \right) ... \varphi \left(u_{\zeta_4} \right) \varphi \left(u_{\alpha_1} \right) ... \varphi \left(u_{\alpha_5} \right) du$$

Here $\varphi(.)$ is the density of the standard normal distribution. This integral is simulated by taking draws of u_α and u_ζ, generating the individual and sequence effects, calculating L_i for each draw and taking the average:

$$SL_i = \frac{1}{R} \sum_{r=1}^{R} L_i^{cr} \left(x_{it}; \Lambda_\alpha u_\alpha^r, \Lambda_\zeta u_\zeta^r \right)$$

where $u_\alpha^1, ..., u_\alpha^R$ and $u_\zeta^1, ..., u_\zeta^R$ are vectors with components drawn from independent standard normal distributions. In order to reduce the variance of our estimates, we take Halton draws to generate u_α^r and u_ζ^r (rather than draws generated by a pseudorandom algorithm). If R tends to infinity at a fast enough rate, the maximum simulated likelihood estimator is asymptotically equivalent to exact maximum likelihood (Train 2003).

3.B Alternative model: non-monotonic sequences interpreted as non-informative

The model presented in the main text and described in detail in the Appendix 3.A assumes that all reported probabilities contain information about the answering process and underlying beliefs, regardless of whether probabilities are (weakly) decreasing within each sequence. In our sample, 22.2 percent of all sequences violate the monotonicity requirement. Our main model interprets such sequences as informative but noisy. An alternative interpretation is to regard them as completely uninformative with respect to rounding and subjective expectations. Van Santen et al. (2012) model logically inconsistent responses as a process of sample selection and collapse them with non-response into a single category of non-informative answers. They then estimate Heckman selection models for the parameters of the individual-specific subjective distributions. In order to check the robustness of our estimates, we follow a similar approach within the modeling framework presented in the main text, including non-monotonic responses as an additional form of sample selection at the level of the response sequence. Also, we adjust the model to take into account that remaining probability sequences are restricted to be monotonic. In initial estimation rounds the intercept of the equation for focal answers tended to negative infinity when the sample was restricted to logically consistent sequences, indicating that the share of focal answers among these responses is close to zero. Hence, we estimate this model without the possibility of non-informative focal answers. In the remainder of this appendix we first explain the likelihood of this alternative model specification. Then we present a brief discussion of the results, comparing them with the model presented in the main text.

3.B.1 The likelihood

Viewing non-monotonic sequences as the non-informative result of a selection process, rather than noisier yet potentially useful data on the sample distribution of subjective expectations, complicates the likelihood function in two ways. Firstly, sample selection is no longer a binary decision between either reporting a full sequence of probabilities and not reporting anything. Instead it becomes a multinomial decision that may result in a monotonic sequence,

a non-monotonic sequence or no response at all. We model this decision as a Random Effects (RE) multinomial logit, with effects corresponding to unobserved heterogeneity at the level of the respondent (a direct extension of the RE binary logit that was used to model sample selection in the main model):

$$
\begin{aligned}
&u_{it}^{j} = x_{it}'\beta_{Sj} + \alpha_i^{Sj} + \varepsilon_{it}^{Sj},\ j = 1,2,3 \\
&y_{it}^{S} = j \text{ if } u_{it}^{j} \geq u_{it}^{k},\ k \neq j \\
&\beta_{S1} = 0;\ \alpha_i^{S1} = 0 \\
&\varepsilon_{it}^{Sj} \sim iid \text{ GEV(I) independent of } \alpha_i^{S2},\ \alpha_i^{S3} \text{ and } x_{it}
\end{aligned}
$$

We can summarize the consequences of this addition for the likelihood function as follows:

$$
L_{it}^{c}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) = \begin{cases} \Pr\left(y_{it}^{S} = 1|x_{it},\alpha_i^{S1}\right) \times \prod_{k=1}^{6} L_{itk}^{c}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) & \text{if } y_{it}^{S} = 1 \\ \Pr\left(y_{it}^{S} = 2|x_{it},\alpha_i^{S2}\right) & \text{if } y_{it}^{S} = 2 \\ \Pr\left(y_{it}^{S} = 3|x_{it},\alpha_i^{S2}\right) & \text{if } y_{it}^{S} = 3 \end{cases}
$$

$$
\begin{aligned}
&y_{it}^{S} = 1 \text{ if a monotonic sequence of probabilities is reported:} \\
&P_{itk} \leq \min\left[P_{itk-1},\ldots,P_{it1}\right]\ \forall k \in \{2,\ldots,6\} \\
&y_{it}^{S} = 2 \text{ if no probabilities are reported: } P_{itk} = . \\
&\quad \text{for } k = 1,\ldots,6 \\
&y_{it}^{S} = 3 \text{ if reported probabilities are non-monotonic:} \\
&\quad \exists K \in \{2,\ldots,6\} : P_{itk} > \min\left[P_{itk-1},\ldots,P_{it1}\right]
\end{aligned}
$$

Note that the vector of individual effects now contains an additional effect for the equation corresponding to reporting a non-monotonic sequence:

$$
\bar{\alpha}_i = \left(\alpha_i^{\mu},\alpha_i^{\sigma},\alpha_i^{R},\alpha_i^{F},\alpha_i^{S2},\alpha_i^{S3}\right)'
$$

The second complication arises from limiting the set of probabilities to be analyzed to monotonic sequences. We incorporate this restriction into the likelihood function as additional censoring, on top of the censoring that occurs at 0 and 100. Due to the sequential nature of the answering process, the upper limit for the reported probabilities is reduced from 100 for P_{it1} to P_{itk-1} for all $k \in \{2,\ldots,6\}$ (the lower limit remains zero for all probabilities). Since this

censoring applies to reported rather than latent probabilities, we have to take into account that this monotonicity requirement applies after rounding takes place. Hence, we consider reported probabilities to be subject to censoring if:

$$P_{itk} = 0 \tag{i}$$

$$0 < P_{itk} < 100 \text{ and } \begin{cases} P_{itk-1} - P_{itk} < 1 \ \& \ R_{itk} = 1 \\ P_{itk-1} - P_{itk} < 5 \ \& \ R_{itk} = 2 \\ P_{itk-1} - P_{itk} < 10 \ \& \ R_{itk} = 3 \\ P_{itk-1} - P_{itk} < 25 \ \& \ R_{itk} = 4 \\ P_{itk-1} - P_{itk} < 50 \ \& \ R_{itk} = 5 \end{cases} \tag{ii}$$

$$P_{itk} = 100 \tag{iii}$$

(R_{itk} reflects the degree of rounding)

This censoring scheme implies that an individual who rounds to a multiple of R, with $R \in \{1, 5, 10, 25, 50\}$, reports probability P_{itk} $(k = 1, 2, ..., 6)$, such that:

$$P_{itk} = n \times R, \text{ with } n \text{ an integer}$$
$$\text{and } 0 \leq P_{itk} \leq P_{itk-1} \leq 100$$

The additional censoring required to guarantee monotonicity of reported probability sequences complicates the question-specific probabilities presented in Appendix 3.A. The cases discussed in that appendix remain valid and correspond to the contributions in absence of censoring except for probabilities equal to zero or 100. In order to aid intuition, we give some examples of the additional censoring introduced by the monotonicity restriction:[10] Rounding to multiples of 1:

$$\begin{aligned} L^c_{itk}\left(x_{it}; \bar{\alpha}_i, \bar{\zeta}_{it}\right) &= \Pr\left(y^F_{itk} = 0 | x_{it}, \alpha^F_i, \zeta^F_{it}\right) \times \Pr\left(R_{itk} = 1 | x_{it}, \alpha^R_i, \zeta^R_{it}\right) \\ &\quad \times \Pr\left(P_{itk} - 0.5 < P^*_{itk} \leq P_{itk} + 0.5 | x_{it}, \alpha^P_i, \zeta^P_{it}\right) \\ &\text{if } P_{itk} \in \{1, ..., 100\}\,;\ P_{itk} \notin \{0, 5, 10, ..., 95, 100\} \end{aligned}$$

[10] As in Appendix 3.A, we use the shorthand notation $\alpha^P_i = \left(\alpha^\mu_i, \alpha^\sigma_i\right)'$ and $\zeta^P_i = \left(\zeta^\mu_i, \zeta^\sigma_i\right)'$. Also, the possibility of focal answers is ignored, because the intercept of that equation converged to minus infinity when estimated from the monotonic responses alone. Hence we concluded that focal answers are unimportant in the monotonic answers and estimated the model under the restriction $\Pr\left(y^F_{it} = 0 | x_{it}, \alpha^F_i, \zeta^F_{it}\right) = 1$.

$$\text{and } P_{itk-1} - P_{itk} \geq 1$$

$$\begin{aligned} L^c_{itk}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) &= \Pr\left(y^F_{itk}=0|x_{it},\alpha^F_i,\zeta^F_{it}\right)\times\Pr\left(R_{itk}=1|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P^*_{itk}>P_{itk}-0.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\\ &\text{if } P_{itk}\in\{1,...,100\}\,;\, P_{itk}\notin\{0,5,10,...,95,100\}\\ &\text{and } P_{itk-1}-P_{itk}<1 \end{aligned}$$

Rounding to multiples of 1 or 5:

$$\begin{aligned} L^c_{itk}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) &= \Pr\left(y^F_{itk}=0|x_{it},\alpha^F_i,\zeta^F_{it}\right)\\ &\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P_{itk}-0.5<P^*_{itk}\leq P_{itk}+0.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\\ &+\Pr\left(R_{itk}=2|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P_{itk}-2.5<P^*_{itk}\leq P_{itk}+2.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\Big]\\ &\text{if } P_{itk}\in\{5,15,35,45,55,65,85,95\}\ \text{ and } P_{itk-1}-P_{itk}\geq 5 \end{aligned}$$

$$\begin{aligned} L^c_{itk}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) &= \Pr\left(y^F_{itk}=0|x_{it},\alpha^F_i,\zeta^F_{it}\right)\\ &\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P_{itk}-0.5<P^*_{itk}\leq P_{itk}+0.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\\ &+\Pr\left(R_{itk}=2|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P^*_{itk}>P_{itk}-2.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\Big]\\ &\text{if } P_{itk}\in\{5,15,35,45,55,65,85,95\}\ \text{ and } 1\leq P_{itk-1}-P_{itk}<5 \end{aligned}$$

$$\begin{aligned} L^c_{itk}\left(x_{it};\bar{\alpha}_i,\bar{\zeta}_{it}\right) &= \Pr\left(y^F_{itk}=0|x_{it},\alpha^F_i,\zeta^F_{it}\right)\\ &\times\Big[\Pr\left(R_{itk}=1|x_{it},\alpha^R_i,\zeta^R_{it}\right)\\ &\times\Pr\left(P^*_{itk}>P_{itk}-0.5|x_{it},\alpha^P_i,\zeta^P_{it}\right)\\ &+\Pr\left(R_{itk}=2|x_{it},\alpha^R_i,\zeta^R_{it}\right) \end{aligned}$$

$$\times \Pr\left(P^*_{itk} > P_{itk} - 2.5 | x_{it}, \alpha^P_i, \zeta^P_{it}\right)\Big]$$
$$\text{if } P_{itk} \in \{5, 15, 35, 45, 55, 65, 85, 95\} \text{ and } P_{itk-1} - P_{itk} < 1$$

Due to our distributional assumptions all probabilities featuring in the conditional likelihood contribution are either univariate normal or logistic and therefore easy to calculate. As described in the Appendix 3.A, we simulate the unconditional likelihood contribution by taking draws of the individual and sequence effects, calculating the likelihood contribution conditional on those draws and averaging the resulting conditional contributions.

3.B.2 Estimation results

In order to assess model fit, we compare observed sample statistics and their simulated counterparts for versions of this model with and without covariates in Table 3.14. We restrict the sample to monotonic sequences when calculating the summary statistics in Table 3.14, since those are the only ones to be used in the estimation of the expectations and rounding parts of the model. Overall, the average reported probability is under-predicted by 4.2 percentage points, while the original model reproduced the average reported probability up to a single percentage point. This is largely due to underestimating the fraction of 100s and overestimating the proportion of 0 percent answers by about 5 percentage points (the model with covariates over-predicts the proportion of zeros by 8.4 percentage points). Also, the model that interprets non-monotonic sequences as uninformative over-predicts the fraction of 50s by 4.6 – 6 percentage points. In these dimensions it performs slightly worse than the baseline specification described in the text. Hence, our specification of a single censoring mechanism that applies at zero, one hundred and at intermediate values may be inappropriate.

Table 3.15 compares the simulated probabilities of rounding, non-response and non-monotonic probabilities for the alternative model (model B) and the original model (model A, reproduced from Table 7). Rounding patterns appear to be qualitatively similar in the subsample of monotonic responses according to model B compared to the complete sample of probabilities in model A. Rounding to multiples of 10 is by far the most prevalent form of rounding in both samples, close to half of the respondents fall in that category, followed

Table 3.14: Model fit for model B: observed vs. simulated samples

	Simulated observed/latent probabilities								
		Percentage equal to					Percentage multiples of		
	Mean	P = 0	P = 25	P = 50	P = 75	P = 100	1	1; 5	1; 5; 10
				i. Sample statistics					
Selection	7.4								
Inconsistent	22.2								
Probabilities	47.3	19.1	1.7	9.1	1.9	18.3	4.2	8.2	37.6
				ii. Simulations: constant-only[a]					
Selection	7.3								
Inconsistent	21.6								
Probabilities	46.0	23.8	1.3	15.0	1.7	14.3	3.4	6.1	34.3
Latent probs.	50.4	12.9	–	–	–	12.8	–	–	–
				iii. Simulations: specification with covariates[b]					
Selection	7.8								
Inconsistent	22.0								
Probabilities	43.1	27.5	1.3	13.7	1.6	13.7	3.3	6.0	32.9
Latent probs.	48.2	15.3	–	–	–	13.7	–	–	–

[a] Model estimated using 500 Halton draws.
[b] Model estimated using 100 Halton draws.
Fraction of non-response, average probabilities and proportions equal to 0, 25, 50, 75 and 100 all reported in percentages.
Simulated P^*_{itk} censored between 0 and 100.
Simulations are based on 1,000 samples.

by another quarter that round to multiples of 5. Towards the extremes of the rounding scale, Model B suggests that answers are more precise in the subsample of monotonically decreasing probabilities: the simulated probability of rounding to multiples of 1 is about 3 percentage points higher and the probability of rounding to multiples of 50 is 3 to 4 percentage points lower compared to the full sample of probabilities. Logically consistent sequences tend to consist of more precisely reported probabilities, though there is still substantial rounding.

Table 3.15: Simulated probabilities from models A and B

	Simulated probabilities			
	Model A		Model B	
	No covariates[a]	Covariates[b]	No covariates[a]	Covariates[b]
Rounding: multiples of				
1	5.8	7.3	8.7	10.1
5	22.4	25.0	24.3	26.5
10	48.9	48.2	46.5	46.7
25	3.8	3.1	4.3	4.2
50	19.2	16.4	16.2	12.5
Non-response				
Prob. non-response	7.9	7.5	7.3	7.8
Non-monotonic sequence				
Prob. non-monotonic sequence	–	–	21.6	22.0
Focal 50/50				
Prob. focal 50/50	0.0	0.0	0.0	0.0

[a] Model estimated using 500 Halton draws.
[b] Model estimated using 100 Halton draws.
Simulated probabilities are reported as percentage.
Rounding probabilities are conditional on giving an informative answer.
Simulations are based on 1,000 samples.

Analogously to the results in the main text, we find quantitatively important unobserved heterogeneity at the level of the individual in all parts of the model.[11] For response behavior, sample selection and rounding, we find that the magnitudes of the effects are very close to those reported in the main text. Moreover, the tendency to give logically inconsistent responses is persistent, with individual effects accounting for 33 percent of the error term. Expectations

[11] Estimates available on request.

are more persistent in the logically consistent subsample: individual effects make up 65 percent of error variance in the μ-equation and 99 percent in the σ-equation (compared to 56 and 90 percent respectively in the preferred model).

Estimation results for both models are presented in Table 3.16. The estimates of the rounding and selection equations are mostly similar to those from the baseline model, but there are some exceptions. For example, ignoring the logically inconsistent sequences (model B) no longer gives the result that single men round more than single women (model A). Logically inconsistent answers are disproportionally given by lowly educated respondents, as is evident from the negative and significant coefficients on the education dummies in the equation for sequence consistency.

The second page of Table 3.16 presents the equations for μ and σ as well as the heteroskedasticity equation of both models. It shows that limiting the estimation sample to respondents who report internally consistent probabilities also changes some of the conclusions with regard to the determinants of the location and dispersion measures of the subjective expectations (8 coefficients on socio-economic covariates change sign). For example, home ownership enters the μ-equation with a significantly positive coefficient in model A yet turns up significantly negative in model B. The comparison with linear RE models estimated on the same subsample gives the same result as for the model discussed in the main text: the estimates are significantly different for both μ and σ.

Table 3.16: Models of subjective expectations and response behavior

	Model A[a]		Model B[a]		
	Selection	Rounding	Selection	Consistent	Rounding
HH. income €1800-2600	-0.379*	-0.0195	-0.421*	-0.0499	0.103
	(0.223)	(0.0722)	(0.227)	(0.142)	(0.0724)
HH. income >€2600	-0.0475	-0.172**	-0.151	-0.000995	0.0843
	(0.230)	(0.0836)	(0.235)	(0.151)	(0.0790)
Age	0.0330	-0.0214	0.0143	-0.0743	0.0183
	(0.0725)	(0.0256)	(0.0730)	(0.0467)	(0.0244)
Age sqrd/100	-0.0551	0.0419	-0.0300	0.0819	0.00448
	(0.0816)	(0.0288)	(0.0822)	(0.0517)	(0.0275)
Education middle	-0.236	0.320***	-0.385*	-0.231*	0.192**
	(0.208)	(0.0913)	(0.212)	(0.127)	(0.0809)
Education high	-0.274	0.143*	-0.457**	-0.875***	0.113
	(0.209)	(0.0857)	(0.210)	(0.131)	(0.0763)
Male	-0.679**	0.308***	-0.563*	-0.0532	0.0106
	(0.303)	(0.116)	(0.306)	(0.203)	(0.0985)
HH. head	0.169	-0.00956	0.238	0.286*	-0.171**
	(0.255)	(0.101)	(0.258)	(0.172)	(0.0868)
Partner	0.0361	0.230*	0.110	0.313	-0.107
	(0.318)	(0.130)	(0.323)	(0.224)	(0.112)
Partner*male	-0.121	-0.476***	-0.148	-0.260	0.0444
	(0.384)	(0.153)	(0.390)	(0.258)	(0.126)
Wave 2007	1.072***	0.0204	1.179***	0.366***	-0.0182
	(0.224)	(0.0554)	(0.226)	(0.13)	(0.0610)
Wave 2008	-0.0315	0.163***	0.0741	0.124	0.0780
	(0.289)	(0.0624)	(0.291)	(0.154)	(0.0662)
Wave 2009	0.136	-0.159***	0.208	0.0976	-0.143**
	(0.279)	(0.0603)	(0.280)	(0.155)	(0.0654)
Wave 2010	0.552**	-0.107*	0.657**	0.204	-0.225***
	(0.264)	(0.0637)	(0.266)	(0.159)	(0.0691)
Constant	-2.955*		-2.360	0.346	
	(1.597)		(1.610)	(1.032)	
Number of obs.	2,931			2,931	
Number of respondents	1,246			1,246	
Log likelihood	-40,588.96			-26,348.49	

[a] Model estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 3.16: Models of subjective expectations and response behavior (continued)

	Model A[a]			Model B[a]		
	Heterosc.	Mu	Sigma	Heterosc.	Mu	Sigma
Expected ret. age 50-60		-0.0594***	0.206***		-0.118***	0.153***
		(0.0104)	(0.0656)		(0.00860)	(0.0567)
Expected ret. age 61-64		-0.00381	-0.0476		-0.0304***	-0.0609
		(0.00609)	(0.0458)		(0.00548)	(0.0392)
Expected ret. age 66-70		0.0182*	0.0563		0.0322***	0.0201
		(0.00982)	(0.0610)		(0.0104)	(0.0594)
HH. income €1800-2600	-0.0221	-0.0706***	-0.272***	-0.0270	-0.0231***	-0.0995*
	(0.0268)	(0.0109)	(0.0605)	(0.0393)	(0.00675)	(0.0526)
HH. income >€2600	0.0174	-0.0746***	-0.342***	-0.134***	-0.0351***	-0.104*
	(0.0291)	(0.0113)	(0.0693)	(0.0406)	(0.00799)	(0.0563)
Prob. pp. increase		-0.00545***	0.00192		-0.00147*	0.00736
		(0.00112)	(0.00760)		(0.000777)	(0.00655)
Prob. pp. decrease		-0.0146***	0.0217***		-0.0102***	0.0135**
		(0.000956)	(0.00641)		(0.000777)	(0.00557)
Part-time pension		0.0356***	-0.0432		0.0525***	-0.0852**
		(0.00646)	(0.0398)		(0.00459)	(0.0336)
Age	-0.00569	-0.0150***	0.146***	-0.0272**	-0.0269***	0.163***
	(0.00879)	(0.00317)	(0.0230)	(0.0125)	(0.00231)	(0.0228)
Age sqrd/100	0.0266***	0.0149***	-0.209***	0.0488***	0.0291***	-0.225***
	(0.00993)	(0.00350)	(0.0267)	(0.0142)	(0.00263)	(0.0265)
Education middle	-0.0484**	-0.0581***	-0.00720	-0.134***	-0.0955***	-0.0156
	(0.0243)	(0.0113)	(0.0722)	(0.0382)	(0.00815)	(0.0611)
Education high	-0.106***	-0.117***	-0.254***	-0.140***	-0.0765***	-0.116*
	(0.0238)	(0.0101)	(0.0720)	(0.0367)	(0.00800)	(0.0595)
Male	-0.0352	-0.0153	-0.0711	-0.213***	0.0352***	-0.0940
	(0.0342)	(0.0138)	(0.0848)	(0.0482)	(0.00800)	(0.0796)
HH. head	0.138***	-0.0138	0.120*	0.0503	-0.0194**	0.0299
	(0.0325)	(0.0106)	(0.0708)	(0.0431)	(0.00815)	(0.0620)
Number of children		0.00342	-0.0902***		0.0113***	-0.0327
		(0.00296)	(0.0250)		(0.00258)	(0.0207)
Partner	0.103**	0.0256*	0.203**	0.123**	-0.00123	0.0763
	(0.0410)	(0.0151)	(0.0931)	(0.0578)	(0.0111)	(0.0905)
Partner*male	-0.0949**	-0.0757***	-0.114	0.154**	-0.0457***	0.149
	(0.0455)	(0.0169)	(0.106)	(0.0645)	(0.0121)	(0.0960)
Homeowner		0.0430***	0.156***		-0.0380***	0.0120
		(0.00764)	(0.0528)		(0.00542)	(0.0456)
Wave 2007	0.207***	-0.00633	0.0120	0.0105	0.0170**	-0.212***
	(0.0281)	(0.00991)	(0.0554)	(0.0479)	(0.00729)	(0.0426)
Wave 2008	0.0515	-0.0214**	-0.171***	0.160***	0.0285***	-0.0875*
	(0.0321)	(0.00968)	(0.0584)	(0.0467)	(0.00788)	(0.0496)
Wave 2009	0.00927	-0.0370***	0.0175	0.0966**	0.0139**	0.00419
	(0.0303)	(0.00961)	(0.0569)	(0.0448)	(0.00661)	(0.0473)
Wave 2010	-0.0817**	-0.0142	-0.120**	0.315***	-0.0124	0.191***
	(0.0323)	(0.0103)	(0.0604)	(0.0443)	(0.00996)	(0.0570)
Constant	2.842***	4.964***	-3.368***	3.089***	5.075***	-4.122***
	(0.191)	(0.0770)	(0.503)	(0.276)	(0.0495)	(0.468)
Number of observations		2,931			2,931	
Number of respondents		1,246			1,246	
Log likelihood		-40,588.96			-26,348.49	

[a] Models estimated using 100 Halton draws.
Standard errors in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%

Table 3.16: Models of subjective expectations and response behavior (continued)

	Model A[a]		Model B[a]	
	Mu	Sigma	Mu	Sigma
Sector: construction	0.0670***	-0.0677	-0.0532***	0.0705
	(0.0192)	(0.118)	(0.0156)	(0.104)
Sector: trade/transport	0.00671	0.0130	0.00585	-0.0775
	(0.0170)	(0.0828)	(0.0100)	(0.0711)
Sector: financial services	0.0213	-0.0722	0.0124	-0.267***
	(0.0153)	(0.0823)	(0.00848)	(0.0643)
Sector: education	-0.0920***	-0.245***	-0.0236***	-0.00490
	(0.0150)	(0.0831)	(0.00831)	(0.0699)
Sector: healthcare	0.00777	-0.337***	-0.0156*	-0.239***
	(0.0150)	(0.0766)	(0.00825)	(0.0668)
Sector: governance	0.0163	-0.244***	0.00397	-0.312***
	(0.0144)	(0.0804)	(0.00948)	(0.0672)
Sector: other	-0.0277	-0.470***	-0.0353***	-0.270***
	(0.0177)	(0.105)	(0.0105)	(0.0870)
Number of observations	2,931		2,931	
Number of respondents	1,246		1,246	
Log likelihood	-40,588.96		-26,348.49	

[a] Models estimated using 100 Halton draws.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Chi-squared goodness of fit tests 3.C

In the text we asses model fit by means of a comparison of simulated data with the actual sample. Another way of doing so is through goodness-of-fit tests along the lines proposed by Andrews (1988). Such tests divide the outcome variable in partitions and compare the empirical and model-implied frequencies of responses in each partition.[12] Table 3.17 reports test statistics and critical values. Our outcome variable is the reported probability, which is either an integer between 0 and 100 or missing. In addition to the natural division of non-response versus response, partition a) in Table 3.17, we divide the set of reported probabilities in two ways. Firstly, partition b) consists of 10 brackets equal to $(0,10),(11,20),\ldots(91,100)$. Secondly, in partition c) we divide reported probabilities in groups that can only result from particular combinations of rounding (as is done in the construction of the likelihood, see Appendix 3.A). In order to apply the tests with these partitions of the response, we need to reduce our panel to a cross-section that includes all individuals used in estimation. This is achieved by including only the first observation for each respondent in our data. If such first observation is a full response we observe six probabilities for that individual, one of which is randomly selected so as to cover the distribution of reported probabilities across all thresholds.[13] Note that our choice for the first observation for each individual implies that the relative frequencies of non-response and other categories need not be equal to those reported in Table 3.6 based on the full dataset. Especially for non-response we find that the proportions in the full sample and the test-sample differ markedly (7.4% for the full sample against 9.4% for the test sample). Such differences bias our tests towards rejecting the model.

As can be seen in the top panel of Table 3.17, partition a) implies that our model is rejected based on the simple division between response and non-response. This is due to the higher fraction of non-response in the test sample compared to the full sample and illustrates the conservative nature of our tests. Moreover, partitions b) and c) also lead to rejection of the model,

[12] Another possibility is to define partitions across sub-samples defined by covariates as well as the outcome variable. However, we do not pursue this possibility here.

[13] The intention is to provide test statistics for equality of the simulated and observed fractions in Table 3.6, which are pooled across thresholds. Moreover, Andrews (1988) notes that random cells are allowed in his testing framework.

though including covariates improves model fit according to these statistics. The lower panels report tests from binary partitions taken from b) and c), in order to get a more detailed idea of where the model fails to reproduce the data. Also, given the non-representative fraction of non-response in the test sample, more detailed tests for particular groups of observations are fairer to the model. We find that for detailed decompositions of the observed response, the model is only rejected for 2 out of 10 cells for partition b) and 2 out of 7 cells for c). In the latter case the rejections formalize the notions from Table 3.6 that the model tends to produce too many zeros and too few answers that are multiples of 1, 5 and 10, but not 25 and 50. Hence, we conclude that although the general tests reject the model, this is at least partly due to the selection of our testing sample. Conditional on observing a response, the model fits the data well, as is evident from Table 3.17.

Table 3.18 presents similar model-fit tests for the specification described in Appendix 3.B. As noted in the previous paragraph, due to relatively high rate of non-response in our test sample we reject the part of the model that explains sample selection. Non-monotonic responses, on the other hand, appear to be modeled adequately, since the test sample is representative of the full sample in this respect. More detailed tests, reported in the second and third panel of 3.18, confirm that the model fits the data for 7 out of 10 partitions along the lines of b) and 4 out of 7 partitions according to c). Lack of model fit is caused by over prediction of the incidence of zeros and 50s, and to a lesser extent by under prediction of probabilities that are multiples only of 1, 5 and 10. All in all, model fit appears worse than for the model that interprets non-monotonic answers as resulting from reporting error.

Table 3.17: Goodness of fit: Chi-squared tests, model A

		Chi-squared tests, model A		
Partitions of Y	$J-G$	Critical value ($\alpha = 0.01$)	No covariates[a]	Covariates[b]
		i. Overall		
a)	1	6.6	10.9	25.8
b)	10	23.2	44.0	65.4
c)	7	18.5	41.0	49.7
		ii. Intervals of the reported probability		
b) - 0/10	1	6.6	0.9	2.3
b) - 11/20	1	6.6	1.5	0.6
b) - 21/30	1	6.6	0.06	3.6
b) - 31/40	1	6.6	9.0	5.1
b) - 41/50	1	6.6	10.6	9.7
b) - 51/60	1	6.6	4.1	8.1
b) - 61/70	1	6.6	0.003	2.0
b) - 71/80	1	6.6	8.3	7.0
b) - 81/90	1	6.6	2.7	2.2
b) - 91/100	1	6.6	0.004	1.7
		iii. Different categories of rounding		
c) - 0	1	6.6	18.9	17.3
c) - 100	1	6.6	1.7	0.001
c) - 50	1	6.6	7.0	4.8
c) - 25, 75	1	6.6	1.4	0.05
c) - k*1; 5; 10	1	6.6	12.6	9.8
c) - k*1; 5	1	6.6	0.57	0.3
c) - k*1	1	6.6	0.09	0.8

[a] Model estimated using 500 Halton draws.
[b] Model estimated using 100 Halton draws.
Partition a): non-response vs. response
Partition b): non-response vs. $[0,10],[11,20],\ldots,[81,90],[91,100]$
Partition c): non-reponse vs. $\{0\}$, $\{50\}$, $\{100\}$, $\{25,75\}$, $\{10,20,30,40,60,70,80,90\}$, $\{5,15,35,\ldots,85,95\}$, $\{1,2,3,\ldots,98,99\}$
Simulations are based on 1,000 samples

Table 3.18: Goodness of fit: Chi-squared tests, model B

	Chi-squared tests, model B			
Partitions of Y	$J-G$	Critical value ($\alpha = 0.01$)	No covariates[a]	Covariates[b]
		i. Overall		
a)	1	6.6	19.6	25.9
b)	1	6.6	0.8	1.4
c)	2	9.2	21.9	21.4
d)	11	24.7	93.9	389.6
e)	8	20.1	84.1	121.0
		ii. Intervals of the reported probability		
d) - 0/10	1	6.6	0.6	0.4
d) - 11/20	1	6.6	0.1	0.03
d) - 21/30	1	6.6	1.6	0.05
d) - 31/40	1	6.6	17.6	16.4
d) - 41/50	1	6.6	51.2	0.7
d) - 51/60	1	6.6	3.2	93.6
d) - 61/70	1	6.6	0.2	239.7
d) - 71/80	1	6.6	0.8	37.0
d) - 81/90	1	6.6	1.0e-04	2.4
d) - 91/100	1	6.6	9.6	84.2
		iii. Different categories of rounding		
e) - 0	1	6.6	14.9	1.9
e) - 100	1	6.6	2.9	30.9
e) - 50	1	6.6	50.3	5.9
e) - 25, 75	1	6.6	1.1	12.5
e) - k*1; 5; 10	1	6.6	5.3	57.9
e) - k*1; 5	1	6.6	0.5	3.8
e) - k*1	1	6.6	0.2	18.4

[a] Model estimated using 500 Halton draws.
[b] Model estimated using 100 Halton draws.
Partition a): non-response vs. response
Partition b): response vs. non-monotonic response
Partition c): response vs. non-response vs. non-monotonic response
Partition d): non-response vs. non-monoptonic response vs. $[0,10],[11,20],...,[81,90],[91,100]$
Partition e): non-reponse vs. non-monotonic response vs. $\{0\}$, $\{50\}$, $\{100\}$, $\{25,75\}$,$\{10,20,30,40,60,70,80,90\}$, $\{5,15,35,...,85,95\}$, $\{1,2,3,...,98,99\}$
Simulations are based on 1,000 samples.

4 Eliciting Subjective Survival Curves: Lessons from Partial Identification

This chapter is based on Bissonnette and De Bresser (2013).

4.1 Introduction

Human beings take into account the future consequences of their behavior. In economics this notion is formalized through models in which rational agents fully grasp the inter-temporal effects of their decisions, integrating the present with expectations of what is to come. For instance, when deciding how much to save for retirement one combines preferences over current and future consumption with ideas of how long this future might be: with beliefs about one's longevity (e.g. Hurd et al. 2004).

Without data on subjective expectations, economists impose identifying assumptions when taking theoretical models to the data. The dominant approach is to assume rational expectations: the subjective probability distributions that are expectations are equal to actual distributions observed in data. In the specific context of survival expectations rationality usually means that the probability to survive another year, conditional on having reached a certain age, is equal to that listed in actuarial tables (taking into account birth cohort and sex).

Though convenient, this rationality assumption is restrictive and may lead to misguided inference if individuals' subjective beliefs do not correspond to the official life tables. Therefore, researchers have started to analyze expectations directly by means of survey questions eliciting subjective beliefs (see

Manski 2004, for an overview). We propose a new method for the analysis of subjective expectations of a continuous outcome such as next year's wages or life expectancy. For continues variables, the data usually contain points on the cumulative distribution function over future events (in our application the events are survival past five age thresholds). Such data are often analyzed by assuming some underlying parametric distribution, fitting distributions to individual observations and then analyzing the estimated parametric distributions as measures of expectations (e.g. Dominitz and Manski 1997, and Dominitz 1998, on income expectations; Dominitz and Manski 2006, on pension benefit expectations and Perozek 2008, on mortality). We are interested in the extent to which analysis of this type of data requires parametric restrictions on beliefs. We analyze the importance of parametric assumptions by contrasting that approach to one that does not impose any restrictions on expectations. This novel method derives bounds on life expectancy that take into account that surveys only collect information on some points on the CDF, and that we do not have information on the CDF between those points. We show that these non-parametric bounds can easily be generalized to allow for rounding in the reported probabilities. Moreover, they can be combined with restrictions on beliefs or interpolation between data points to produce narrower admissible sets for subjective life expectancy.

Using a sample of Dutch adults, we show that the bounds are wide if we do not smooth expectations between data points. Under those minimal assumptions the bounds are 11 years wide on average and reduce the width of the uninformative interval, defined as the difference between the maximum age of 110 and a respondent's current age, by 77 percent. Without further assumptions, such wide bounds do not allow for useful inference: partially identified models do not corroborate the strong relationship between life expectancy and the covariates *age* and *self-reported health* found in point identified linear models. Allowing for rounding increases the width of the bounds, making them even less informative. However, we can narrow the bounds substantially by simultaneously interpolating beliefs between data points and allowing for rounding. Such interpolation using linear or cubic splines does not impose a parametric form, but does assume subjective survival functions are continuous and piecewise linear (linear splines) or continuous and smooth (cubic spline). Interpolation reduces the average width of the bounds

to 3 years if we allow for rounding but assume that each probability reported by a respondent is rounded to the same extent. The average width is 7 years if we allow each probability to be rounded differently, still a large improvement compared to bounds without smoothing or rounding. Under the common rounding scheme the bounds derived using interpolated survival functions are sufficiently tight to corroborate the correlations between life expectancy and age and health. A second approach to produce tighter bounds is to impose restrictions on survival functions without any smoothing. We show that assuming that the hazard of death, the probability of dying today conditional on surviving up to today, is continuous and weakly increases with age cuts the intervals for life expectancy in half. It does not, however, yield sufficiently narrow intervals to estimate informative intervals for the relationships between life expectancy and covariates.

Finally, we match our sets of bounds on life expectancy with cohort life tables produced by Statistics Netherlands. The extent to which expectations correspond to actuarial forecasts on average is relevant, because economists often use life tables as substitutes for subjective expectations thanks to the easy availability of actuarial figures (Peracchi and Perotti 2011). Our results suggest that some conclusions with regard to expectations disappear once we let go of point identification. If we point identify life expectancy, either through parametric restrictions on beliefs or through linear or cubic splines, we find that the expectations of men correspond closely to actuarial tables on average, but that women expect to live much shorter than the life tables suggest. The same discrepancy has been documented for American women by Perozek (2008). However, if we do not interpolate expectations, the bounds show that beliefs are on average consistent with actuarial forecasts. This conclusion emerges even more strongly when we allow for rounding and remains unchanged if we restrict expectations to exhibit a weakly increasing hazard of death. If we do smooth beliefs, we cannot reject consistency of women's expectations with life tables if we allow each individual probability to be rounded to the maximum extent. If we impose that all probabilities reported by a respondent follow a common rounding scheme, we restore the result that women on average expect to die younger than life tables suggest for limited age-ranges around age 30 and 60.

These findings offer some insights into the nature of subjective probabilities that may be relevant for a broader group of researchers than those interested in subjective survival. Firstly, point identification of expectations by parametric distributions yields almost identical results to approximation by (linear or cubic) splines, at least if a sufficient number of subjective probabilities are available. We observe five probabilities and the correlations between the calculated life expectancies are above 0.98. This result is in line with the findings of De Bresser and Van Soest (2013a), who document that their summary measures of expectations of one's retirement income are very similar for cubic spline approximation of beliefs compared to an imposed parametric form. Therefore, if we do not mind interpolating expectations between reported probabilities and if we are comfortable to ignore rounding, results are robust with respect to the specific form used. Secondly, in terms of limiting the informativeness of the data we find that the coarseness of the grid on which expectations are elicited is much more important than the extent of rounding in the probabilities. In particular, even though we have a relatively rich set of five probabilities that we use to approximate expectations, we cannot learn much about subjective life expectancies if we do not smooth expectations between those points. Unfortunately, that is the case even if we do not allow for any rounding and we could not save informativeness by imposing restrictions on expectations. Thirdly, if we do interpolate expectations, rounding does not necessarily make the data uninformative. However, the type of rounding matters: a common rounding scheme as proposed in Manski and Molinari (2010) leads to much more informative bounds on expectations than does a general "worst case"-scheme in which individual probabilities are rounded to the maximum extent. From that perspective it is reassuring that De Bresser and Van Soest (2013b) find that individuals tend to stick to the same rounding rule, suggesting that a single rounding rule may be an appropriate assumption.

The structure of the paper is as follows. The next section briefly summarizes the literature on subjective survival expectations and mentions the papers on partial identification on which we base our methodology. Section 4.3 introduces the type of data that we use and explains the parametric and non-parametric methods that we apply to approximate expectations. The data are described in section 4.4 and section 4.5 presents our results. Finally, section 4.6 concludes.

Details regarding the bounds on expectations under the increasing hazard restriction are given in Appendix 4.A.

Literature 4.2

Our paper ties in with two strands of research. The first literature concerns the elicitation of subjective survival questions, their correspondence to actuarial tables and their predictive power for subsequent mortality. Research on these issues started after the 1992 wave of the Health and Retirement Study (HRS) was released, which was the first large scale survey in which subjective longevity expectations were elicited by means of probabilistic questions. The HRS remains the most widely used source of subjective survival expectations even today. In 1992 it included two questions that asked for the likelihood of survival past target ages that depend on the current age of the respondent. However, answers were limited to multiples of 10 percent (see Juster and Suzman 1995, for more details on the HRS and Hurd and McGarry 1995, for a thorough description of the survival questions). Early research based on this data concluded that respondents were willing to answer questions on their beliefs about their own survival and that the average of the reported probabilities was close to the corresponding life table survival rate (Hurd and McGarry 1995). Moreover, the subjective probabilities were found to vary across subpopulations in the same way that actual mortality does, establishing predictive validity for actual survival (Hurd 2009). When the second wave of the HRS was released in 1994, longitudinal analysis became possible. Using the first two waves of the HRS, Hurd and McGarry (2002) found that individuals who reported higher probabilities of survival in 1992 were more likely to survive to 1994 and that changes in health that are known to increase the risk of dying lowered the probabilities reported in 1994.

Recent research based on the HRS data and observed mortality over a longer period suggests that men are on average optimistic and women pessimistic relative to actual survival (Bissonnette et al. 2011, Hurd 2009). Similar results were documented by Perozek (2008), who adds that subsequent revisions of actuarial forecasts of life expectancy were in line with discrepancies between subjective expectations and life tables in 1992. Predictive validity for mortality

at the individual level has been confirmed: subjective beliefs about longevity are consistent with individuals' observed survival patterns (Smith et al. 2001). Furthermore, the subjective probabilities of individuals who die between later survey waves show a consistent decline over time and respond negatively to new health shocks and increases in individuals' functional limitations (Smith et al. 2001). Two examples of Dutch datasets that include similar items are the 2011 wave of the *Pensioenbarometer*, collected from the CentERpanel and analyzed in this paper, and the health section of the LISSpanel.[1] Kutlu and Kalwij (2012) analyze longitudinal data from the CentERpanel, which we use as well, but they only observe two age thresholds per respondent/wave because they analyze data from a different survey. They fit parametric models of survival and compare the median expected remaining lifetime with within-sample objective survival. They find that subjective survival predicts actual mortality, even after controlling for risk factors such as smoking and being obese. Furthermore, subjective survival is correlated with covariates of actual survival, such as gender, age, smoking and self-rated health. However, education and income are not significant once health and risk factors are controlled for. Finally, Kutlu and Kalwij (2012) show that both men and women underestimate their life expectancy relative to actual life duration in the sample and that this underestimation is more severe for women.

The second literature that we draw on is that on partial identification. Manski and Molinari (2010) focus on rounding of subjective probabilities in the HRS. They use several reported probabilities to infer the type of rounding a respondent applies when answering probabilistic questions and re-interpret the exact probabilities as indicative only of intervals for the true subjective probabilities. We apply the Manski and Molinari (2010) common rounding scheme to our subjective survival questions and analyze its restrictiveness by contrasting it with a more conservative model that allows for differential rounding within the question sequence. Another related set of papers develop an econometric framework for the analysis of interval-level data (see Tamer 2010, for a summary of that literature). Imbens and Manski (2004) and Beresteanu and Molinari (2008) develop the theory of best linear predictors of interval-censored dependent variables. We apply their models to our non-parametric bounds on life expectancy to investigate differences across socioeconomic groups.

[1] See www.centerdata.nl for more details on both surveys.

Methods 4.3

Survival questions 4.3.1

The subjective longevity questions that we analyze are similar to those found in the HRS: they also ask for the probability of surviving past certain age thresholds. However, in contrast to the HRS we consider questions that refer to a maximum of five thresholds, depending on the age of the respondent at the time of the questionnaire. The items are phrased as follows:

> Please indicate on a scale from 0 to 100 how likely you think it is that you live to:
>
> 1. *[if age < 69]* age 70
> 2. *[if age < 74]* age 75
>
> ...
>
> 5. *[if age < 89]* age 90

Age-eligibility requires a respondent to be at least 2 years younger than a particular age threshold in order for that question to be presented. One example of possible answers for a respondent who is 68 or younger is given in Table 4.1. Depending on the estimation method and the rounding rules we will interpret these probabilities as perturbed parametric probabilities, as exact points on the subjective survival function, or as rounded approximations to the survival function (see section 4.3.5 for an explanation of the different forms of rounding that we distinguish).

Table 4.1: Hypothetical data

Event	Prob. (%)
Survive to age 70	90
Survive to age 75	70
Survive to age 80	45
Survive to age 85	30
Survive to age 90	15

The fact that we observe five probabilities for many respondents means that our data on expectations is relatively rich compared to the HRS, which only

contains two questions on subjective longevity. This facilitates the estimation of parametric models of survival and increases the informativeness of the bounds that we derive.

4.3.2 Parametric survival functions

One commonly used approach when analyzing subjective data of the type described in the previous section is to fit parametric, individual-specific survival functions by non-linear least squares. This methodology was introduced by Dominitz and Manski (1997) in the context of expectations about future income and has been applied to subjective survival probabilities by Perozek (2008). The latter takes two popular functional forms for survival functions, the Gompertz and Weibull distributions, and obtains parameters for each individual by solving:

$$\min_{\theta_i^1,\theta_i^2} \sum_{t \in T_i} \left[P_{it} - S\left(t, age_i; \theta_i^1, \theta_i^2\right) \right]^2 \tag{4.1}$$

Where P_{it} is the subjective probability of survival past age t reported by individual i; T_i is the (age-dependent) set of all age thresholds t for which the questions are asked (in the context of our data $T_i = \{70, 75, 80, 85, 90\}$ for respondents aged 68 or younger); $S(.)$ is the parametric form imposed on expectations and θ_i^1 and θ_i^2 are the parameters of $S(.)$ (these parameters are specific to individual i). Estimation of subjective survival curves by non-linear least squares assumes that the subjective probabilities of individuals are equal to the parametric probabilities $S(.)$, perturbed by mean zero IID errors.

In order to enforce that expectations are such that the probability of survival tends to zero for extremely old ages, we apply the technique introduced by Perozek (2008) and construct an additional probability for survival to age 98. This extra probability consists of the product of the subjective probability of survival past age 90 and the actuarial estimate of the likelihood of surviving to age 98, conditional on survival to age 90. For the hypothetical data from Table 4.1, this auxiliary probability would be equal to $P\left(90|age_i\right) \times P\left(98|90, age_i\right) =$

$15 \times P\left(98|90, age_i\right)$.[2] However, given that we observe five probabilities in the data, our estimates are not sensitive to this additional probability. After estimating the individual-specific parameters, we can easily calculate remaining life expectancy from the parametric survival curves.

Non-parametric survival functions 4.3.3

An alternative approach that also allows for the point identification of expectations is to trace the points on the subjective survival function that are given in the data, without assuming those probabilities stem from a parametric model. This method constructs a flexible (piecewise) function that passes exactly through the reported probabilities; it does not invoke implicit error terms as does the non-linear least squares method explained in the previous subsection. We tried approximating expectations with linear and cubic splines; the latter are explained by Bellemare et al. (2012). In the remainder of the paper we report only results from cubic splines, since Bellemare et al. (2012) show that those functions are best able to approximate a wide range of parametric models. Moreover, in the specific context of human mortality it seems unlikely that individuals have piecewise linear expectations, especially in the tails of the distribution.[3] However, cubic splines lead to complications when constructing intervals for life expectancy under the general rounding scheme described in section 4.3.5: the upper and/or lower bound cross the point estimate for 22% of the sample. Therefore, it is reassuring to find that the results from linear spline functions, for which the bounds never cross each other, are almost identical to those reported below. The solid line in the right panel of Figure 4.1 shows the fitted survival function using cubic splines for the example data given in Table 4.1.

The right panel of Figure 4.1 also illustrates that the probabilities reported in the data, shown as circles in the figure, do not suffice to characterize expectations completely if we follow the spline approach. Only if the respondent is certain to survive past the youngest age threshold for which the respondent

[2] We only construct the probability of living past age 90 for the part of the sample that is at least 50 years old, because Statistics Netherlands does not yet offer forecasts of survival probabilities to that age for younger cohorts.

[3] We expect the survival function to be concave for ages close to the current age of the respondent and convex for the oldest ages.

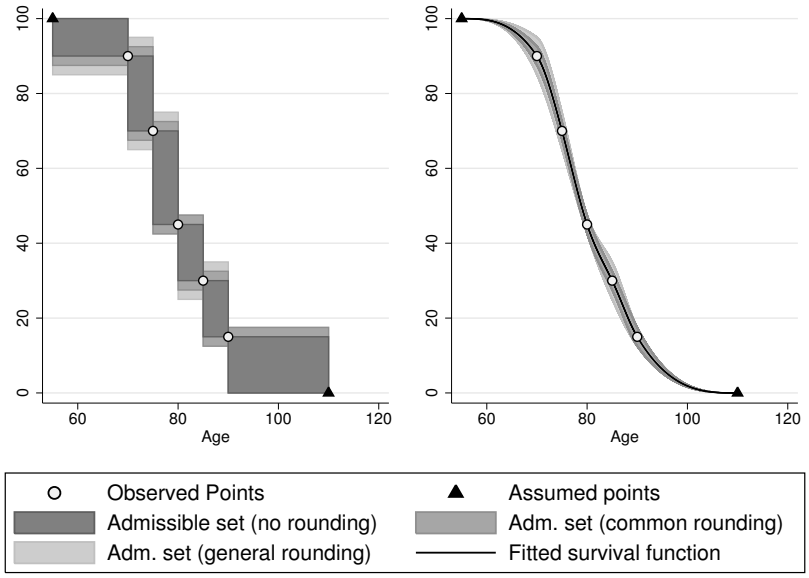


Figure 4.1: Admissible set for the survival curve with and without rounding (left panel) and spline interpolation approach (right panel)

is eligible, $P_{it_{i1}} = 100$, and certain to die before the age of 90, $P_{i90} = 0$, do the data tie down expectations. This is an important consideration since 84 percent of respondents indicate less than complete certainty about their survival past the the first age threshold that was presented to them, so that these respondents require an imposed lower bound on the age of death. An even larger fraction of 87 percent indicate that the probability of living past the age of 90 is more than zero. The auxiliary probabilities that we construct are given by the triangles in Figure 4.1. For the minimum age of death we take the current age of the respondent (age 55 for the example in Figure 4.1). It is less straightforward to come up with a credible maximum age of death. We experimented with the ages 120, 110 and 100; in Figure 4.1 and the remainder of this paper we impose 110 as the maximum attainable age. Note that the influence of this maximum lifespan on the calculated life expectancies depends on the level of the final reported probability for age 90: larger probabilities will result in more sensitive life expectancies. Moreover, cubic spline interpolation tends to be more robust than linear splines with respect to the age of certain death, since the interpolated functions are usually convex in the final interval (see Figure

4.1). Results using the alternative maximum ages of 120 and 100 are similar to those reported in the text and are available on request.

Non-parametric bounds on life expectancy 4.3.4

Approximation of beliefs by non-linear least squares or spline functions is convenient, because both methods give us point estimates of the life expectancy of each individual in the sample (and of any other interesting moments of the distribution that characterizes expectations). However, neither road to point identification is free of potholes. For the non-linear least squares approach, there is some tension between first invoking an estimation procedure for the parametric distributions, relying explicitly on the idea of some kind of reporting error, and then analyzing the estimated distributions as if they are identically equal to expectations. Such a two-step approach does not take into account the effect of reporting errors on the subsequent analysis. Another disadvantage of the non-linear least squares method is that it requires one to impose some parametric form on expectations. Though the Gompertz and Weibull distributions might adequately model human survival, there is no guarantee that such descriptive accuracy for objective survival translates into descriptive adequacy for subjective expectations. Neither of these objections applies to flexible spline approximation, since the constructed survival functions follow the data exactly and impose no parametric restrictions on expectations. However, spline interpolation does smooth expectations between the reported probabilities in the data. Though smooth functions may reflect beliefs accurately, we want to investigate how much we can learn about expectations without relying on any parametric assumptions or interpolation.

In order to derive bounds on remaining life expectancy, we start from the same point as for the spline interpolation approach and interpret the reported probabilities as points on the subjective survival function. However, we do not assume any particular form of expectations between those points. The implication of this is that our data identifies boxes within which the subjective survival curve lies, but contains no information on the location of the curve within the boxes. The darkest grey area in the left panel of Figure 4.1 is the admissible set for survival expectations under the assumption that reported probabilities lie exactly on the subjective survival curve. Note that any function

is allowed, even a step function, as long as it passes through the points given in the data. Though this makes the procedure very flexible, it does not by itself allow us to calculate bounds on life expectancy for a large majority of cases: in order to construct the bounds on life expectancy we need to impose a minimal and maximal age of death. As in the case of cubic spline interpolation explained in section 4.3.3, we assume death occurs between the current age of the respondent and age 110 (results with maximum ages equal to 100 and 120 are similar and available on request). Given the observed subjective probabilities and our constructed age-range across which death may occur, we simply trace the bottom edges of the darkest boxes in Figure 4.1 to calculate the lower bound on life expectancy. Likewise, the upper edges yield the highest life expectancy that is consistent with the subjective data.

4.3.5 Rounding

Common rounding scheme

Another methodological consideration in our analysis of subjective longevity is the effect of rounding of subjective probabilities on our inference. Rounding means that we cannot interpret the reported probabilities as points that are exactly on the subjective survival curve. Instead, rounded probabilities are informative of intervals within which the true subjective probabilities fall. For instance, a probability equal to 20 percent that is rounded to a multiple of five indicates that the true probability is in the interval $[17.5, 22.5]$, and the same probability rounded to a multiple of ten means that the true probability is in $[15, 25]$. This example shows that different rounding rules are observationally equivalent when we consider one single probability. Therefore, we apply the strategy proposed in Manski and Molinari (2010) to infer the degree of rounding from a set of related probability questions under the assumption that they result from the same rounding rule.[4] That is, we assume that the answers

[4]There is some evidence that supports the assumption of a common rounding rule for different questions that describe a single probability distribution is plausible. For instance, De Bresser and Van Soest (2013b) estimate a model that fits subjective distributions for the replacement rate of income at retirement from 6 probabilistic questions that are comparable to the survival questions analyzed in this paper. They allow different rounding for questions within a single sequence, but include survey and individual effects to describe persistence in

to all subjective longevity questions from a given respondent are rounded similarly and select the most conservative rounding rule that is consistent with the reported probabilities. Following Manski and Molinari (2010), we allow for the following rounding rules:

- All probabilities equal to 0 or 100: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = \left[\max\left\{0, P_{ik} - 50\right\}, \min\left\{P_{ik} + 50, 100\right\}\right]$
- All probabilities equal to 0, 50 or 100: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = \left[\max\left\{0, P_{ik} - 25\right\}, \min\left\{P_{ik} + 25, 100\right\}\right]$
- All probabilities multiples of 10: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = \left[\max\left\{0, P_{ik} - 5\right\}, \min\left\{P_{ik} + 5, 100\right\}\right]$
- All probabilities multiples of 5: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = \left[\max\left\{0, P_{ik} - 2.5\right\}, \min\left\{P_{ik} + 2.5, 100\right\}\right]$
- Some probabilities in $\{1, 2, 3, 4\}$ or $\{96, 97, 98, 99\}$:
  - $P_{ik} \in \{0, 1, 2, 3, 4, 5\} \cup \{95, 96, 97, 98, 99, 100\}$: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = P_{ik}$
  - $P_{ik} \in [6, 94]$: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = \left[\max\left\{0, P_{ik} - 2.5\right\}, \min\left\{P_{ik} + 2.5, 100\right\}\right]$
- Not in any of the categories above: $P_{ik} \in \left[P_{LB}, P_{UB}\right] = P_{ik}$

The left panel of Figure 4.1 shows how rounding changes the construction of the identified region for the hypothetical data from Table 4.1. Without rounding, the identified region for mortality expectations is that indicated by the darkest boxes. All probabilities in Table 4.1 are multiples of 5 and two probabilities are not multiples of 10, so according to the Manski and Molinari rounding scheme probabilities are rounded to multiples of 5. The lighter grey bands indicate how the area of the identified region increases when we allow for rounding, assuming that all probabilities are rounded to the same extent.

Analogously, the right panel of Figure 4.1 shows how we combine a common rounding scheme with cubic spline interpolation of expectations. The darkest

rounding behavior. Together these effects account for 55-60 percent of total error variance, the remaining 40-45 percent being idiosyncratic variation that differs between questions in a sequence.

shaded area in the figure corresponds to the identified region for the survival function under common rounding. We construct that region by tracing out the upper and lower limits on the individual probabilities, which are equal to $P_{it} + 2.5$ and $P_{it} - 2.5$ respectively, by means of cubic splines. Figure 4.1 shows that for our hypothetical data the resulting region in which the survival function lies is considerably narrower if we interpolate between reported probabilities (right panel) compared to the situation in which we do not interpolate (left panel), even if we do not allow for rounding in the latter scenario. The difference suggests that rounding is not as important a limitation of the informativeness of the data as is the coarseness of the grid on which we elicit expectations.

General rounding scheme

According to the model of rounding explained in the previous section, probabilities reported by an individual respondent are all generated by a common rounding rule (which is identified using all probabilities in the sequence). An alternative approach allows the extent of rounding to vary across the probabilities that describe an individual's expectations. Following this logic, we assume that reported probabilities are all rounded to the crudest extent possible. We allow for rounding to multiples of 100, 50, 25, 10 and 5. That is, a reported probability of 100 is interpreted as evidence that the true probability lies in the interval $[50, 100]$ and a reported 35 implies the interval $[32.5, 37.5]$, regardless of the other probabilities reported by that respondent. Probabilities that can only result from rounding to a multiple of 1 are interpreted as indicative of an interval with width 5 (we do not interpret any probabilities as exact representations of underlying expectations).

This rounding scheme leads to broad intervals for some probabilities, especially for those equal to 0, 50 and 100. However, imposing monotonicity on the true, partially identified, probabilities helps to narrow down the bounds on reported probabilities to informative levels. Monotonicity implies that the lower bound on the true probability for any threshold can never be below the lowest possible probability at the following threshold. Similarly, the upper bound at any threshold can never be above that at the previous threshold. For instance, if we observe a probability equal to 50% for the first age threshold of

70, the general rounding scheme interprets that probability as indicative of an interval equal to $[25, 75]$ for the corresponding true probability. However, if the reported probability for age 75 is 40%, we know that the true probability for age 70 cannot be smaller than the lower bound of 35% (since that is the lower bound on the probability for age 75). The lightest bands in the left panel of Figure 1 illustrate the general rounding procedure for the hypothetical data from Table 4.1. Note that the general rounding rule does not affect the admissible set for those age brackets at which the reported probability was rounded to a multiple of 5 but not of 10, since the common rounding rule assumes that all probabilities are rounded to multiples of 5. In this case general rounding only increases the size of the admissible set at those thresholds for which the reported probability is a multiple of 10.

Similarly, the lightest band in the right panel of Figure 4.1 shows the identified region for the survival function under general rounding if we interpolate between reported probabilities. As is the case for the approach without interpolation, allowing for different extents of rounding within the reported probability sequences changes the width of the identified region only at those thresholds for which the reported probability is also a multiple of 10 (the ages 70, 75 and 85).

Non-parametric bounds under the monotonic hazard restriction 4.3.6

The bounds without smoothing presented in the previous section embody a worst case scenario of what we can identify about beliefs if we cannot, or will not, impose structure beyond what is given in the data. A drawback of this generality is that the admissible set for the survival curve includes functions that are unlikely to represent actual expectations, such as the step functions mentioned above. We can reduce the size of the admissible set, and consequently the width of the bounds on life expectancy, by imposing restrictions on expectations.

Two plausible assumptions are that survival curves are continuous and that the hazard of dying, e.g. the probability of dying today conditional on having survived up to today, is weakly increasing in age. We construct admissible sets for the survival curves under these assumptions, allowing for rounding according to the schemes explained in sections 4.3.5 and 4.3.5. As explained in

Appendix 4.A, the reported probabilities are only consistent with a monotonic hazard of death if we allow for the general rounding scheme from section 4.3.5. However, in that case all results are similar to those obtained if we do not restrict expectations. Therefore, we do not report them in the main text and refer the reader to Appendix 4.A for more information and additional results.

4.4 Data quality and descriptives

We use the 2011 wave of the yearly *Pensioenbarometer*, a survey administered to the respondents of the CentERpanel. Data collection was financed by Netspar, Network for Studies on Pensions, Aging and Retirement. The CentERpanel is administrated by CentERData and is representative of the Dutch adult population. The sample consists of approximately 2500 respondents age 16 and older, but due to the focus on pensions the *Pensioenbarometer* surveys are only elicited from respondents who are older than 24. All CentERpanel questionnaires are administered via the Internet and members of the panel without Internet access are provided with a set-top box to maintain representativeness.

The 2011 *Pensioenbarometer* was distributed to 2,396 potential respondents and was returned by 1,577 of them (65.8 percent survey response). Item non-response to the subjective survival questions is not an issue: 94.6 percent of the panel members who filled out the questionnaire provided an answer to all survival questions. Furthermore, violations of the arithmetic of probabilities are rare despite the fact that no safeguards were applied to ensure logical consistency of the reported probabilities: 97.1 percent of the complete responses are weakly monotonically decreasing (that is: the probabilities weakly decrease for older age thresholds). Because of these observations, we do not model item non-response or logical inconsistencies. After removing incomplete or logically inconsistent response sequences, we are left with 1,381 observations.

Table 4.2 presents descriptive statistics of the subjective probability questions. Due to the age-eligibility criteria described in section 4.3.1, the sample sizes are larger for questions referring to older ages. Nonetheless, 1,157 respondents (84 percent of the sample) answer all five questions. Though the average probability of survival decreases with the age thresholds, the sample mean subjective probability of surviving past the age of 90 is rather high at

27 percent. Table 4.2 shows that the top quantile of the distribution of the subjective probabilities decreases less across thresholds than the lowest one, both in absolute and relative terms. Respondents use the full response scale for all questions.

Table 4.2: Descriptive statistics of the reported probabilities

	N	Mean	S. D.	Min	Q25	Mdn	Q75	Max
Age 70	1,157	75.86	21.23	0	60	80	90	100
Age 75	1,262	67.42	23.69	0	50	70	85	100
Age 80	1,335	55.25	25.95	0	40	60	75	100
Age 85	1,374	41.52	26.47	0	20	40	60	100
Age 90	1,381	27.16	25.23	0	5	20	50	100

We illustrate the importance of rounding in Table 4.3. The top panel of Table 4.3 contains a tabulation of the different categories of rounding according to the common rounding scheme described in section 4.3.5. Frequencies are presented for the entire sample, by sex and by level of education. Rounding appears quite important: about half of the sample reports only probabilities that are multiples of 5, with at least one probability that is not a multiple of 10. Another third of the sample provides probabilities that are consistent with rounding to multiples of 10. Cruder forms of rounding do occur but are rare: 2 percent of the reported sequences consist entirely of probabilities equal to zero or one hundred and another 4 percent consist only zeros, fifties and one hundreds. Focal 50/50s are not likely to be an important concern for our data, since no more than 2.5 percent of the respondents answer 50 percent to all questions. Confirming the analysis in Manski and Molinari (2010), we find some evidence that respondents may round probabilities near the extremes of zero and one hundred differently since 8 percent of the sample reports multiples of 1 near the extremities of the scale. Only 3 percent of the responses are incompatible with all other forms of rounding and are thus interpreted as exact answers (though one might argue that these are rounded to a multiple of 1, the most precise level of rounding allowed for in the questionnaire). Interestingly, we do not find any gender nor education pattern in rounding behavior, as can be seen in the remaining columns of the top panel of Table 4.3.

The bottom panel of Table 4.3 describes rounding according to the general rounding scheme explained in section 4.3.5. The general rounding scheme also

Table 4.3: Incidence of rounding

	Common rounding rule (rounding at the level of the individual respondent)						
					By education		
	Frequency	%	Men (%)	Women (%)	Low (%)	Middle (%)	High (%)
All 0 or 100	27	1.96	2.05	1.83	1.92	2.53	1.58
All 0, 50 or 100	52	3.77	3.72	3.83	2.64	5.57	3.34
All multiples of 10	448	32.44	31.54	33.61	34.86	30.38	32.16
All multiples of 5	710	51.41	52.31	50.25	50.96	51.39	51.85
Some in [1,4] or [96,99]	105	7.60	7.69	7.49	6.49	8.10	7.91
Other	39	2.82	2.69	3.00	3.13	2.03	3.16
Total	1,381	100	100	100	100	100	100
	General rounding rule (rounding at the level of individual probability)						
			By age threshold				
Multiples of...	Frequency	%	70 (%)	75 (%)	80 (%)	85 (%)	90 (%)
...100	717	11.02	17.20	10.30	6.89	7.79	13.69
...50	933	14.33	14.17	11.89	16.40	16.74	12.31
...25	574	8.82	7.17	11.65	9.14	9.10	7.02
...10	3,274	50.30	49.70	49.29	54.68	49.34	48.44
...5	808	12.41	9.59	13.79	11.01	14.34	12.96
...1	203	3.12	2.16	3.09	1.87	2.69	5.58
Total	6,509	100	100	100	100	100	100

The numbers in this table are frequencies of the reported probabilities that fall in the various rounding categories.

indicates substantial rounding, since 50 percent of reported probabilities are multiples of 10 (but not of 50 or 100). Another 14 percent of probabilities are equal to 50 and only 3 percent can only result from rounding to multiples of 1. The variation in frequencies of the different rounding categories across the age thresholds accords with intuition. Respondents express greater certainty, report a higher fraction of zeros and one hundreds, near the extreme ends of the age range. Conversely, 50/50s are more prevalent at the age thresholds of 80 and 85, expressing uncertainty about surviving past those ages.

Some demographic characteristics of the sample are shown in Table 4.4. Men and women are evenly represented and the average age is 56. Three quarters of the sample lives with a partner. The sample is mostly healthy: 75 percent rates their own health either as "excellent" or "very good", while only 8 percent is in "bad" health. 41 Percent of the sample has finished some form of higher education (university or an applied college). A further breakdown by sex shows that the sample is better educated than the Dutch average: 44 percent of men and 38 percent of women have completed higher education compared to nationwide averages of 31 and 26 percent respectively in 2009.

Furthermore, 50 percent has a monthly net household income above 2,600 euro. The average gross personal income in our sample is close to that of the population at large: the economically active within the sample earn 2,978 euro per month compared with a national average of 2,900 euro in 2010.

Table 4.4: Descriptive statistics of demographic variables

	N	Mean	S.D.	Min	Max
Demographics					
Gender: male	1,381	0.56	0.50	0	1
Age	1,381	56.12	13.07	25	88
Lives with partner	1,381	0.76	0.43	0	1
Health					
Excellent	1,381	0.14	0.35	0	1
Good	1,381	0.61	0.49	0	1
Fair	1,381	0.18	0.38	0	1
Bad	1,381	0.08	0.26	0	1
Education					
Primary	1,380	0.05	0.21	0	1
Lower secondary	1,380	0.25	0.44	0	1
Higher secondary	1,380	0.11	0.31	0	1
Lower vocational	1,380	0.18	0.38	0	1
Higher vocational	1,380	0.27	0.45	0	1
University	1,380	0.14	0.35	0	1
Income					
HH. income <eu. 1,151	1,381	0.07	0.25	0	1
HH. income eu. 1,151 - 1,800	1,381	0.16	0.37	0	1
HH. income eu. 1,801 - 2,600	1,381	0.27	0.44	0	1
HH. income >eu. 2,600	1,381	0.50	0.50	0	1
Gross personal income:					
Economically active (eu/month)	695	2,978.35	1,713.08	0	13,500

Results 4.5

Point- and interval estimates of life expectancy 4.5.1

Table 4.5 contains descriptive statistics of the expected age of death calculated from the individual-specific parametric survival curves and the spline survival functions. Men expect to live to the age of 82 on average, which is slightly

below the average prediction of 82.85 years found in the cohort life tables. For women we find a larger discrepancy between the average subjective life expectancy and the average actuarial forecast: women expect to live to age 82-83 while the average actuarial prediction is 85.72. This larger discrepancy for women was also observed by Perozek (2008) for the HRS and Kutlu and Kalwij (2012) for the same panel that we analyze (but a different dataset, see section 4.2). If we divide the sample in age groups, we find that men and women of all ages expect to live shorter than predicted in their cohort life tables. We will analyze the statistical significance of the differences between average subjective life expectancy and the forecasts in cohort life tables in section 4.5.3 below.

Unsurprisingly, subjective mortality expectations exhibit much more variation than the life tables (the latter only vary with gender and year of birth). The standard deviations of the expected age of death are around 8, while that of the actuarial forecasts is 2.15 for men and 1.42 for women. In the next subsection we analyze whether this variation in expectations is related to covariates such as health and socio-economic status.

The life expectancies calculated from parametric models are very similar to those derived from spline survival functions: all correlations between the expected ages of death are above 0.98. Hence, given sufficiently rich data, computed mortality expectations are robust with respect to the choice for a (non-)parametric model for subjective survival. This finding is in line with De Bresser and Van Soest (2013a), who document that the first two moments of their subjective distributions of the expected replacement rate of income at retirement are very similar for a parametric (log-normal) specification versus cubic splines.

Now we turn to the bounds that we derive using the methodology described in sections 4.3.4 and 4.3.5. Table 4.6 presents sample averages of the bounds on life expectancy. Expectations of men and women are summarized in the top and bottom panel of the table respectively. Table 4.6 reports descriptives for the case in which we do not interpolate expectations (*no smoothing*) and for that in which we do interpolate using cubic splines (*smoothing using cubic splines*). The bounds derived using linear spline interpolation are very similar to those discussed here and are available upon request. Table 4.6 contains averages for bounds under the assumptions of no rounding (columns under *no*) and the

Table 4.5: Point estimates of life expectancy

	Men					
		By age bracket				
	Overall	25-35	36-45	46-55	56-65	66+
Parametric approximations						
Gompertz	81.71	81.90	80.94	79.73	80.95	84.48
	(7.61)	(7.43)	(10.48)	(8.21)	(6.98)	(4.98)
Weibull	82.34	83.10	82.16	80.35	81.40	84.93
	(7.65)	(7.65)	(10.65)	(8.19)	(7.10)	(4.94)
Non-parametric approximations						
Linear splines	82.11	80.72	79.08	79.59	82.02	85.94
	(8.09)	(7.78)	(9.29)	(8.10)	(7.91)	(5.99)
Cubic splines	81.95	81.24	79.33	79.40	81.82	85.55
	(8.29)	(7.76)	(9.62)	(8.36)	(8.30)	(6.12)
Cohort life table forecasts						
Statistics Netherlands life tables	82.85	80.56	80.84	81.39	82.61	85.62
	(2.15)	(0.07)	(0.11)	(0.23)	(0.47)	(1.91)
N	780	36	104	185	233	222

	Women					
		By age bracket				
	Overall	25-35	36-45	46-55	56-65	66+
Parametric approximations						
Gompertz	81.99	81.35	83.08	80.73	80.76	84.63
	(8.39)	(9.86)	(10.77)	(8.68)	(7.05)	(5.51)
Weibull	82.63	82.52	83.85	81.40	81.25	85.07
	(8.38)	(9.77)	(10.64)	(8.74)	(7.17)	(5.49)
Non-parametric approximations						
Linear splines	81.78	79.29	80.72	80.28	81.61	86.03
	(8.33)	(9.18)	(8.91)	(8.27)	(7.86)	(6.51)
Cubic splines	81.61	79.32	80.73	80.10	81.33	85.76
	(8.61)	(9.78)	(9.22)	(8.65)	(8.04)	(6.80)
Cohort life table forecasts						
Statistics Netherlands life tables	85.72	84.29	84.49	84.95	86.04	87.99
	(1.42)	(0.04)	(0.08)	(0.21)	(0.37)	(1.24)
N	601	61	107	137	178	118

Reported numbers are averages; standard deviations in parentheses.

common rounding rule described in section 4.3.5 (columns under *common*). Analogous results for the general rounding rule, section 4.3.5, are given in Appendix 4.B. The bounds under general rounding show similar patterns to those discussed in the main text, but are always slightly wider than the bounds under the common rounding scheme.

According to the bounds without smoothing, which do not impose any restrictions on expectations beyond being consistent with the data, men expect to live to age 76-88 on average. We find very similar figures for women. Expectations are conditional on the current age of the respondent, so the lower bounds of the intervals for life expectancy increase with age. The upper bound, on the other hand, appears to follow a U-shaped pattern and is the lowest for the 46-55 age bracket. In addition to summary statistics of the bounds, Table 4.6 also contains information on the average width of the intervals for life expectancy and of the relative decrease in width compared to the uninformative interval (which sets the minimum life expectancy of respondent i to zero and the maximum to $T_{\max} - age_i$, where $T_{\max} = 110$). The former provides an absolute measure of the informativeness of the bounds, while the latter takes into account that the width of the uninformative interval is smaller for older respondents (since the maximum lifespan does not vary with age). The intervals for life expectancy are about 12 years wide on average which translates into a reduction relative to the uninformative interval of 78 percent. The average absolute width drops with age from 17-18 years for the youngest age group to 9 years for those of age 66 and older. The percentage decrease in width relative to the uninformative interval does not show any clear pattern with age.

By definition allowing for rounding makes the estimated intervals wider and thus less informative. Assuming a common rounding rule results in bounds with an average width of 15 years, or an average reduction from the uninformative interval of 72 percent.

Table 4.6: Sample averages of bounds on life expectancy derived under absence of rounding and common rounding

	Men											
			By age bracket									
	Overall		25-35		36-45		46-55		56-65		66+	
	No	Common	No	Common	No	Common	No	Common	No	Common	No	Common
No smoothing												
LB	76.43	75.03	72.46	69.31	70.93	68.62	73.31	71.73	77.12	75.82	81.55	80.88
UB	87.91	89.41	88.98	91.09	87.22	89.33	85.87	87.41	86.91	88.41	90.45	91.91
Width	11.37	14.39	16.52	21.78	16.29	20.71	12.56	15.68	9.80	12.59	8.90	11.03
Percentage decrease[a]	77.88	72.13	78.80	71.98	76.42	70.00	78.90	73.61	80.13	74.47	75.22	69.48
N	780		36		104		185		233		222	
Smoothing using cubic splines												
LB	–	80.44	–	78.54	–	77.20	–	77.84	–	80.32	–	84.54
UB	–	83.48	–	83.58	–	81.77	–	80.96	–	83.16	–	86.67
Width	–	3.04	–	5.04	–	4.57	–	3.12	–	2.85	–	2.13
Percentage decrease[a]	–	94.18	–	93.49	–	93.35	–	94.71	–	94.23	–	94.19
N	778		36		104		184		232		222	

	Women											
			By age bracket									
	Overall		25-35		36-45		46-55		56-65		66+	
	No	Common	No	Common	No	Common	No	Common	No	Common	No	Common
No smoothing												
LB	75.66	73.99	70.08	66.77	72.66	70.20	73.91	72.10	76.83	75.68	81.52	80.83
UB	87.93	89.59	88.49	90.43	88.77	90.50	86.66	88.97	86.40	88.00	90.65	91.81
Width	12.27	15.60	18.41	23.66	16.12	20.30	12.74	16.57	9.57	12.32	9.13	10.98
Percentage decrease[a]	78.09	72.27	76.42	69.67	76.87	70.86	78.75	72.33	80.59	75.04	75.51	70.67
N	601		61		107		137		178		118	
Smoothing using cubic splines												
LB	–	79.90	–	76.37	–	78.45	–	78.29	–	80.03	–	84.73
UB	–	83.29	–	81.75	–	82.84	–	82.15	–	82.73	–	86.64
Width	–	3.38	–	5.28	–	4.39	–	3.87	–	2.70	–	1.91
Percentage decrease[a]	–	94.10	–	93.06	–	93.70	–	93.51	–	94.55	–	95.01
N	600		61		106		137		178		118	

[a] Percentage decrease in width of the interval for life expectancy relative to the uninformative interval. The uninformative interval of respondent i is equal to $110 - age_i$.

We obtain much tighter bounds on life expectancy if we interpolate expectations between the elicited survival probabilities. Interpolation allows one to focus on the ambiguity that results from rounding, isolating it from the coarseness inherent in eliciting expectations by means of a limited number of probabilities. Under the common rounding assumption, interpolation reduces the average width of the interval from 15 years to 3 years. The reduction is apparent in all age groups. The average for the 25-36 year olds, for instance, is diminished from 22 to 5 years, while that among respondents aged 66 and older drops from 11 to 2 years.

4.5.2 Linear models

Next we investigate the relationship between life expectancy and demographic variables. We estimate linear regressions with the point identified life expectancies as dependent variables. For the bounds we apply partially identified models according to the methods presented in Imbens and Manski (2004).[5] Estimation results are presented in Table 4.7. The model specifications in that table pool men and women, because we could not reject the null hypothesis of equal coefficients for the sexes. Point identified models show that age and self-reported health are the most important covariates of subjective life expectancy. As expected, remaining life expectancy decreases non-linearly with age. Health too is strongly related to subjective life expectancy: compared to the baseline of people in excellent health, those in bad health expect to live 7 years shorter on average. Respondents in fair or good health also expect to live shorter than their healthier peers. Like Kutlu and Kalwij (2012), we do not find significant associations between life expectancy and education and income if we also condition on subjective health. If we remove health from the equation, we find that respondents from the lowest income group and those with the poorest education expect to die younger (estimates available on request). Note that the estimates are not sensitive to the way in which expectations are approximated: similar conclusions emerge whether we fit Weibull distributions or cubic splines.

[5] We estimate the interval regressions using the Stata program CI1D, presented in Beresteanu et al. (2010).

Table 4.7: Point and partially identified models of the remaining life expectancy

			Partially identified models					
			No smoothing		Cubic splines			
	Point identified models		No rounding		Common rounding		General rounding	
	Weibull	Cubic spline	LB	UB	LB	UB	LB	UB
Age	-1.740***	-1.539***	-3.278	2.256	-2.257	-0.796	-3.176	0.266
	(0.106)	(0.114)	(-3.526; 2.504)		(-2.475; -0.578)**		(-3.446; 0.535)	
Age squared/100	0.761***	0.670***	-1.762	3.079	0.0264	1.290	-0.883	2.103
	(0.0890)	(0.0978)	(-1.981; 3.298)		(-0.144; 1.461)		(-1.115; 2.334)	
Male	-0.808*	-0.616	-12.571	11.185	-3.802	2.663	-8.397	6.935
	(0.430)	(0.441)	(-13.423; 12.037)		(-4.687; 3.547)		(-9.436; 7.974)	
Educ. - primary school	-1.092	-0.789	-13.151	11.713	-3.811	1.940	-9.338	5.529
	(0.937)	(1.020)	(-14.734; 13.297)		(-5.527; 3.655)		(-11.642; 7.833)	
Educ. - higher secondary	-0.239	-0.0520	-12.898	12.681	-3.602	3.062	-8.626	7.707
	(0.735)	(0.781)	(-14.184; 13.968)		(-5.210; 4.671)		(-10.344; 9.425)	
Educ. - lower vocational	0.143	-0.144	-14.136	13.911	-4.349	3.811	-9.803	9.065
	(0.696)	(0.705)	(-15.540; 15.314)		(-5.717; 5.179)		(-11.482; 10.744)	
Educ. - higher vocational	-0.515	-0.415	-13.706	12.771	-4.008	2.939	-9.247	7.592
	(0.575)	(0.592)	(-14.861; 13.926)		(-5.134; 4.064)		(-10.643; 8.988)	
Educ. - university	0.456	0.164	-14.281	14.712	-3.961	3.808	-9.151	9.085
	(0.680)	(0.674)	(-15.609; 16.040)		(-5.192; 5.039)		(-10.822; 10.755)	
Income < 1150 euro/month	-0.336	-0.438	-15.306	14.327	-4.586	3.708	-9.628	9.741
	(1.125)	(1.087)	(-17.247; 16.268)		(-6.639; 5.762)		(-12.091; 12.203)	
Income 1801-2600 euro/month	-0.282	-0.120	-12.737	12.744	-3.724	3.291	-8.101	8.252
	(0.703)	(0.716)	(-13.908; 13.915)		(-4.928; 4.495)		(-9.816; 9.968)	
Income > 2600 euro/month	-0.331	-0.0576	-13.122	13.059	-3.637	3.343	-8.263	8.421
	(0.668)	(0.671)	(-14.461; 14.398)		(-4.890; 4.596)		(-9.965; 10.122)	
Health - bad	-6.713***	-7.668***	-23.717	8.920	-11.606	-2.449	-16.660	3.988
	(1.039)	(1.082)	(-25.682; 10.886)		(-13.736; -0.319)**		(-19.009; 6.337)	
Health - fair	-5.378***	-6.351***	-21.302	8.746	-10.190	-2.236	-14.821	3.312
	(0.764)	(0.798)	(-22.797; 10.242)		(-11.764; -0.662)**		(-16.661; 5.152)	
Health - good	-1.366**	-2.155***	-15.403	11.416	-5.773	1.670	-10.151	6.549
	(0.660)	(0.697)	(-16.734; 12.746)		(-7.092; 2.989)		(-11.820; 8.217)	
Constant	101.9***	93.39***	14.901	169.661	72.721	113.506	41.289	139.392
	(3.249)	(3.396)	(7.710; 176.852)**		(66.029; 120.198)**		(32.937; 147.744)**	
N	1,380	1,380	1,380		1.377		1,073	
R-squared	0.735	0.680	–		–		–	

Standard errors in parentheses for point identified models
95% confidence sets in parentheses for the partially identified models
***significant at 1%; **significant at 5%; *significant at 10%

The rightmost columns of Table 4.7 present estimates from models with interval-censored life expectancy as the dependent variable. The estimates from partially identified models estimated on the bounds without smoothing show that little can be learned about differences in life expectancy across the sample if one is unwilling to interpolate expectations between data points. There are clear differences between the identified sets that are in line with the coefficient estimates from point identified models. For instance, the set-estimate of the coefficient of being in bad health, which is associated with a 7-year lower life expectancy relative to the baseline in point identified models, is $(-23.7; 8.9)$. Though that interval suggests a negative correlation, zero is included in all identified sets (and by implication in all 95% confidence sets) except for the constant. Hence, we are not able to draw conclusions about the sign of any of the coefficients: a clear indication that the bounds are too wide for useful inference. Since rounding only makes the bounds on life expectancy wider, this holds even more strongly if we allow for rounding. Without additional assumptions there is not enough information in the data to make precise inference on differences in life expectancy across different socio-demographic groups.

Table 4.7 shows that if we do smooth expectations by means of cubic splines and allow for rounding according to a common rule, we find evidence for similar correlations that are apparent in non-censored models. In particular, zero is not included in the 95% confidence sets for the coefficients on the indicators for bad and fair health. These results show that the limited number of elicited points on the survival functions is a more important limitation on the informativeness of the data than (common) rounding. However, if we simultaneously interpolate and allow for general, worst-case, rounding, the identified sets again become too wide to draw any conclusions (though they remain narrower than in the no-smoothing, no-rounding case). Therefore, the informativeness of the data hinges not only on our willingness to smooth beliefs, but also on the particular type of rounding that we assume. Finally, note that the sample sizes for the partially identified models of smoothed expectations are smaller than those for the other models. This is due to the fact that the cubic spline functions that trace the upper and lower bounds for the true, non-rounded probabilities sometimes cross each other, in which case we drop that observation. This is rare for the case of common rounding, only

3 observations are lost this way, but quite common for general rounding for which we lose 307 observations (or 22% of the sample). This complication does not occur with linear splines, so we verified the patterns from Table 4.7 using linear splines. All estimates are similar to those reported here, results are presented in Appendix 4.C.

Consistency of expectations with life tables 4.5.3

As mentioned in the introduction, consistency of subjective expectations with published life-tables has been an active area of research since the first subjective longevity questions were posed in the HRS in 1992 (Gan et al. 2005). Most recently, Perozek (2008) found that in the 1992 wave of the HRS women expected to live shorter than the 1992 actuarial forecasts predicted and that these expectations were in line with subsequent downward revisions in forecasts of female life expectancy (though even the most recent forecasts still implied longer life expectancy compared to the expectations). Similar patterns are found for a Dutch sample by Kutlu and Kalwij (2012). We now investigate the correspondence between the Dutch cohort life tables assembled by Statistics Netherlands, as of December 2010, and subjective expectations. First we look at the point identified life expectancies. Figure 4.2 presents kernel regressions of subjective life expectancy, computed from cubic splines, and the actuarial forecast. Analogous figures based on Weibull or Gompertz distributions and linear splines are similar and are available upon request. The left panel of Figure 4.2 shows that for men expectations and actuarial forecasts are very much in line on average: the actuarial tables lie within the 95% confidence interval for the average subjective life expectancy for all ages. For women, on the other hand, we find that official forecasts are above the 95% confidence interval for all ages between 30 and 70 (see the right graph in Figure 4.2). Hence, women indeed expect to die significantly younger than the actuarial estimates predict. The size of this difference is large: close to 5 years around the age of 60. Note, however, that these estimates do not take into account missing information between elicited probabilities or rounding.

Figure 4.3 plots 95% confidence bands for subjective remaining life expectancy for men (left panel) and women (right panel). These bands are based on bounds derived without smoothing expectations and span the width

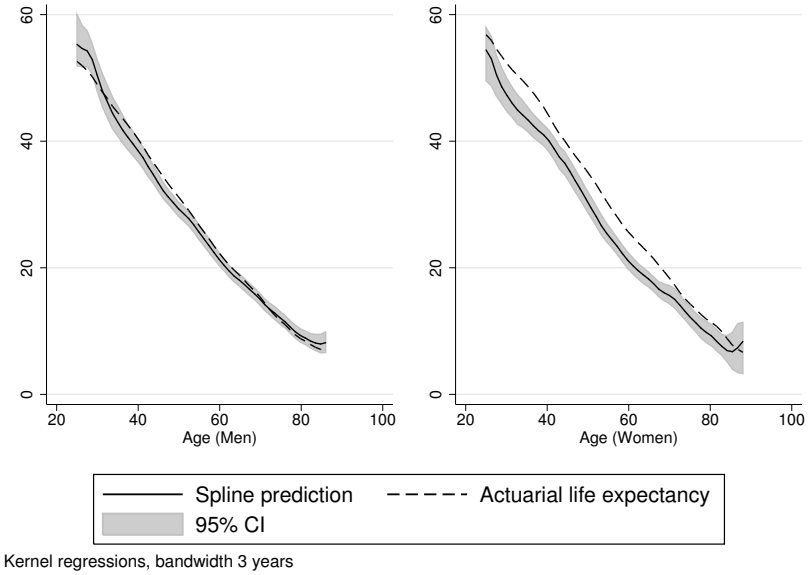


Figure 4.2: Actuarial forecasts and subjective life expectancy (expectations approximated using cubic splines)

from the lower end of the 95% confidence interval for the lower bound on life expectancy to the higher end of the 95% confidence interval for the upper bound. The solid lines are the corresponding predictions from life-tables. Even without rounding, the data for men are consistent with the actuarial forecasts across all ages. Moreover, allowing for rounding does not affect the average bounds much, though the effect is slightly larger at younger ages. The right panel of Figure 4.3 contains the same information for women. Contrary to the findings in Figure 4.2, the subjective data of women are on average consistent with the life tables once we take into account that we only observe a few points on the subjective survival function. Though we find that the actuarial forecasts from Statistics Netherlands are on the high end of the confidence interval for subjective life expectancy, they are between the bounds that do not allow for rounding. Even around age 60, where Figure 4.2 indicates that the difference between life tables and expectations is substantial, the non-parametric bounds do not allow us to conclude that the actuarial forecasts are above the average upper bound on subjective life expectancy. When we take rounding into account, the actuarial figures lie well within the interval for all ages. Hence, we conclude that without additional assumptions we cannot reject the hypothesis that expectations are on average consistent with the life tables

for both men and women in our sample. Robustness checks indicate that these conclusions remain largely unchanged when we lower the maximum lifespan to 100 years: for men the actuarial forecasts remain well within the subjective bounds for all ages, while for women the actuarial forecasts remain within the 95% confidence bands if we allow for rounding.[6] Moreover, the bounds on life expectancy remain consistent with the actuarial forecasts if we restrict the hazard of death to be monotonically increasing with age (see Appendix 4.A).

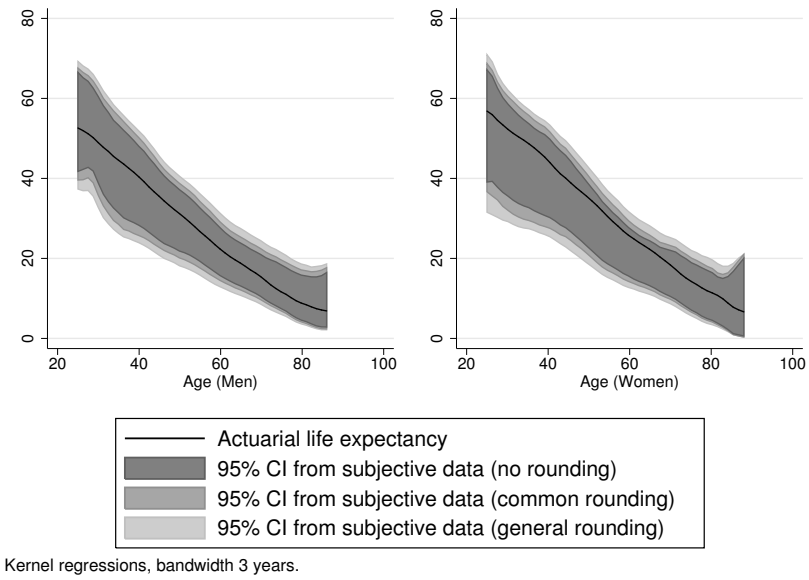


Figure 4.3: Non-parametric bounds on life expectancy without interpolation between reported probabilities

In the previous sections we have shown that the width of the bounds can be reduced substantially by simultaneously allowing for rounding of the reported probabilities and interpolating expectations between the elicited points on the subjective survival functions. A natural question is whether rounding alone can close the gap between average expectations and the actuarial forecasts for women that is evident in Figure 4.2. The right panel of Figure 4.4 shows that common rounding closes the gap for all ages except for 50 to 70 year olds and around the age of 30, for which small differences remain. The more conservative general rounding scheme, on the other hand, removes those discrepancies. Under general rounding, we cannot reject the null hypothesis

[6]Estimates available upon request.

that the average upper bound of the intervals is equal to the corresponding life table forecast for any age. The same conclusions emerge if we approximate expectations by means of linear splines.

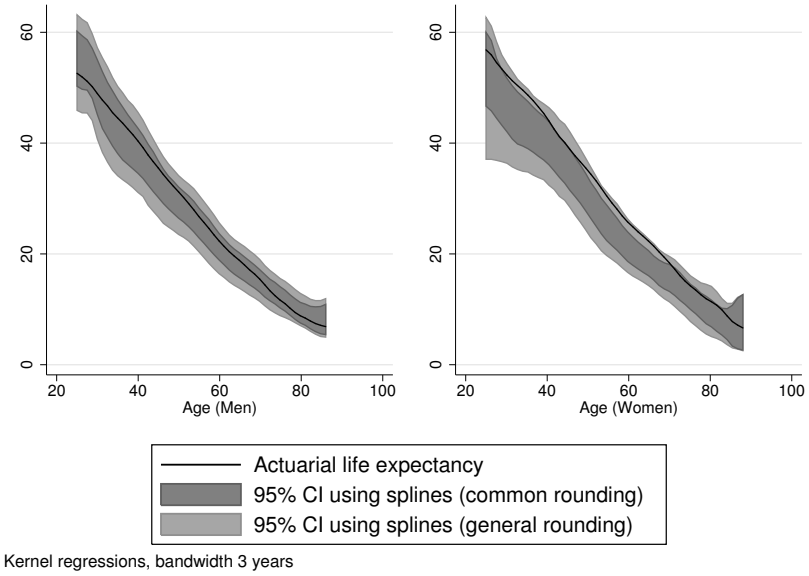


Figure 4.4: Non-parametric bounds on life expectancy with cubic interpolation between reported probabilities

4.6 Conclusion

When investigating the subjective expectations held by survey respondents, researchers often use parametric forms to describe beliefs (see Dominitz and Manski 1997, or Dominitz 1998, for examples using income expectations and Perozek 2008, and Kutlu and Kalwij 2012, for recent examples using survival expectations). We analyze the consequences of such assumptions by contrasting non-linear least squares estimates of parametric subjective probability distributions with other approaches that are more flexible with respect to the form of expectations. In particular, we compare the parametric approach with cubic spline approximation of expectations as proposed by Bellemare et al. (2012). Moreover, we propose two methods to partially identify expectations, constructing identified sets within which subjective survival functions must fall to be consistent with the data. The first is a "worst-case" scenario in which

we neither assume any functional form for expectations nor do we interpolate between the elicited points on the subjective survival functions. We show that this procedure can easily be generalized to allow for rounding of the subjective probabilities reported in surveys. Furthermore, we illustrate how beliefs can be restricted in order to increase the precision of the bounds. The second approach focuses exclusively on the ambiguity in the data that is due to rounding of the reported probabilities. This method interpolates expectations between the elicited points, which in combination with an assumed rounding scheme yields identified regions for expectations that are much narrower than those that do not interpolate. Both methods we propose construct bounds on relevant aspects of beliefs.

Our application concerns mortality expectations from a representative sample of Dutch adults. We find that subjective life expectancies are similar regardless of whether we approximate expectations by parametric models or by (linear or cubic) spline interpolation: correlations between the resulting life expectancies are above 0.98.

If we do not interpolate expectations between elicited points on the survival functions, the bounds on life expectancy are too wide to be informative. In particular, models for interval-censored dependent variables fail to corroborate the associations between life expectancy and the covariates *age* and *health* that are highly statistically significant in non-censored models. Rounding makes the intervals even less informative and imposing that the hazard of death, the probability of dying today conditional on survival up to today, weakly increases with age does not narrow the admissible sets sufficiently to estimate differences in life expectancy between socio-demographic groups. However, if we do smooth survival functions between observed points, the intervals can be narrowed enough to allow for useful inference. Under the assumption that all probabilities reported by a given respondent are rounded to the same extent, the same assumption made by Manski and Molinari (2010), partially identified models show the same patterns that emerge from point identified models. Allowing for more conservative rounding, every probability is rounded to the maximum extent, again yields wide and uninformative intervals.

We match our subjective life expectancies to cohort life tables constructed by Statistics Netherlands. If we point identify expectations, either parametrically or non-parametrically, we confirm the finding from Perozek (2008) and Kutlu

and Kalwij (2012) that women expect to live shorter on average than the published life-tables predict. The difference is largest around the age of 60. Men, on the other hand, have expectations that are on average in line with the actuarial forecasts. The bounds on life expectancy confirm that beliefs held by men are consistent with life-tables in the aggregate, regardless of whether or not we interpolate between data points. In contrast to the case in which we point identify expectations, we can no longer reject that expectations of women are consistent with the forecasts if we do not smooth expectations. This remains true if we impose that the risk of death should be increasing in age and is even starker if we allow for rounding in the reported probabilities. If we interpolate expectations and simultaneously allow for common rounding, expectations of women are inconsistent with the life tables around the ages of 30 and 60 (the size of the remaining difference is much smaller). If we allow for the more conservative rounding scheme that does not impose common rounding, we cannot reject consistency of women's expectations with actuarial forecasts for any age.

The general idea that emerges from this paper is that it is possible to learn about subjective expectations without imposing parametric restrictions on beliefs or even point identifying them. The methods we propose are sufficiently flexible to take into account rounding issues that are relevant for survey data of many types. Moreover, they can be combined with plausible assumptions on beliefs to allow more precise inference that remains less restrictive than fully characterizing expectations by parametric distributions. Our partial identification framework yields new insights into the influence of parametric assumptions on the analysis of this important and increasingly popular type of data.

4.7 Acknowledgements

We thank Netspar for financially supporting the data collection and Arthur van Soest, Frederic Vermeulen, Martin Salm, Marcel Das, Pierre-Carl Michaud and Rob Alessie for their insightful comments.

Monotonically Increasing Hazard of Death 4.A

In the main text we analyze subjective expectations under two extreme sets of assumptions: either we characterize them fully by a parametric model, or we assume nothing beyond the probabilities given in the data. As a middle road, we now consider what we can learn if we impose that the subjective survival curve is continuous and that the probability of dying today is weakly increasing in age.

Algorithms 4.A.1

In order to analyze the admissible functions with these conditions, let us recast the problem in terms of the integrated hazard:

$$\Lambda(t) = -\ln S(t) \tag{4.2}$$

We can recast the set of admissible survival functions in terms of a set of admissible integrated hazards. Doing so already helps to illustrate conditions under which our assumptions are not compatible with the observed points. For our data with age brackets of equal size, it will be the case if the increase in integrated hazard between observed ages t and $t+5$ is larger than the increase between observed ages $t+5$ and $t+10$ etc.

The assumption that we want an increasing hazard implies that the integrated hazard function has to be convex. Hence, the most pessimistic survival function (assuming that the set of admissible function is not empty) consists in the case where the integrated hazard grows linearly between the observed points in the function. However, it is harder to characterize optimistic curves. Consider the simple 3-interval case shown in Figure 4.5 as an example.

In this example, the most pessimistic survival curve is defined by the case where hazard increases linearly between the points. This is the solid line on both panes of the figure. However, we cannot draw a simple envelope for the most optimistic case. We know that integrated hazard in the middle interval increases at most with the linear rate of the most pessimistic hazard. Then, we would know that the hazard rate cannot increase faster than this within the

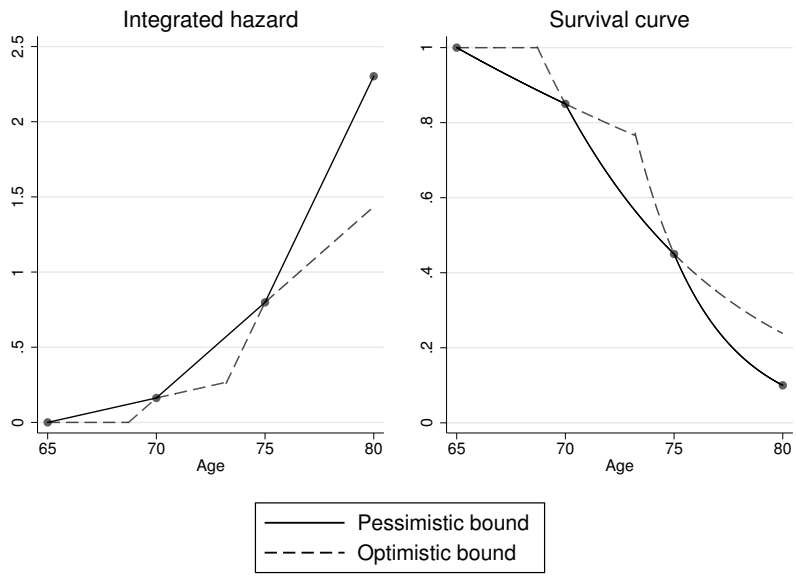


Figure 4.5: Non-parametric bounds under monotonicity and continuity: the case without rounding

first interval. The limiting case is given by the dashed line. In this case, the hazard does not increase at all during the first segment, and then, there is a linear increase in the second part only. Any curve more optimistic than that would necessarily lead to a non-convex function in the second interval. Then what about the second interval? We can use a similar logic. This time we know that the hazard cannot increase slower than the fastest rate of the first interval, or faster than the most pessimistic rate of the following one. Again, the dashed line corresponds to this scenario. Note however that the most optimistic case in the first interval necessarily leads to the most pessimistic case in the second interval. The dashed line is not an admissible integrated hazard function. This dashed line only provides a piecewise interval for the integrated hazard Đ and thus to our survival curve. Nevertheless, the set of the admissible functions is contained within the set of functions defined by these lines. For the sake of simplicity, we can run our analysis considering this larger set, though the bounds we derive using this method are not sharp.

Rounding

The case with rounding is more complicated. Let's denote Λ_i^{min} and Λ_i^{max} the minimum and maximum hazard at the ith observed point in the distribution, such that $\Lambda_0^{min} = \Lambda_0^{max} = 0$ correspond to the integrated hazard conditional on surviving to the first observed age (the current age of the respondent). Finally, suppose that these points are for ages $t_i, ..., t_n$, where i indexes the number of the age threshold (so for our data i runs from 1 for age 70 to 5 for age 90).

Most pessimistic case

Obtaining the upper bound of the interval is more challenging than simply considering the most pessimistic hazard, given that this scenario may not be admissible. However, we know that the hazard cannot decrease in subsequent intervals. It follows that the slope of the most pessimistic admissible case for interval i is given by:

1. Start from $i = 1$
2. Define the most pessimistic hazard as defined from the slope:

$$m_i^{max} = \min\left(\frac{\Lambda_{i+1}^{max} - \Lambda_i^{max}}{t_{i+1} - t_i}, ..., \frac{\Lambda_n^{max} - \Lambda_i^{max}}{t_n - t_i}\right)$$

3. If needed, redefine the maximal integrated hazard at a given point $j > i$, which will be the value:

$$\widetilde{\Lambda}_j^{max} = \min\left(m_i^{max}\left(t_j - t_i\right) + \Lambda_i^{max}, \Lambda_j^{max}\right)$$

   If $\widetilde{\Lambda}_j^{max} < \Lambda_j^{min}$, there is no admissible function.
4. Increase i by 1 and start from point 2 until you reach $i = n - 1$.

Most optimistic case

As it was the case without rounding, we will limit ourselves to finding piecewise intervals on integrated hazard. LetÕs define m_i^{slow} and m_i^{fast} the slowest and fastest increase in hazard within interval i. Let $m_1^{slow} = 0$. When we compute the slowest increase, we will need to consider the origin of this increase. We

define α_i as the age from which originate this steady increase in hazard. We start from $\alpha_1 = 1$ and thus, the corresponding integrated hazard is given by $\Lambda_1^{\alpha} = \Lambda_1^{min} = 0$.

1. Start from $i = 1$.
2. Determine the fastest rate at which the integrated hazard can grow within this interval. This is given by:

$$m_i^{fast} = \min\left(\frac{\Lambda_{i+1}^{max} - \Lambda_i^{min}}{t_{i+1} - t_i}, ..., \frac{\Lambda_n^{max} - \Lambda_i^{min}}{t_n - t_i}\right)$$

3. (Backward correction, only if $i > 2$) Once the steepest increase is determined, we may realize that previous increases are too steep to be admissible. We may therefore want to correct previous increases:

$$\begin{aligned} \widetilde{\Lambda}_{i-1}^{max} &= \max\left(m_i^{fast}\left(t_{i-1} - t_i\right) + \Lambda_i^{min}, \Lambda_{i-1}^{min}\right) \\ \widetilde{m}_{i-1}^{fast} &= \min\left(m_{i-1}^{fast}, \widetilde{m}_{i-1}^{fast}\right) \end{aligned}$$

   which in turn can lead to further backward corrections.
4. Increase i by 1. When you reach $i > (n-1)$ you are done.
5. Determine the slowest rate at which the integrated hazard can grow within this interval. This supposes that the increase in the previous interval was linear. In principle, the slowest increase in hazard is a steady increase from the origin, should this solution be admissible. It will not be admissible if such an increase would lead to an integrated hazard below one of the lowest possible hazard below. Hence:

$$\alpha_i = \begin{cases} \alpha_{i-1} & \text{if } m_{i-1}^{slow}\left(t_i - \alpha_{i-1}\right) + \Lambda_i^{\alpha} > \Lambda_i^{min} \\ t_{i-1} & \text{otherwise} \end{cases}$$

$$\Lambda_i^{\alpha} = \begin{cases} \Lambda_{i-1}^{\alpha} & \text{if } m_{i-1}^{slow}\left(t_i - \alpha_{i-1}\right) + \Lambda_i^{\alpha} > \Lambda_i^{min} \\ \Lambda_{i-1}^{min} & \text{otherwise} \end{cases}$$

$$m_i^{slow} = \frac{\Lambda_i^{min} - \Lambda_i^{\alpha}}{t_i - \alpha_i}$$

6. (Forward correction) If needed, redefine the minimal integrated hazard at a given point $j > i$, which will be the value:

$$\widetilde{\Lambda}_j^{max} = \max\left(m_i^{slow}\left(t_j - t_i\right) + \Lambda_i^{min}, \Lambda_j^{min}\right)$$

   If $\widetilde{\Lambda}_j^{min} > \Lambda_j^{max}$, there is no admissible function.
7. Go to 3.

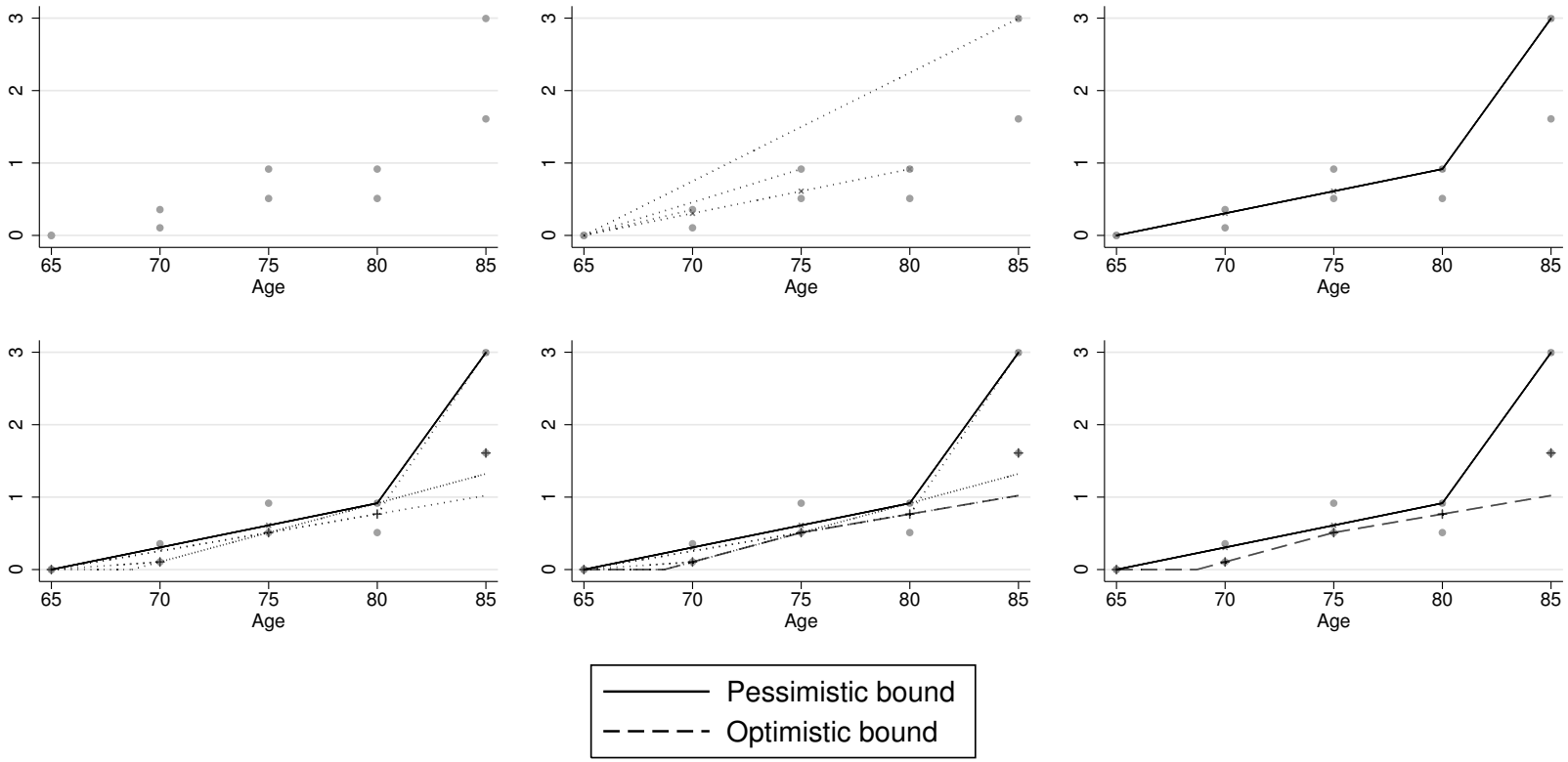


Figure 4.6: Non-parametric bounds under monotonicity and continuity: the case with rounding

Results 4.A.2

Only 35 percent of the sample report probabilities that are consistent with this assumption if we do not allow for rounding. Hence, if we maintain that the data correspond exactly to expectations, we have to conclude that the assumption of increasing death hazards does not fit subjective beliefs very well. This could be due to respondents believing in "critical ages", after which the risk of dying subsides. The picture changes once we allow for rounding: with a common rounding scheme 73 percent of observations are consistent and under general rounding this fraction increases further to 88 percent. We checked whether the tendency to report probabilities that are consistent with an increasing risk of death is related to socio-demographic covariates. For men we find that age is the single most important predictor and that the age pattern is strongly non-linear with a peak around 35 and trough at 65. Moreover, men who are in poor health are significantly less likely to have a monotonic death hazard, but this effect disappears once we allow for rounding. For women too we find that age matters: women around the age of 65 are less likely to hold monotonic expectations.

The bottom panel of Table 4.8 presents descriptive statistics of the bounds derived under the monotonic hazard assumption in combination with different forms of rounding (the other panels contain the same descriptives we reported in the main text and are included here only to aid comparisons). The bounds are approximately half as wide on average as those described in the middle panel: without rounding their width reduces from 12 to 6; with common rounding the reduction is from 15 to 9; and with general rounding from 19 to 11. This increase in precision is so large that the average interval under a monotonic hazard and the more conservative general rounding scheme is narrower than that which neither imposes any restrictions on expectations nor allows for rounding.

We analyze variations in life expectancy across the sample using this new set of bounds by means of the same partially identified models that we used in section 4.5.2. Despite the fact that the new bounds are considerably tighter, our estimates show that they are still not sufficiently informative to allow for inference with regard to differences in the expected age of death across

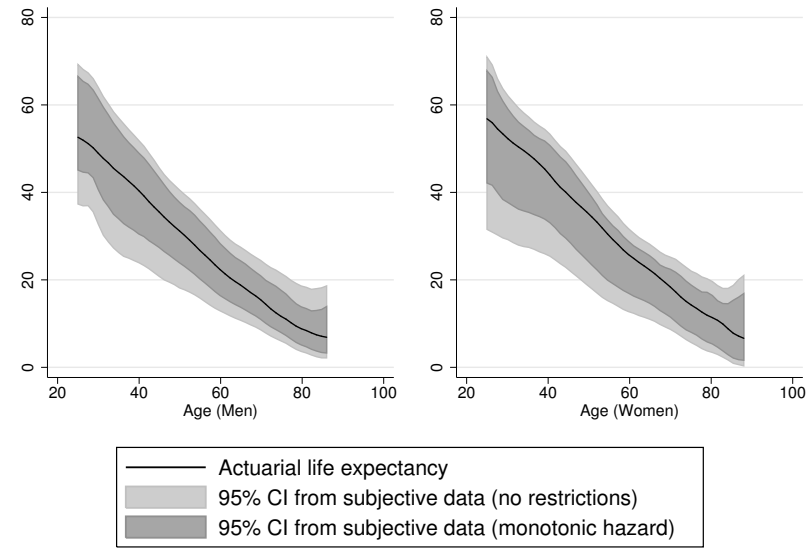
Table 4.8: Point estimates and bounds on life expectancy

	Men				Women			
	No rounding		Rounding		No rounding		Rounding	
	Mean	(S. D.)	Common	General	Mean	(S. D.)	Common	General
				Parametric survival functions				
Gompertz	81.71	(7.61)	-	-	81.99	(8.39)	-	-
Weibull	82.34	(7.65)	-	-	82.63	(8.38)	-	-
				Non-parametric bounds				
LB expected age death	76.43	(8.15)	75.03	73.45	75.66	(8.62)	73.99	72.20
UB expected age death	87.81	(8.75)	89.41	91.32	87.93	(8.90)	89.59	91.68
Width	11.37	(4.84)	14.39	17.87	12.27	(5.41)	15.60	19.48
Percentage decrease	77.88	(9.13)	72.13	65.31	78.09	(8.61)	72.27	65.15
N	780				601			
				Non-parametric bounds under increasing hazard assumption				
LB expected age death	79.88	(6.91)	78.05	77.42	80.36	(6.76)	77.64	76.85
UB expected age death	85.62	(8.37)	86.38	88.07	86.72	(8.52)	87.24	88.70
Width	5.74	(4.44)	8.34	10.65	6.36	(4.67)	9.60	11.86
Percentage decrease	88.92	(9.31)	84.38	79.72	88.78	(8.94)	83.47	79.21
N	288		569	679	194		438	539

socio-demographic categories, regardless of whether and how we allow for rounding.[7]

Finally, we investigate whether expectations and actuarial forecasts of life expectancy are on average consistent if we impose that the hazard of death increases with age. Figure 4.7 shows 95 percent confidence bands for life expectancy as a function of age both with and without the monotonicity assumption. The bands in Figure 4.7 allow for rounding of a general form. As expected, the monotonicity assumption reduces the width of the bounds at all ages. This reduction is mostly due to an upward shift in the lower bound at younger ages and to a downward shift of the upper bound at older age. Even when we impose monotonicity, we cannot reject that expectations are on average in line with actuarial forecasts. Robustness checks reveal that this result holds up regardless of the rounding rule.

[7]Estimates available on request.



Kernel regressions, bandwidth 3 years. Both sets of bounds allow for general rounding.

Figure 4.7: Non-parametric bounds on life expectancy with and without monotonic hazard assumption

4.B Descriptives of Bounds under General Rounding

Table 4.9: Sample averages of bounds on life expectancy derived under absence of rounding and general rounding

	Men											
			By age bracket									
	Overall		25-35		36-45		46-55		56-65		66+	
	No	General	No	General	No	General	No	General	No	General	No	General
No smoothing												
LB	76.43	73.45	72.46	66.24	70.93	66.85	73.31	69.76	77.12	74.24	81.55	79.96
UB	87.91	91.32	88.98	92.62	87.22	91.06	85.87	89.24	86.91	90.63	90.45	93.70
Width	11.37	17.87	16.52	26.39	16.29	24.20	12.56	19.48	9.80	16.39	8.90	13.74
Percentage decrease	77.88	65.31	78.80	66.12	76.42	64.94	78.90	67.20	80.13	66.74	75.22	62.29
N	780		36		104		185		233		222	
Smoothing using cubic splines												
LB	–	78.19	–	74.38	–	74.30	–	75.04	–	78.15	–	83.02
UB	–	85.34	–	85.37	–	84.53	–	82.88	–	85.03	–	87.95
Width	–	7.14	–	10.99	–	10.23	–	7.84	–	6.88	–	4.93
Percentage decrease	–	86.35	–	85.88	–	85.14	–	86.80	–	86.01	–	86.92
N	611		29		76		144		182		180	

	Women											
			By age bracket									
	Overall		25-35		36-45		46-55		56-65		66+	
	No	General	No	General	No	General	No	General	No	General	No	General
No smoothing												
LB	75.66	72.20	70.08	63.98	72.66	68.15	73.91	69.85	76.83	74.22	81.52	79.78
UB	87.93	91.68	88.49	91.95	88.77	92.18	86.66	91.06	86.40	90.41	90.65	93.71
Width	12.27	19.48	18.41	27.97	16.12	24.03	12.74	21.21	9.57	16.18	9.13	13.93
Relative width	78.09	65.15	76.42	64.20	76.87	65.50	78.75	64.60	80.59	67.15	75.51	62.95
N	601		61		107		137		178		118	
Smoothing using cubic splines												
LB	–	77.44	–	71.24	–	75.32	–	75.44	–	78.36	–	82.99
UB	–	85.51	–	83.89	–	84.92	–	84.95	–	85.18	–	87.96
Width	–	8.07	–	12.65	–	9.60	–	9.51	–	6.82	–	4.97
Percentage decrease	–	85.65	–	83.80	–	86.23	–	84.07	–	86.16	–	87.03
N	463		44		76		104		148		91	

[a] Percentage decrease in width of the interval for life expectancy relative to the uninformative interval. The uninformative interval of respondent i is equal to $110 - age_i$.

Point and partially identified models using linear splines 4.C

Table 4.10: Point and partially identified models of the remaining life expectancy

			Partially identified models					
			No smoothing		Linear splines			
	Point identified models		No rounding		Common rounding		General rounding	
	Gompertz	Linear spline	LB	UB	LB	UB	LB	UB
Age	-1.677***	-1.501***	-3.278	2.256	-2.195	-0.787	-2.921	-0.0381
	(0.105)	(0.109)	(-3.526; 2.504)		(-2.407; -0.575)**		(-3.117; 0.157)	
Age squared/100	0.720***	0.648***	-1.762	3.079	0.0278	1.248	-0.629	1.888
	(0.0882)	(0.0930)	(-1.981; 3.298)		(-0.149; 1.425)		(-0.796; 2.055)	
Male	-0.851**	-0.692	-12.571	11.185	-3.769	2.456	-7.192	5.862
	(0.428)	(0.423)	(-13.423; 12.037)		(-4.580; 3.266)		(-7.971; 6.641)	
Educ. - primary school	-1.200	-0.712	-13.151	11.713	-3.655	2.009	-7.348	5.642
	(0.929)	(0.995)	(-14.734; 13.297)		(-5.474; 3.828)		(-9.090; 7.384)	
Educ. - higher secondary	-0.268	-0.116	-12.898	12.681	-3.526	2.948	-7.353	6.630
	(0.730)	(0.746)	(-14.184; 13.968)		(-4.915; 4.337)		(-8.705; 7.983)	
Educ. - lower vocational	0.191	-0.120	-14.136	13.911	-4.115	3.657	-8.186	7.822
	(0.704)	(0.670)	(-15.540; 15.314)		(-5.382; 4.924)		(-9.437; 9.074)	
Educ. - higher vocational	-0.614	-0.477	-13.706	12.771	-3.926	2.810	-7.882	6.751
	(0.564)	(0.573)	(-14.861; 13.926)		(-4.997; 3.880)		(-8.899; 7.769)	
Educ. - university	0.405	0.211	-14.281	14.712	-3.726	3.793	-7.766	8.004
	(0.670)	(0.646)	(-15.609; 16.040)		(-4.910; 4.976)		(-8.982; 9.220)	
Income < 1150 euro/month	-0.389	-0.495	-15.306	14.327	-4.488	3.468	-8.497	7.707
	(1.144)	(1.037)	(-17.247; 16.268)		(-6.233; 5.213)		(-10.394; 9.604)	
Income 1801-2600 euro/month	-0.411	-0.00699	-12.737	12.744	-3.522	3.257	-7.163	6.694
	(0.701)	(0.678)	(-13.908; 13.915)		(-4.811; 4.545)		(-8.296; 7.827)	
Income > 2600 euro/month	-0.497	-0.0403	-13.122	13.059	-3.498	3.228	-7.287	6.917
	(0.664)	(0.636)	(-14.461; 14.398)		(-4.675; 4.405)		(-8.321; 7.951)	
Health - bad	-6.614***	-7.439***	-23.717	8.920	-11.216	-2.465	-15.027	2.553
	(1.038)	(1.042)	(-25.682; 10.886)		(-13.069; -0.612)**		(-16.911; 4.438)	
Health - fair	-5.443***	-6.285***	-21.302	8.746	-9.987	-2.294	-13.590	2.241
	(0.758)	(0.762)	(-22.797; 10.242)		(-11.326; -0.955)**		(-14.895; 3.546)	
Health - good	-1.327**	-1.999***	-15.403	11.416	-5.492	1.645	-8.790	5.241
	(0.658)	(0.656)	(-16.734; 12.746)		(-6.757; 2.910)		(-9.923; 6.375)	
Constant	99.20***	92.07***	14.901	169.661	72.196	111.488	51.036	131.610
	(3.221)	(3.230)	(7.710; 176.852)**		(66.036; 117.648)**		(45.568; 137.078)	
N	1,380	1,380	1,380		1,380		1,380	
R-squared	0.730	0.691	–		–		–	

Standard errors in parentheses for point identified models
95% confidence sets in parentheses for the partially identified models
***significant at 1%; **significant at 5%; *significant at 10%

5 Can the Dutch Meet Their Own Retirement Expenditure Goals?

This chapter is based on De Bresser and Knoef (2013).

5.1 Introduction

The question whether people save enough for retirement is not a new one. Most research on retirement preparedness has focused on the US, because the US pension system places responsibility for securing an income after retirement with the individual. In the absence of generous, universal public pensions one naturally worries about savings decisions and their implications for eventual retirement income. Pensions in the Netherlands, on the other hand, cover almost the entire population and have traditionally succeeded to ensure an adequate income during retirement. As in most developed countries, however, Dutch pensions are not immune to the combined forces of population aging and weak financial market performance. Maintaining sustainability of the system will necessitate a combination of raising the pension eligibility age[1], lower or no price indexation of pensions, and perhaps even cuts in pension payments. Against this backdrop, this paper investigate the retirement readiness of the Dutch in January 2008, at the eve of the downturn in the financial markets. Our aim is to describe whether the Dutch were sufficiently prepared according

[1] Until 2013 the statutory retirement age in the Netherlands was 65. As of January 2013 the statutory retirement age increased by one month, and will gradually increase to 66 in 2019 and 67 in 2023.

to their own standards, to identify vulnerable groups, and to examine the consequences of disappointing pensions and decreasing housing prices.

Our approach differs from previous efforts in that we adopt as our yardstick for savings sufficiency self-reported measures of the minimal and adequate level of expenditures during retirement. The rationale for this approach is that preferences and constraints are likely to vary across individuals and households. Measuring readiness against a single universal threshold fails to capture relevant differences in coping strategies. Moreover, allowing required expenditures to vary between individuals provides insight into how goals and resources interact and which people are financially prepared for retirement. Simultaneously analyzing both aims and means yields new policy implications. On the one hand there are groups that have a modest expected retirement income but also low perceived needs. These people will not change their saving behavior when confronted with a realistic assessment of their financial position. On the other hand, financial information can motivate saving behavior of groups with high expected retirement incomes who also have high perceived needs. Our focus on attaining consumption goals after retirement means that we do not take into account other reasons to save, such as precautionary or bequest motives. If such additional rationales exist, our analysis should be interpreted as an upper bound on preparedness.

Another distinguishing feature of this paper is our combined use of survey and administrative data. For the subjective assessments of minimal and adequate expenditure levels during retirement we draw survey data from a representative sample of the Dutch population. Tax records and data from pension funds, on the other hand, allow us to construct a complete and precise measure of the resources available to households.

We find that in the aggregate the Dutch are well prepared for retirement. The median difference between the after-tax annuity that can be obtained at age 65 and the individual-specific level of minimal expenditures is 27% if we ignore all non-pension wealth, and the median difference relative to adequate expenditures is 6%. Still, there is a sizable minority of close to 20 percent of the sample for whom the annuity falls short of minimum expenditures even if we include housing wealth. The size of those deficits is large enough to be problematic, with a median shortfall of around 30% (even if we include housing wealth). Multivariate analysis reveals that variation in needs and

desires interacts with accumulated resources to produce interesting patterns. For instance, we find that the highly educated both accumulate more wealth and are more demanding in terms of their minimal retirement income. As a result they are more likely to reach their goals only if we control for the level of their needs. The self-employed and the divorced also stand out as vulnerable groups.

The remainder of the paper is organized as follows. The next section provides a short overview of pension arrangements in the Netherlands. Section 5.3 presents a literature review on retirement readiness. Section 5.4 describes the sources from which we gather survey data on expenditure needs during retirement and administrative data on the resources available to meet them. Descriptive statistics of consumption floors and resources are provided in section 5.5. Our methodological approach for simulating retirement readiness is explained in section 5.6. Section 5.7 combines the data on resources and consumption floors and analyzes who can and cannot look forward to comfortable retirement. The final section concludes.

The Dutch pension system 5.2

As in many European countries the Dutch pension system consists of four pillars, all of which will be taken into account in this paper. The first pillar is a flat-rate public pensions for all residents financed by a pay-as-you-go scheme. This public pension aims to provide retirees with a subsistence income during retirement. Its level is set in relation to the minimum wage[2] and depends only on the number of years spent abroad during the accumulation period (payments are cut with 2 percent for each year spent abroad between age 15 and 65). For people with a low pension income and almost no wealth, the first pillar is topped up with social assistance to guarantee a social minimum. The second pillar is that of occupational pensions that cover 90 percent of Dutch workers (Bovenberg and Meijdam 2001). The level of occupational pensions depends on the average or final wage of the individual worker throughout the accumulation phase. Though occupational pensions are mostly defined benefit,

[2]Single pensioners who have lived in the Netherlands between the ages of 15 and 65 receive 70% of the minimum wage. Couples receive 100% of the minimum wage.

the possibilities of non-indexation and pension cuts introduce uncertainty in payments. Together the first two pillars of the pension system aim to replace 70 percent of the final or average wage. The third pillar offers saving vehicles such as life annuities. They are fiscally attractive for and especially used by the self-employed and those who have not accumulated (enough) occupational pensions. In contrast to the first two pillars, third pillar pensions are voluntary and usually of the defined contribution type. The fourth pillar contains all other assets that individuals may decumulate to generate income during retirement, such as savings accounts and housing wealth.

5.3 Literature

This paper compares available resources with self-reported minimal and adequate retirement expenditures to assess whether the Dutch are ready to meet their expenditure goals after retirement. It fits in with the large literature on retirement savings adequacy, which has focused mostly on the US. In the US responsibility for maintaining one's standard of living after retirement has long been allocated primarily to the individual, with social security replacing 40% of final income on average (Beshears et al. 2009). Research in the 1990s found that the introduction and growing importance of defined contribution personal savings accounts had lead to large increases in overall wealth available for retirement (Venti and Wise 1997). However, in an influential article Banks et al. (1998) show that consumption drops around retirement, a finding that goes against the consumption smoothing implications of lifecycle models and that cannot be explained fully by changes in labor supply. Instead, it suggests that households are surprised to find their savings inadequate to maintain consumption. Similarly, Mitchell and Moore (1998) warn that the median US household at the verge of retirement has accumulated insufficient funds to sustain consumption close to the pre-retirement level for another 20+ years. Excluding housing wealth, Skinner (2007) argues that rising out of pocket medical expenses threaten even the affluence of households with post-graduate degrees. Engen et al. (1999) reach a more comforting conclusion. Comparing observed wealth data with optimal behavior in a lifecycle model, they show that more than half of their survey respondents have wealth-earnings ratios

at or above the median optimal ratio for their socioeconomic characteristics. Furthermore, their simulation model underestimates the actual wealth among households with high ratios of wealth to earnings, suggesting that wealth accumulation is adequate for a majority of households. Scholz et al. (2006) compare optimal savings from a lifecycle model with household-specific wealth profiles. They find that their model explains 80 percent of the variation in wealth holdings, that fewer than 20 percent of the households have below optimal wealth and that wealth deficits are small.

In any case, Americans are not convinced that they will be able to afford retirement. Yakoboski and Dickemper (1997) document that while 69% of workers set aside some money for retirement, only 25 percent are very confident that their savings will allow them to live comfortably throughout retirement. Such worries persist into retirement, since 23% of retirees are not confident that their savings will allow comfortable living until death. With regard to maintaining savings adequacy after retirement, Haveman et al. (2007) find that a stable fraction of 34 percent cannot meet their own pre-retirement consumption levels and 5 percent does not meet the official US poverty standard during the first decade of retirement. However, at the household level large fluctuations do occur.

Brugiavini and Padula (2001) look at saving in Italy and provide interesting insight into the differences between the US and Europe. In Italy, as in the Netherlands, mandatory contributions to the welfare state account for a large fraction of savings. In return, severance pay and social security are generous, the latter replaces 60-70 percent of pre-retirement after tax income, so there is little need for additional saving. Despite that institutional framework, the authors find no discretionary dissaving at any age. The only decumulation of capital that does occur takes the form of drawing a public pension. Other authors have focused specifically on the Netherlands. Alessie et al. (1997a) look at the effect of social security and pensions on private savings and find no significant effect for pension wealth but some, less than perfect, replacement of private savings by social security. In another study, Alessie et al. (1997b) document that a large fraction of Dutch households do not dissave during retirement, perhaps due to bequest motives. The study that is closest to ours in terms of focus and data is that by Knoef et al. (2013). They use a much larger sample of the same administrative dataset on assets that we use and provide a

detailed description of the wealth holdings in the Dutch population. However, they do not have access to survey data on desired retirement expenditures, so they cannot evaluate the sufficiency of savings using that reference point.

Previous papers mostly use one of three definitions for adequate retirement savings: ability to maintain pre-retirement consumption (or a fraction thereof); ability to meet some official poverty line; or ability to meet wealth holdings predicted by a lifecycle model. The literature shows that the choice of benchmark against which to measure retirement readiness is not without consequences. For the US, about a third of the retiring households may not be able to consume as much during retirement as they did while still working (Haveman et al. 2007). However, that need not be problematic, since consumption needs are likely to decline when individuals retire. Indeed, a much smaller fraction will drop below the poverty line, but income at the poverty threshold will not be satisfactory for individuals used to high consumption levels. If optimal savings are derived from lifecycle models, the picture changes to one in which households are saving adequately. However, lifecycle models may not accurately reflect the decision process and preferences of real individual households. Our approach to retirement readiness compares annuitized wealth with minimal and adequate expenditures reported by survey respondents. One important advantage of this method is that it allows consumption needs to differ at the level of the individual household, depending on preferences and constraints that are likely to be household-specific. In the survey respondents are encouraged to think about options for reducing their expenditures when evaluating their minimal needs (more details can be found in section 5.5.1).

5.4 Data

5.4.1 Data sources

As explained in the introduction, we combine survey data on expenditures during retirement with tax data on assets to investigate whether the Dutch are sufficiently prepared to meet their own goals. Survey data are taken from the LISS panel (Longitudinal Internet Study in the Social Sciences), gathered

by CentERdata.[3] This panel is recruited through address based sampling (no self-selection), and households without a computer and/or internet connection receive an internet connection and computer for free. This roughly nationally representative household panel (Van der Laan 2009) receives online questionnaires each month, on different topics. When respondents complete a questionnaire they receive a monthly incentive. A variety of data is available from studies conducted in the LISS panel.

In this paper we use a single wave study on minimal and adequate pension expenditures. These data were elicited from LISS-respondents in January 2008 on the initiative of Johannes Binswanger and Daniel Schunk (see Binswanger and Schunk 2012, for their analysis of a similar questionnaire distributed to the CentERpanel and Binswanger et al. 2013, for an analysis of panel conditioning that uses the same data we use).

Administrative data are taken from the 2008 Complete Asset data of the Netherlands (CAD, Integraal-Vermogensbestand, CBS 2008a), the 2008 Public Pension Entitlements data (PPE, Pensioenaansprakenstatistiek, CBS 2008b), the 2008 Occupational Pension Entitlements data (OPE, Pensioenaansprakenstatistiek, CBS 2008c), the 2008 Public Pension Benefits data (PUBLB, AOW-jaarbedrag, CBS 2008d) and the 2008 Private Pension Benefits data (PRIVB, Pensovjaarbedrag, CBS 2009e) all gathered by Statistics Netherlands.

The CAD consist of all households in the Netherlands and contains data about financial wealth (savings accounts, stocks, and securities), property, business wealth [4] and debt. Debt is categorized in mortgage and other debt. Although most of these data are derived from tax records, banks also provide information about bank accounts. Banks only have to report accounts with a balance of 500 euro or more (or 15 euro in interest payments), which means that we miss small amounts of money held in bank accounts.

PPE and OPE contain information about public and occupational pension entitlements for the whole Dutch population between the ages of 21 and 64. PPE is based on data from the organization that implements national insurance schemes in the Netherlands (SVB) and OPE is based on data from pension

[3] For more information, see http://www.lissdata.nl/lissdata/.

[4] In this way we also take into account the so called 'fiscal old age reserve' for the self-employed. The fiscal old age reserve is a latent tax liability that allows the self-employed to set aside part of their profits to provide additional income during retirement.

funds. Finally, public and private pension benefits received by all retirees in the Netherlands are available in PUBLB and PRIVB.

Third pillar pensions (e.g. life annuities) are, unfortunately, only observed in administrative data once they are claimed, because they are subject to taxation only in the payout phase. Third pillar pensions are likely to be important for those groups that cannot rely on occupational pensions. Therefore, we need to take them into account if we are to credibly assess retirement preparedness. In order to construct a complete picture of the resources available to each household, we supplement the administrative data of pre-retirees with information on third pillar pension entitlements from the 2008 wave of the LISS Assets Survey.

In the remainder of this section we describe the sample (5.4.2) and show descriptives of self-perceived minimal and adequate retirement expenditures and wealth. Descriptives of retirement planning can be found in Appendix 5.B.

5.4.2 Sample selection

In January 2008, 2,405 LISS panel-members received the questionnaire on minimal and adequate expenditure levels during retirement. Panel members that received the questionnaire were household heads and their spouses, with a reported net monthly income higher than 800 euros[5] and age greater than or equal to 25. Out of these 2,405 individuals, 2,005 answered at least 1 question (83% survey response). However, a much smaller number of 1,300-1,500 respondents, 54-62% of the potential sample, answered the questions about expenditures. If the underlying selection process is related to consumption needs, such a large fraction of item non-response is problematic.

To link administrative data to the survey data, opt-out consent was obtained for all panel members in September 2011. Unfortunately, 1,292 (54%) of the potential sample of 2,405 respondents were not participating in the panel anymore in September 2011. For these people we have no consent and because of ethical considerations we could not link their survey answers to the administrative data. In addition, (only) 134 of these panel members objected

[5] In this way students, for example, are excluded.

against linkage and administrative records could be retrieved for the remaining 1,158 individuals. Figure 5.1 presents a graphical representation of the sample.

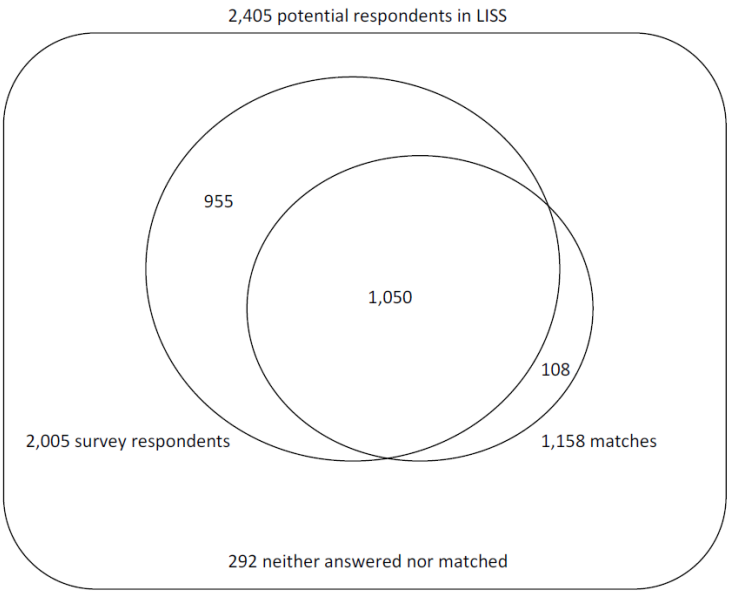


Figure 5.1: Survey response and merge with administrative records

Table 5.1 describes the three samples: the potential sample, the survey respondents and the sample that we could matched with administrative records. It is reassuring that the descriptives of the observed characteristics of the three samples are about the same. However, net personal income is relatively low in the matched sample. Our goal is to analyze the retirement readiness of a representative sample of the Dutch population. Section 5.6 describes how we achieve this goal.

Variable definitions and descriptive statistics 5.5

Retirement expenditures 5.5.1

The yardstick against which we measure retirement savings adequacy is the respondent's own subjective evaluation of what would be a minimal and adequate level of expenditure during retirement. After some questions about

Table 5.1: Descriptive statistics

	Potential sample[a]		Survey respondents[b]	Matched[c]
	Mean	SD	Mean	Mean
			Demographics	
HH head	0.58	0.49	0.58	0.60
Male	0.48	0.50	0.48	0.49
Birth year	1958	13.12	1958	1958
# children	0.92	1.11	0.87	0.91
Homeowner	0.77	0.42	0.78	0.75
Lives with partner	0.83	0.37	0.83	0.82
Married[d]	0.71	0.45	0.71	0.70
Separated/divorced	0.08	0.28	0.08	0.10
Widowed	0.03	0.17	0.03	0.03
Never married	0.18	0.38	0.17	0.17
			Education	
Primary[d]	0.09	0.29	0.09	0.10
Intermediate secondary	0.26	0.44	0.26	0.24
Higher secondary	0.08	0.27	0.08	0.09
Intermediate vocational	0.25	0.43	0.25	0.26
Higher vocational	0.23	0.42	0.24	0.24
University	0.08	0.28	0.08	0.07
			Primary activity	
Employed[d]	0.58	0.49	0.57	0.58
Self-employed	0.08	0.28	0.07	0.07
HH work	0.12	0.32	0.12	0.11
Retired	0.15	0.36	0.17	0.17
Disabled	0.03	0.17	0.03	0.04
Other	0.04	0.19	0.04	0.03
Net personal income[e]	1818	6912	1862	1644
Has simPC	0.04	0.20	0.04	0.06
N	2,405		2,005	1,158

[a] The potential sample contains all LISS panel-members that received the questionnaire on minimal and adequate expenditures during retirement.
[b] The sample of survey respondents contains all respondents to the questionnaire on expenditures during retirement, regardless of item (non-)response.
[c] The matched sample contains all panel members of the potential sample that could be matched with administrative records.
[d] Baseline for categorical variables.
[e] For income the sample sizes are 2381, 1988 and 1151.

retirement planning and housing expenditures, subjective minimal levels of household expenditures were elicited by means of an open-ended question:

> This question refers to the overall level of spending that applies to you [*and your partner/spouse*] during *retirement*. What is the minimal level of monthly spending that you would never want to fall below during retirement, at all costs? Please think of all your expenditures, such as food, clothing, housing, insurance etc. Remember, please assume that prices of the things you spend your money on remain the same in the future as today (i.e. no inflation).
>
> ... per month
>
> don't know

The phrasing of this question helped respondents to keep in mind an inclusive view of their monthly budget by emphasizing the wide variety of expenditures that need to be covered. Housing is especially important in this respect, since the primary residence often is the most important discretionary asset held around retirement. The survey primes respondents to take this into account by first presenting questions on housing before moving on to retirement expenditures. In the actual question the importance of housing is emphasized again.

In contrast to minimal expenditures, adequate expenditures during retirement are anchored on current net household income in order to ensure that the answers are meaningful for the respondents. Moreover, instead of answering a single open-ended question, respondents are guided to their answer by means of 1 or 2 multiple choice questions. In each question the respondent is shown scenarios that consist of a certain level of expenditures during working life and during retirement, with replacement rates of 50, 64, 76, 88, 100, and 140 percent (the scenarios are roughly actuarially neutral). For example, the following questions were asked to a respondent with a household income of 3,500 euro per month[6]:

> [*Please assume that you are not retired yet*]

[6]Consumption levels are chosen such that monthly consumption during working life in option D do not exceed current income. For retirees current income is divided by 0.85, to take into account the income drop due to retirement.

Next you will find four options for how you could spend your money over your lifetime. For each option the first column indicates how much [*you* (if respondent has no spouse/partner) / *your household* (if respondent has a spouse/partner)] could spend on average per month from age 25 until retirement. Thus, this refers to your total (working) time from age 25 until retirement, [add only if not retired: *NOT just the remaining (working) time*]. The second column indicates how much [*you* (if respondent has no partner)/ *your household* (if respondent has a partner)] could spend during retirement. Please think of all your expenditures, such as food, clothing, housing, insurance, traveling etc. Assume that the numbers below show what you can spend after having already paid for taxes. Assume also that prices of the things you spend your money on remain the same in the future as today (i.e., no inflation). If you had a choice, which option would you like most?

a. 3,000 during working life; 3,000 during retirement

b. 3,150 during working life; 2,750 during retirement

c. 3,300 during working life; 2,500 during retirement

d. 3,450 during working life; 2,200 during retirement

don't know

If the respondent indicates that she would prefer one of the extreme options (a. or d.), a second question is asked in order to give respondents more choice without confusing them with too many options at once:

[*If the answer is a.:*]

You chose option A [*3,000*] euro during working life and [*3,000*] euro during retirement. If you had a choice between this option and a further new option (Z, see the table below), which one would you choose?

a. 3,000 during working life; 3,000 during retirement

z. 2,650 during working life; 3,700 during retirement

don't know

[*If the answer is d.:*]

You chose option D [*3,450*] euro during working life and [*2,200*] euro during retirement. If you had a choice between this option and a further new option (Z, see the table below), which one would you choose?

d. 3,450 during working life; 2,200 during retirement

z. 3,650 during working life; 1,800 during retirement

don't know

The questionnaire includes two such sets of questions, one that assumes a real interest rate of 1% and another that assumes a 6% real interest rate. In this paper we only analyze adequate expenditures based on 1% real interest, because that scenario is closest to the current market conditions. Furthermore, the order of the scenarios is randomized across respondents: half see expenditures during working life in ascending order (as shown above) and the other half in descending order. Binswanger and Schunk (2012) provide a more detailed description of the questions. In this paper we standardize expenditures to single person households using the equivalence scale constructed by Statistics Netherlands (Siermann et al. 2004). This equivalence scale assumes that a couple without children needs 37% more income than a single person household.

The top panel of Table 5.2 contains descriptive statistics for minimal expenditures during retirement. The question is answered by 1,483 respondents (out of the 2,405 respondents that received the questionnaire). On average the respondents would like to spend at least 1,542 euro per month during their retirement. The median respondent would like to spend at least 1,460 euro per month. This is about 50% more than the public state pension, which aims to provide a basic subsistence level of income to all retired residents. In Table 5.2 we also divide minimal individual assessed expenditures by current income. We find an average and median ratio of 74 percent. This is in line with the 70 percent replacement rate that is often mentioned as reasonable by financial planners and it is in line with the average replacement rate provided by the first two pillars of the Dutch pension system (Bovenberg and Meijdam 2001).

Self perceived adequate expenditures during retirement are somewhat higher, with a median of 1,606 euros per month. Across different age groups,

we find that minimum expenditures are constant but median adequate expenditures are higher for middle aged respondents. The distribution of the underlying self assessed adequate replacement rates, however, is about the same for all age categories. Most people would like to have a replacement rate of 76 or 88%. However, a substantial percentage of 12% prefer a replacement rate of 100% (and are willing to sacrifice consumption during working life for that).

Table 5.2: Descriptive statistics of minimum expenditures during retirement and adequate replacement rates

	All respondents						By age				
							25-34	35-44	45-54	55-64	65+
	N	Mean	p25	Mdn	p75	SD	Mdn	Mdn	Mdn	Mdn	Mdn
Self-assessed minimum retirement expenditures											
Minimum exp.[a]	1,483	1,542	1,095	1,460	1,825	781	1,357	1,460	1,460	1,460	1,460
Min. exp. / current income	1,483	74.3	55.9	73.6	90.2	36.1	74.5	72.5	66.8	77.1	80.9
Self-assessed adequate retirement expenditures											
Adequate exp.[a,b]	1,289	1,889	1,277	1,606	2,044	3,725	1,460	1,569	1,715	1,642	1,606
Adeq. exp. / current income	1,289	78.1	71.2	78.6	83.9	10.7	77.8	77.8	78.1	79.2	84.5
Distribution of self-assessed adequate replacement rates											
Adequate RR 50%[b]	1,289	0.03	0	0	0	0.17	0.04	0.04	0.03	0.02	0.02
RR 64%	1,289	0.06	0	0	0	0.24	0.05	0.07	0.05	0.05	0.09
RR 76%	1,289	0.39	0	0	1	0.49	0.38	0.38	0.40	0.38	0.39
RR 88%	1,289	0.37	0	0	1	0.48	0.37	0.37	0.39	0.39	0.33
RR 100%	1,289	0.12	0	0	0	0.32	0.13	0.11	0.11	0.12	0.13
RR 140%	1,289	0.03	0	0	0	0.17	0.03	0.02	0.03	0.03	0.04

[a] Retirement expenditures are equivalized to 1-person household without children.
[b] Adequate expenditures are elicited using a 1% real interest rate.

The standard deviations of minimum and adequate expenditures show the added value of our subjective benchmarks compared to one-size-fits-all approaches (such as poverty lines or fixed replacement rates for everyone).

Figure 5.2 shows kernel regressions of minimal and adequate expenditures during retirement on income and age. The left panel contains the regressions on standardized household income and includes a dashed line corresponding to a 70% replacement rate with regard to current income. Both expenditure levels increase with income but the overall gradient is steeper for adequate expenditures than for minimal expenditures. Adequate expenditures closely follow the 70% line, suggesting that a replacement rate of 70% is a good approximation for the average desired consumption level. Minimum expenditures are larger than adequate expenditures only for respondents with monthly net household income below 1,700 Euro, perhaps because the latter are anchored on current income while the former are not. Within-respondent comparisons of minimum

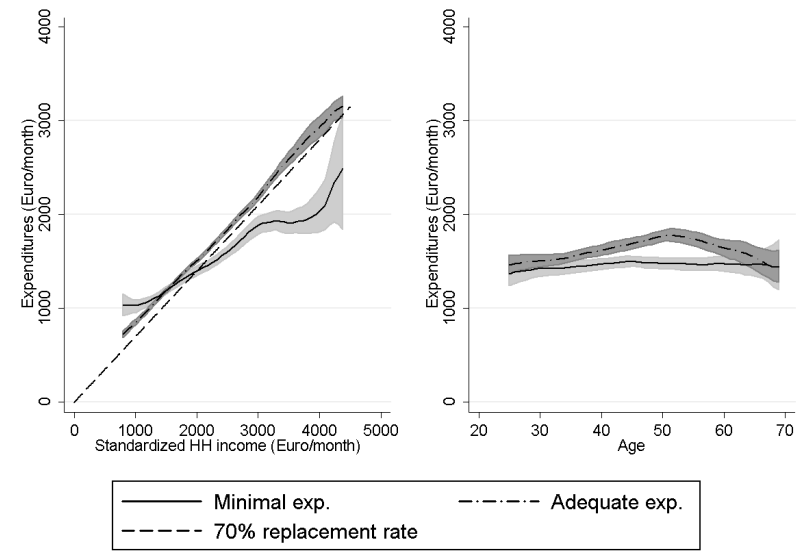


Figure 5.2: Kernel regressions of minimal and adequate expenditures during retirement on income and age (consumption floors and income are standardized to 1-person household)

and adequate expenditures reveal that 58 percent of the respondents to both questions report minimum expenditures below adequate expenditures. Since adequate expenditures are anchored to current income, people with higher minimal expenditures than adequate expenditures probably do not regard their current expenditures as sufficient for retirement.

The right panel of Figure 5.2 shows that minimal expenditures are constant across ages, at a level close to 1,400 Euro/month. Adequate expenditures follow a hump-shape, with a peak at 1,700 Euro/month around the age of 50. Table 2 shows us that this hump in adequate expenditures is due to income differences, since self assessed adequate replacement rates are constant across age groups. The level of adequate expenditures around the age of 25 and 70 is very similar to that of minimal expenditures.

Respondents are likely to find the questions on expenditures during retirement challenging. For instance, younger respondents have not thought much about retirement: of the respondents aged 25-34, 86% report having thought "only a little" or "not at all" about retirement. For the pre-retirees, age 55-64, this fraction is 51%. In addition to questionable salience of retirement to young

respondents, 63% of individuals find the questions difficult to answer and 35% find it difficult to assess their expenditure needs during retirement. As a result, our measures of expenditures may be noisy. In our models we analyze expenditures as dependent variables, so random measurement error will not bias our coefficient estimates or our simulations. However, if poor-quality answers tend to be systematically higher or lower than others, our simulations of the extent to which the Dutch can meet their retirement goals will be affected. The survey items on question salience and difficulty allow us to control for these factors. Descriptives of these questions can be found in Appendix 5.B.

5.5.2 Assets

The quality of any evaluation of retirement readiness depends on the analyst's ability to measure financial resources. Survey reports of assets are known to suffer from substantial non-response and under-reporting, in particular when it comes to ownership of categories like stocks and savings accounts (Bound et al. 2001, Johansson and Klevmarken 2007).[7] For these reasons we prefer to draw on more reliable administrative sources.

Table 5.3 shows descriptives of assets at the household level, for all panel members that received the questionnaire about pension expenditures and could be matched to the administrative data. We observe large differences in ownership rates across the various categories: savings accounts are held by almost every household in the sample while non-residential real estate and business wealth are only held by 5-8 percent. Apart from pensions, the most important asset for most households is their house: 77 percent of the households own real estate and their median net housing wealth is equal to 142,860 euro). Actually, this is an underestimation, since about 30% of the mortgages in the Netherlands are endowment- or investment-based mortgages (Dijkhuizen et al. 2013). For these mortgages payments are not used to repay the debt directly, but instead are paid to the endowment insurance company or saved on separate account

[7]Non-response is common in the LISS questionnaires on assets. Moreover, a comparison of administrative records with survey data reveals that many respondents wrongly indicate that they do not own savings accounts and investments (ownership rates are 10 percentage points lower in the survey compared to tax records). Home ownership, on the other hand, is reported accurately. Among those who indicate ownership, we find that the balance of savings accounts, the value of the primary residence and that of the corresponding mortgage are understated.

for fiscal reasons. The entire debt is then paid off at once at the end of the mortgage contract. Since the savings held in such mortgage-specific accounts are not taxed, they are not available in our administrative data. Unfortunately, there is no additional survey information about them in the LISS panel. Hence, we underestimate retirement preparedness for those who own these kind of mortgages. 10% of the households have non-mortgage debt. With a median of 22,442 euro (conditional on ownership) the size of those debts may be problematic for the households who own them.

Table 5.3: Descriptive statistics of household assets and pension entitlements in 2008.

			Conditional on ownership		Unconditional		
	N (HHs)	Ownership	Mean	Median	Mean	Median	SD
Financial assets							
Saving accounts	955	0.94	39,939	19,223	37,716	17,629	56,255
Stocks and bonds	955	0.34	64,629	16,976	22,130	0	120,921
Business assets	955	0.06	45,893	15,118	2,226	0	29,467
Other assets	955	0.04	165,887	8,000	6,774	0	156,978
Debt (other than mortgage)	955	0.10	50,424	22,442	5,174	0	28,651
Net non-housing wealth	955	0.95	66,747	22,751	63,672	20,855	212,962
Housing wealth							
Residential real estate	955	0.76	295,442	253,973	225,216	222,227	197,853
Non-residential real estate	955	0.08	197,488	165,000	15,509	0	74,444
Mortgage debt	955	0.69	152,716	138,856	106,021	80,294	112,851
Net housing wealth	955	0.77	174,077	142,860	134,705	85,720	197,051
Mortgage/property			0.49	0.44	0.49	0.44	0.36
Gross pension entitlements (standardized annuities)							
Public pension	976	1.00	1,018	1,040	1,018	1,040	74
Private pensions[a]	976	0.98	1,285	1,122	1,264	1,111	1,057
Share of private pension			0.49	0.52	0.48	0.52	0.19
Net standardized annuities							
Pensions	976	1.00			1,838	1,749	686
Percentage total					0.78	0.79	0.21
Pensions + wealth	955	1.00			2,089	1,917	1,058
Percentage total					0.85	0.86	0.18
Pensions + wealth + housing	955	1.00			2,616	2,328	1,567

[a] Private pensions are the sum of forecasted occupational pensions and self-reported third pillar pensions. For current claimers private pensions are the sum of all non-public pensions received.

Table 5.4 shows asset holdings, debt and pension entitlements for different age groups in our cross-section (households are grouped based on the age of the head of the household). Up to age 55-64 we find that older age groups hold more financial assets of all types compared to younger respondents, but the 65+ subsample is slightly less wealthy (these are combinations of age and cohort effects). The incidence of homeownership is roughly the same for the different cohorts, around 75-81 percent, except for the oldest respondents of

whom only 55 percent own a house. The median value of property conditional on ownership increases with age, from 214,857 euro in the 25-34 category to 325,403 for those respondents of age 65 or older. We find that the oldest group has only a small remaining mortgage debt: conditional on having a mortgage, the value of the remaining mortgage relative to the value of the house declines from a median of 92 percent for respondents below age 35 to only 13 percent for the oldest respondents. Not only have older households paid off a larger part of their mortgage, but they also benefited from increasing housing prices which have decreased their loan to value.

Table 5.4: Assets for different age groups (ownership rates and median amounts conditional on ownership)

	By age group											
	25-34		35-44		45-54		55-64		65+		Total	
	Owner	Mdn	Owner	Mdn	Owner	Mdn	Owner	Mdn	Owner	Mdn	Owner	Mdn
Financial assets												
Saving accounts	0.91	12,033	0.93	16,206	0.96	20,519	0.95	25,228	0.98	23,395	0.94	19,223
Stocks and bonds	0.20	6,075	0.34	14,911	0.38	19,575	0.36	21,666	0.38	26,339	0.34	16,976
Business assets	0.07	14,951	0.08	19,081	0.08	11,441	0.04	8,706	0.01	18,636	0.06	15,118
Other assets	0.01	64,397	0.03	2,831	0.04	6,352	0.06	32,944	0.08	125	0.04	8,000
Debt (other than mortgage)	0.03	9,866	0.09	12,127	0.14	30,171	0.12	21,234	0.10	27,834	0.10	22,442
Net non-housing wealth	0.92	13,569	0.95	20,857	0.97	23,431	0.95	32,960	0.98	23,395	0.95	22,751
Housing wealth												
Residential real estate	0.75	214,857	0.81	250,572	0.80	253,973	0.78	283,453	0.55	325,403	0.76	253,973
Non-residential real estate	0.10	199,000	0.04	165,000	0.12	135,000	0.07	169,760	0.09	115,866	0.08	165,000
Mortgage debt	0.75	192,300	0.77	176,470	0.70	106,797	0.69	88,488	0.41	60,503	0.69	138,856
Net housing wealth	0.76	22,185	0.81	82,233	0.81	164,399	0.79	193,868	0.57	250,042	0.77	142,860
Mortgage/property		0.90		0.66		0.37		0.25		0.10		0.44
Gross pension entitlements (standardized annuities)												
Public pensions	1.00	1,040	1.00	1,040	1.00	1,040	1.00	1,040	1.00	1,040	1.00	1,040
Private pensions[a]	0.99	1,140	1.00	1,265	0.99	1,215	0.97	867	0.95	950	0.98	1,122
Share of private pensions		0.52		0.55		0.54		0.46		0.48		0.52
Net standardized annuities												
Pensions	1.00	1,764	1.00	1,887	1.00	1,837	1.00	1,601	1.00	1,652	1.00	1,749
Percentage total		0.94		0.85		0.75		0.72		0.73		0.79
Pensions + wealth	1.00	1,872	1.00	1,994	1.00	1,997	1.00	1,819	1.00	1,884	1.00	1,917
Percentage total		0.99		0.91		0.81		0.79		0.82		0.86
Pensions + wealth + housing	1.00	2,002	1.00	2,304	1.00	2,509	1.00	2,504	1.00	2,442	1.00	2,328
N (HHs)	123		262		232		234		104		955	

[a] Private pensions are the sum of forecasted occupational pensions and self-reported third pillar pensions for individuals who do not claim non-public pensions. For current claimers private pensions are the sum of all non-public pensions received.

5.5.3 Annuities

The bottom panel of table 5.3 presents descriptive statistics of pension entitlements and after-tax annuities. We use three definitions of after-tax annuities, (1) annuities based on pensions (both public and private), (2) annuities based on pensions plus non-housing wealth, and (3) annuities based on all wealth, including housing wealth. To annuitize wealth we have to make a number of assumptions which we describe in this section, together with the descriptives.

Pensions, both public and private, form the main source of retirement income for the Dutch. Table 5.3 shows the monthly public pension benefit that people will receive at age 65 based on the assumption that respondents continue to live in the Netherlands until retirement. All households in the sample accumulate at least some public pension rights and the median pre-tax pension annuity equals 1,040 euro per month (benefits have been standardized to a one-person household). With regard to the second pillar, we observe occupational pension rights which assume that people remain employed in their current job with their current wage rate until the age of 65. These occupational pension entitlements are nominal rights with price indexation conditional on the financial situation of the pension fund. Because of the poor financial situation of most pension funds in the Netherlands in recent years, pension funds have been unable to make inflation corrections. For the future we assume that 50% of the inflation will be corrected and that inflation amounts to 2% per year. 98% of the households in the sample have built up occupational pension entitlements, with a median of 1,122 euro per month. As expected, private pensions are much more dispersed than the flat-rate public pensions, with a standard deviation of 1,057 (compared to 74 for public pensions).

Households may also deplete financial assets to finance their retirement. In view of this we annuitize private savings, using a real interest rate of 1% and mortality rates per gender and cohort predicted by Statistics Netherlands (latest version: 17 December 2010). For couples we assume that household wealth is divided equally, taking into account economies of scale and different life expectancies of spouses. Assume, for example, that a 50 year old man and a 45 year old woman have a total household wealth of 50,000 euros. When the man reaches the age of 65 he withdraws a fixed amount of money every year. After five years the wife also reaches the age of 65 and they both start to

withdraw money out of their household wealth. Probably, the wife lives longer than the man and after the death of the man we let her still take money out of the account. We take into account that as a widow, she needs relatively more money to be equally well of as before, because she looses economies of scale. Knoef et al. (2013) explain the calculations in detail. We allow for widowhood, but assume that couples stay together and that singles stay single. Also, we assume that remaining lifetimes of men and women are independent and we do not take into account differential mortality or any bequests.

In the last definition we also assume that households deplete net housing wealth during retirement. We take into account that homeowners who have already paid off part of their mortgage have relatively low housing costs, using an imputed rent of 4% of net capital accrued in property. With an inflation of 2% we have an imputed rent in real terms of 2%, which can be seen as a return on housing wealth. Finally, we assume an average yearly drop in real property prices of 1%.[8]

Finally, since respondents are asked about their preferred consumption levels, we have to take into account that tax pressures are relatively low for individuals aged 65 and over.[9] Therefore, we apply median tax pressures for elderly singles and couples per decile of gross income in 2008.

Table 5.3 shows that the median net annuity from pensions is 1,749 euro per month (standardized to a one-person household). Pensions are by far the most important category of wealth for our sample: on average they account for almost 80 percent of the annuity that can be attained if households were to spend all their wealth. If households would transform all their non-housing wealth into an annuity, the median monthly annuity would increase to 1,956 euro. Annuitizing housing wealth increases the median monthly annuity further to 2,328 euro per month. The question arises, however, whether households are able and willing to deplete housing wealth (e.g. by moving to a rental house, or by taking out a reverse mortgage).[10] The median annuities are generous

[8]This means that an individual of age 40 in 2008 experiences a drop in the real value of his house of 22% between now and the age of 65. We know that between 2008 and the present the average drop in housing prices was already 20% in real terms.

[9]After the age of 65 individuals face different marginal tax rates in the first two brackets of the income tax system and do not pay premiums for social insurance and social security anymore.

[10] Only 13% of the respondents between the ages of 55 and 64 are willing to move during retirement in order to lower housing costs.

compared to the medians of minimal and adequate expenditures (which are 1,460 and 1,606 respectively). Note, however, that we cannot simply use these descriptives to evaluate preparedness for retirement, since expenditure goals and resources are likely to be correlated and selection may play a role.

The bottom of table 5.4 presents after-tax annuities by age group. As expected, the share of private pensions is relatively high for young cohorts, because of the increased pension coverage (especially among women). The median pension annuity equals 1,764 euro for the youngest cohort and 1,652 euro for the oldest cohort distinguished in the table. When we take non-housing wealth into account median annuities increase, especially for the older cohorts: the median annuity for the youngest age group increases with 153 euros to 1,917 euro/month, while the median annuity for households in the age category 55-64 increases with 249 euros to 1850 euro/month. Housing wealth increases the differences between age groups further, since young households have a relatively low net housing wealth.

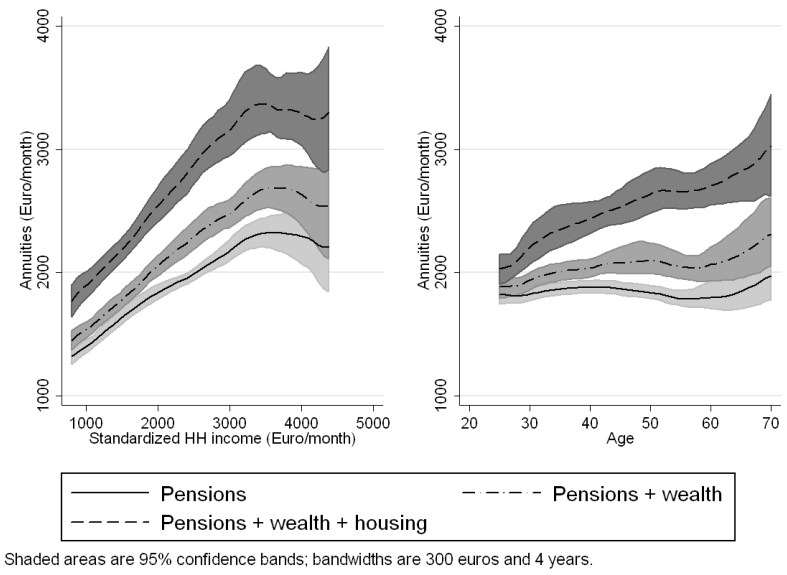


Figure 5.3: Kernel regressions of annuities on income and age of the household head (annuities and income are standardized to 1-person household)

Figure 5.3 shows kernel regressions of standardized annuities, net of taxes, on standardized household income and on the age of the household head. The

left panel shows that annuities increase with income and that this increase is somewhat stronger once we take private savings and housing wealth into account. The average pension annuity increases from 1,400 euro/month for an income of 1,000 euro to 2,300 euro/month for households with an income of 4,500 euro/month, while the total annuity (including all wealth components) increases from 1,900 to 3,300. On average, pension annuities do not increase for households with an income of 3,500 euro/month or more, but the variance increases. The right panel shows that older households have higher annuities, in particular when we take discretionary savings and housing wealth into account.

Measuring retirement readiness 5.6

The purpose of this paper is to evaluate the retirement readiness of a representative sample of the Dutch population. Instead of using a universal threshold, such as the often used 70% of preretirement income (e.g. Haveman et al. 2007), we use self-perceived minimal and adequate retirement expenditures to quantify retirement readiness. In this way we take into account that people may have different preferences. Self-perceived minimal retirement expenditures and self-perceived adequate replacement rates may differ by gender, income (Binswanger and Schunk 2012), and household characteristics. However, also unobserved characteristics may play a role. For example, conditional on observed characteristics, some people like to save more during their working life for (expensive) hobbies after retirement, whereas others can do with a relatively low expenditures level after retirement and prefer to spend more during working life. This may also influence the amount of wealth that we observe. Therefore, it is important to model assets and self-reported retirement goals simultaneously, to allow for correlation between the underlying unobserved heterogeneity. This section describes how we measure retirement readiness. First we consider representativeness (section 5.6.1), which determines which selection equations we have to use in the model. Section 5.6.2 describes the model and section 5.6.3 describes the simulations that we use to judge retirement savings adequacy.

5.6.1 Representativeness

To be able to draw representative conclusions about retirement readiness, item nonresponse and selection of the households that can be linked with the administrative data are important. Though survey response to the retirement expenditures questionnaire is a satisfactory 83 percent, the response rate to the questions that actually elicit expenditures during retirement is only 54-62 percent. Such low rates of item response are typical for these kind of questionnaires and probably due to the difficulty of the questions. Indeed, 33 percent of survey respondents agree that many questions are difficult for them to answer, even though only 13 percent find it very difficult to imagine how much money they would like to have during retirement. Neither survey response, nor item non-response occurred randomly across the potential sample. However, for evaluating the retirement preparedness of the Dutch the vital question is whether these selection effects introduce endogeneity in equations that explain retirement expenditures. In order to test this, we collapse survey and item non-response into a single selection indicator per question and run 2-step Heckman selection models of the logs of the different measures of expenditures on the covariates listed in Table 5.1. Our exclusion restrictions are measures of survey attitude taken from the 2008 personality questionnaire distributed to the LISS panel, supplemented with a dummy that indicates respondents who failed to answer to any of the yearly personality questionnaires in the 2008-2011 period. The explanatory power of those instruments in the selection equation is satisfactory: Wald tests for joint significance all convincingly reject the null at a significance level of 1 percent (test statistics are 227.83 for minimal expenditures and 132.66 for the measure of adequate expenditures against a critical value of 23.2 at a significance level of 1%). We do not find any evidence for significant selection issues with respect to retirement expenditures or income replacement rates (the inverse Mill's ratios are insignificant at the 10% level in all equations and remain insignificant when subsets of the instruments are considered). For all measures of expenditures during retirement we cannot reject that the selection process is independent from expenditures, allowing us to model expenditures without correcting for sample selection through non-response.

As described in section 5.4.2, not all of the 2,405 potential respondents of the expenditures questionnaire could be linked to the administrative data. The main cause for this is panel attrition. Panel attrition is not random. For example, we find that the self-employed are more likely to drop out of the sample and retirees, on the other hand, are less likely to drop out. The descriptive statistics shown in Table 5.1 indicate that the matched subsample has a relatively low average income compared to the potential sample. This raises concerns with respect to endogenous selection in the annuity equations. However, Heckman selection models indicate that sample selection for annuities is ignorable, which we verify using the model presented in the next subsection.

Model 5.6.2

To investigate how (pension) wealth and retirement goals interact we model self perceived minimal income and self perceived adequate replacement rates simultaneously with observed annuitized household wealth and the selection equation for observing wealth in the administrative data. Furthermore, whereas annuitized household wealth is the same for two members of a couple, self-perceived minimal and adequate expenditure levels may be different for men and women. Therefore, we include an equation for men and women and allow for the fact that the error terms for spouses may be correlated.

The model for self perceived minimal expenditures of men and women and annuitized household wealth can be described as follows

$$
\begin{aligned}
M_i &= x'_{mi}\beta_m + \epsilon_{mi} \\
N_i &= x'_{ni}\beta_n + \epsilon_{ni} \\
W_i &= x'_{wi}\beta_w + \epsilon_{wi} \\
d_i^* &= z'_i\gamma + \epsilon_{di}
\end{aligned}
$$

where M_i is self-perceived minimal retirement expenditures reported by a man in household i (if present) and N_i is self-perceived minimal retirement expenditures reported by a woman. W_i is annuitized household wealth and d_i^*

represents a latent variable indicating whether administrative data could be linked or not. The observed counterpart of d_i^* is

$$d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \text{ (household } i \text{ could be linked with wealth records)} \\ 0 & \text{if } d_i^* \leq 0 \text{ (household } i \text{ could not be linked with wealth records)} \end{cases}$$

We assume the error terms to be normally distributed with mean zero and covariance matrix Σ_M and estimate the model using Maximum Likelihood. For our research goal it is important to take into account correlation between annuitized wealth and self perceived adequate retirement expenditures. For example, those individuals who wish relatively high expenditures after retirement (conditional on their observed characteristics), may also have saved relatively a lot already.

5.6.3 Simulation

To judge retirement savings adequacy we simulate for all 1,780 households for whom we observe all covariates annuitized wealth and subjective adequate retirement expenditures, taking into selectivity and taking into account correlations between wealth and self-perceived minimal and adequate replacement rates.

For every individual we do several simulations, using the estimated model and with several draws of the error terms. For all simulations we confront annuitized wealth with self assessed minimal expenditures. Furthermore, we confront annuitized wealth with adequate expenditures, computed with the simulated replacement rates and current income. Instead of current income, later we may estimate average income during working life, to take into account income trajectories over the life cycle.

In addition to simulating various moments of the joint distribution of assets and expenditures, we also construct confidence intervals for those moments by means of a parametric bootstrap over the asymptotic distribution of our estimates.

Results 5.7

Estimation results 5.7.1

Before simulating retirement readiness, we first present estimation results from the models that form the basis for the simulations. As explained in the previous section, the descriptives suggest that endogenous selection may be a relevant concern for our assets data. Therefore, the models reported here correct for selectivity in assets. In the main text we only present estimation results from the equations for annuities and minimal and adequate expenditures. For estimates from the selection equations, see Appendix 5.C.

Annuities

Table 5.5 shows estimation results for the annuity equations of models pertaining to the minimal level of expenditures during retirement. We report estimates from three different measures of annuities: annuities based on public and private pensions; annuities that also include annuitized non-housing wealth; and finally annuities that include pensions, non-housing wealth and the annuitized value of all real estate net of mortgage.[11]

If we disregard the net value of real estate, we find no significant difference between the average standardized annuities accumulated by singles and couples. However, single males accumulate 18% higher standardized annuities on average relative to couples if we do take real estate into account. The difference between average annuities of single women and couples remains insignificant when we allow households to draw down their housing wealth. Households with older heads have slightly larger annuities on average if we take non-pension assets into account (the estimates vary between 0.45% per year if we do not take real estate into account and 0.98% if we do). The lack of a significant age effect in the equation for annuities from pensions alone in Table 5.5 depends on the extent to which we assume future pensions will be corrected for inflation: for 50% indexation there is no covariation of average

[11] In order to save space, we do not report the corresponding estimates from models of adequate expenditures. Those estimates confirm the patterns reported here and are available upon request.

Table 5.5: Joint models of annuities and minimal retirement expenditures - annuity equations

	Pensions		Pensions + Wealth		Pensions + Wealth + Housing	
Single	-0.0124	(0.0367)	0.0517	(0.0451)	0.180***	(0.0483)
Female × single	-0.0464	(0.0370)	-0.0762*	(0.0450)	-0.178***	(0.0483)
Age HH head	9.31e-04	(9.75e-04)	0.00445***	(0.00120)	0.00983***	(0.00129)
Any kids	-0.0885***	(0.0310)	-0.0745**	(0.0377)	-0.0879**	(0.0407)
Number children	0.0220*	(0.0133)	0.0153	(0.0162)	0.0322*	(0.0175)
Homeowner	0.109***	(0.0205)	0.165***	(0.0252)	0.441***	(0.0271)
Log HH income	0.109***	(0.0162)	0.106***	(0.0198)	0.0952***	(0.0213)
Inter. sec. ed.	0.0202	(0.0387	-6.94e-04	(0.0469)	0.0880*	(0.0506)
Higher sec. ed.	0.105**	(0.0450)	0.0931*	(0.0543)	0.215***	(0.0585)
Int. vocational ed.	0.143***	(0.0388)	0.155***	(0.0469)	0.202***	(0.0505)
Higher voc. ed.	0.253***	(0.0385)	0.302***	(0.0467)	0.369***	(0.0503)
University	0.331***	(0.0437)	0.435***	(0.0534)	0.461***	(0.0575)
1 salary worker	0.131***	(0.0329)	0.0341	(0.0399)	0.0394	(0.0430)
All salary workers	0.0806***	(0.0230)	0.0532*	(0.0282)	0.0240	(0.0303)
1 self employed	-0.126***	(0.0329)	-0.143***	(0.0404)	-0.0671	(0.0435)
All self employed	-0.201***	(0.0551)	-0.143**	(0.0673)	-0.179**	(0.0725)
1 retired	0.0858**	(0.0393)	0.0363	(0.0480)	0.0462	(0.0517)
All retired	0.0322	(0.0401)	-0.0326	(0.0487)	-0.0179	(0.0524)
1 disabled	-0.0834**	(0.0375)	-0.161***	(0.0454)	-0.125***	(0.0489)
All disabled	0.0668	(0.101)	0.0195	(0.122)	0.0141	(0.132)
Separated/divorced	-0.0786**	(0.0336)	-0.149***	(0.0417)	-0.130***	(0.0448)
Widow	0.0623	(0.0546)	0.0432	(0.0666)	0.0310	(0.0716)
Never married	-0.0397	(0.0268)	-0.0517	(0.0332)	-0.0211	(0.0357)
Constant	6.236***	(0.131)	6.248***	(0.161)	5.947***	(0.173)
Sigma epsilon	0.239***	(0.00577)	0.291***	(0.00679)	0.311***	(0.00724)
Log likelihood	-2,072.156		-2,255.035		-2,322.148	
N	1,780		1,780		1,780	

Dependent variables are logs of monthly annuities.
Annuities standardized to 1-person household; estimates taken from models of minimal expenditures.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

annuities with age, whereas zero index action yields a significant yet small age gradient for pensions.[12] Furthermore, households with one child have 7% smaller pension entitlements on average. Homeowners accumulate more wealth than renters: their entitlements are 11% larger on average if we only look at pensions, 17% larger if we look at pensions and non-housing wealth and 44% if we also include net housing wealth. Income-rich households tend to accumulate larger annuities, the estimated elasticities are around 0.1-0.11. The small size of the income elasticities may reflect measurement error in the self-reports of household income. This measurement error in income may also underly some of the large differences between the average entitlements of various educational groups. Better educated households are much richer: if we look at pensions alone, households in which at least one adult member has a university degree can look forward to a 33% higher annuity streams on average compared to households in which neither spouse finished secondary school. In addition to measurement error in income, such that the education dummies capture part of the relationship between income and annuities, another complementary explanation for the large disparities between education groups is the distinction between transitory and permanent income. Education is likely to act as a proxy for permanent income since the highly educated can expect to earn more throughout their life. Households with at least one salary worker have 13% larger pension entitlements, but the inclusion of wealth reduces this gap to a statistically insignificant 4%. Households in which at least one adult is self-employed have 13% smaller pension annuities and those for which all adult members are self-employed even have 33% lower pension entitlements compared to households without self-employed adults. Moreover, households for which all adult members are self-employed are poor relative to households with at least one non-self-employed member even if we take (non-)housing wealth into account: their annuity is 25% smaller on average. The finding that households in which spouses are self-employed accumulate less (pension) entitlements is plausible, since the self-employed are not covered by mandatory occupational pensions. The estimates suggest that they compensate the missing occupational pensions only partly with housing wealth. Since we rely on survey data to measure private accounts for those who are not claiming their pensions, a part or perhaps all of the estimated gap may be due to underreporting rather

[12] Estimates available on request.

than truly lower funds. We find little evidence for variation across groups with different marital status. Only, separated and divorced individuals are relatively worse off, especially if we take non-pension wealth into account.

Minimal and adequate expenditures during retirement

Table 5.6 contains estimates from the expenditure equations of the joint models of annuities and expenditures. The left panel refers to minimal and the right panel to adequate expenditures, both sets of estimates are taken from models of annuities based on pensions and wealth other than real estate.[13] We find that older respondents tend to report higher minimal and adequate expenditures (the only exception is that the correlation of minimal consumption with age is absent for men). For men both personal and household income are strongly positively associated with the minimal and adequate expenditure levels during retirement (the elasticities of minimal expenditures w.r.t. income are around 0.15 for both income measures and those of adequate expenditures are 0.16 for personal income and 0.35 for household income). For women, on the other hand, personal income does not affect minimal expenditures but household income does with an estimated elasticity of 0.48 (0.45 for adequate expenditures).

Higher educated men and women report significantly higher levels of minimal expenditures: male university graduates report 49% higher minimal expenditures compared to those who obtained no diploma beyond primary school (for women the corresponding difference is 39%). Note that these differences are even larger than those in annuities, so that highly educated individuals are more likely to feel inadequately prepared for retirement despite the fact that they are doing much better than their poorly educated counterparts in absolute terms. Better educated respondents also report higher adequate expenditures, but that difference is smaller (only 9% for men and 13% for women). One reason for the smaller size of the differences in adequate expenditures may be that they are measured more precisely, because respondents are guided much more in their answer to the adequate expenditures question compared to the item on minimal expenditures (for the former

[13]Estimates are similar to those obtained for different definitions of the annuities. Those estimates are available from the authors on request.

they answer by means of multiple choice scenarios that are designed to fit the personal situation of the respondent, while the latter is elicited through a single open-ended question). However, note that the estimates reported in Table 5.6 control for self-assessed question difficulty and understanding, variables which should reduce the impact of systematic biases in the response to these difficult questions. Alternatively, variation in the true subjective expenditures across education levels may well be larger for minimal than for adequate expenditures, since for a given level of current income the poorly educated may have more experience making ends meet in financially difficult times due to their lower permanent income. Descriptive statistics by educational groups tell us that the median of the minimal expenditures of the least educated is 1,168-1,200 euro per month, which is still well above the subsistence level provided by public pensions. The medians among the best educated, on the other hand, are in the 1,600-2,000 range, or almost twice the level of the public pension. Hence, the data do not indicate that the poorly educated give implausible answers to the minimal expenditures question. Instead, it seems that the best educated are very conservative in their assessment of their consumption floor.

We find some evidence that the self-employed are relatively demanding in terms of their expenditure goals during retirement: self-employed men report 18% higher minimal expenditures than do wage workers and for women the difference in adequate expenditures is 15%. Also, female homemakers are ambitious in their adequate expenditure level, which is 8% higher than that of women in a wage job. We find little evidence for systematic differences along the lines of marital status, except that married women report lower standardized minimal expenditures than the other groups. We find no evidence to suggest that the self-reported salience of retirement is related to the reported consumption level.

The estimates in tables 5.5 and 5.6 reveal that resources and perceived needs vary across the sample in ways that are relevant for policymakers. For instance, the finding that lowly educated respondents are both less demanding and poorer suggests that providing them with accurate information on the status of their retirement funds might not result in substantial changes in savings behavior. Indeed, the poorly educated may be perfectly prepared to meet their own modest goals. In order to induce additional savings one would have to directly target their perception of their consumption needs after

Table 5.6: Joint models of annuities and retirement expenditures - expenditure equations

	Minimal expenditures				Adequate expenditures			
	Men		Women		Men		Women	
Partner	-0.00667	(0.0792)	-0.164**	(0.0743)	-0.124***	(0.400)	0.0425	(0.0562)
Age	3.05e-04	(0.00268)	0.00484**	(0.00207)	0.00289**	(0.00116)	0.00588***	(0.00139)
HH head	0.0126	(0.0854)	-0.0940	(0.0591)	-0.105***	(0.0383)	0.00412	(0.0422)
Number children	-0.00221	(0.0230)	-0.0272	(0.0181)	0.0111	(0.0102)	-0.0176	(0.0126)
Homeowner	0.0215	(0.0571)	0.0954**	(0.0424)	0.119***	(0.0273)	0.0985***	(0.0316)
Log pers. income	0.173***	(0.0463)	-0.00149	(0.00954)	0.161***	(0.0222)	0.0125**	(0.00632)
Log HH income	0.158***	(0.0365)	0.481***	(0.0531)	0.347***	(0.0207)	0.448***	(0.0241)
Has simPC	0.00304	(0.117)	-0.00195	(0.0826)	-0.0784	(0.0658)	-0.119	(0.0781)
Inter. sec. ed.	0.0390	(0.0860)	0.0842	(0.0643)	-0.0280	(0.0367)	0.0286	(0.0438)
Higher sec. ed.	0.283***	(0.108)	0.246***	(0.0803)	0.0298	(0.0464)	0.146***	(0.0545)
Int. vocational ed.	0.224***	(0.0861)	0.220***	(0.0691)	0.0619*	(0.0376)	0.128***	(0.0471)
Higher voc. ed.	0.280***	(0.0863)	0.276***	(0.0696)	0.0537	(0.0371)	0.153***	(0.0479)
University	0.494***	(0.100)	0.392***	(0.0921)	0.0938**	(0.0436)	0.131**	(0.0619)
Self-employed	0.177**	(0.0762)	0.0319	(0.0662)	0.0142	(0.0317)	0.149***	(0.0453)
Home maker	0.278	(0.286)	0.00155	(0.0584)	-0.0503	(0.133)	0.0804**	(0.0390)
Retired	0.109	(0.281)	-0.0600	(0.250)	-0.150	(0.0963)	0.215	(0.163)
Disabled	0.0278	(0.137)	-0.0562	(0.102)	-0.114*	(0.0647)	-0.0310	(0.0648)
Other primary act.	0.272*	(0.143)	0.0899	(0.0861)	0.0721	(0.0842)	0.0255	(0.0604)
Separated/divorced	0.137	(0.0951)	0.151**	(0.0689)	0.0328	(0.0482)	0.0743	(0.0538)
Widow	-0.153	(0.173)	0.204*	(0.113)	0.0726	(0.0789)	0.0504	(0.0873)
Never married	0.0981	(0.0713)	0.107*	(0.0566)	0.0247	(0.0323)	0.0358	(0.0387)
Thought some	-0.0370	(0.0957)	-0.00956	(0.0856)	-0.00799	(0.0415)	0.0786	(0.0591)
Thought a little	-0.0709	(0.0953)	-0.0176	(0.0828)	0.0243	(0.0414)	0.0551	(0.0581)
Hardly thought	-0.134	(0.115)	-0.0334	(0.0899)	-0.0222	(0.0497)	0.0804	(0.0628)
No answer	0.0701	(0.292)	0.105	(0.260)	0.256**	(0.104)	-0.156	(0.171)
Constant	4.257***	(0.447)	3.183***	(0.431)	3.322***	(0.217)	3.338***	(0.222)
Sigma epsilon	0.545***	(0.0143)	0.414***	(0.0116)	0.255***	(0.00760)	0.323***	(0.00884)
Log likelihood	-2,255.035				-1,423.005			
N	1,780				1,780			

Dependent variables are logs of monthly minimal and adequate expenditures.
Expenditures standardized to 1-person household; equations reported from models of annuity excluding housing wealth but including other savings.
We control for self-reported understanding of the questions (estimates available on request).
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

retirement. The self-employed, on the other hand, stand out as a group that is simultaneously demanding in terms of their post-retirement consumption and accumulates relatively little wealth. Hence, they might be expected to be more reactive to information campaigns.

Table 5.7 shows the estimated correlations between the error terms of the models that explain minimal expenditures. We find that, conditional on covariates, annuitized assets and minimal expenditures have a weak positive correlation for men, and a stronger correlation for women. Furthermore, the correlation between the expenditures levels reported by spouses is positive and statistically significant, but also relatively small at 0.13. Most importantly, the correlations between the error terms of the selection equation and the annuity equation are between -0.01 and -0.07 and statistically insignificant, implying that sample selection in annuities is exogenous.

Table 5.7: Error correlations for model of minimal expenditures

	Annuity	Min exp. men	Min exp. women	Selection (annuity)
		Annuity from pensions		
Annuity	1			
Min exp. men	0.0740*	1		
Min exp. women	0.213***	0.128**	1	
Selection (annuity)	0.0685	-0.0620	-0.0840	1
		Annuity from pensions and non-housing wealth		
Annuity	1			
Min exp. men	0.0611	1		
Min exp. women	0.278***	0.131**	1	
Selection (annuity)	-0.0245	-0.0375	-0.0747	1
		Annuity from pensions and all wealth		
Annuity	1			
Min exp. men	0.0888**	1		
Min exp. women	0.180***	0.123**	1	
Selection (annuity)	-0.0140	-0.0388	-0.0686	1

*significant at 10%; **significant at 5%; ***significant at 1%

The estimated correlations for our models of adequate expenditures and annuities including non-real estate wealth are reported in Table 5.8. Similarly to table 5.7, the estimated correlations that capture selectivity in observed annuities are small. The most striking difference between these correlations and those for minimal expenditures is that conditional on covariates there

seems to be much more agreement between spouses on what an adequate expenditure level is compared to minimal expenditures: we estimate the correlation between adequate expenditures to be around 0.87 (compared to 0.13 for minimal expenditures). Note, however, that this may reflect the very different modes of answering the questions: the agreement may be an artefact of the choice between no more than 6 different expenditure levels that respondents are presented with for the adequate expenditures question. Correlations between annuities and adequate expenditures are much stronger than between annuities and minimal expenditures: for men they range from 0.14 if we take all wealth into account to 0.19 if we drop housing wealth while for women they fall between 0.30 and 0.46. Note, finally, that the error term of the selection equation for annuities is significantly negatively correlated with the reported adequate expenditure levels of women. Hence, there are efficiency gains from keeping the selection equation for annuities in the model, despite the lack of evidence for endogenous sample selection.

Table 5.8: Error correlations for model of adequate expenditures

	Annuity	Adequate exp. men	Adequate exp. women	Selection (annuity)
		Annuity from pensions		
Annuity	1			
Adequate exp. men	0.153***	1		
Adequate exp. women	0.303***	0.879***	1	
Selection (annuity)	0.0191	-0.0641	-0.147***	1
		Annuity from pensions and non-housing wealth		
Annuity	1			
Adequate exp. men	0.189***	1		
Adequate exp. women	0.455***	0.858***	1	
Selection (annuity)	-0.105	-0.0875*	-0.157***	1
		Annuity from pensions and all wealth		
Annuity	1			
Adequate exp. men	0.143***	1		
Adequate exp. women	0.405***	0.866***	1	
Selection (annuity)	-0.0800	-0.0867*	-0.145***	1

*significant at 10%; **significant at 5%; ***significant at 1%

5.7.2 Simulations

We use the estimates presented in the previous subsections to simulate the extent to which individuals are able to realize their personal retirement ex-

penditure goals. Table 5.9 presents our simulation results. In addition to comparing annuities with reported minimal consumption levels (top panel), we also compare resources with a poverty line defined by Statistics Netherlands (second panel); with self-reported adequate expenditures (third panel); and with 70% of current standardized household income (bottom panel). Moreover, we run simulations for the annuities in the data (columns labeled “Baseline") and carry out robustness checks for annuities from which we deducted a 20% drop in private pensions (columns labeled “Pensions -20%") and for annuities minus a 20% drop in housing prices (columns labeled “Housing prices -20%").

Looking first at the baseline estimates and self-reported minimal expenditures, we find that though the median individual can expect to exceed his/her minimal expenditures by 25% based on pensions alone, a sizable fraction of 33% will fall short of their minimum unless they fill the gap with private savings. The shortfall is often large, with a median of 31%. Non-pension wealth helps to fill the gap: the fraction that falls short drops to 26% once we take non-real estate savings into account. Net housing wealth further reduces the proportion with insufficient funds to 17%, showing that home ownership is rare among those who fall short of their minimum consumption level even if they would draw down non-housing wealth.

Comparing the second panel baseline simulations to the first, we notice that respondents report much higher consumption floors than the poverty line of 917 euro/month. While 33% will not be able to reach their minimal consumption level based on pensions alone, only 3% falls short of the basic poverty line (the median exceeds it by 65%). If we take into account all three types of wealth, the fraction that falls below the basic poverty line drops to 2%.

Table 5.9: Percentage differences between annuities and consumption floors

	Baseline			Pensions - 20%			Housing prices -20%		
	Median	Fraction <0	Med. shortfall	Median	Fraction <0	Med. shortfall	Median	Fraction <0	Med. shortfall
				Heterogenous targets - minimal expenditures					
Pensions	25	0.33	-31	16	0.38	-32			
	(17, 33)	(0.29, 0.39)	(-37, -30)	(9, 24)	(0.34, 0.44)	(-38, -31)			
Pension + wealth	37	0.26	-28	30	0.30	-30			
	(30, 46)	(0.22, 0.31)	(-35, -27)	(0.22, 0.38)	(0.26, 0.35)	(-35, -28)			
Pensions + wealth + housing	57	0.17	-26	50	0.20	-27	43	0.23	-29
	(48, 65)	(0.15, 0.23)	(-33, -25)	(42, 59)	(0.17, 0.26)	(-33, -26)	(36, 51)	(0.20, 0.29)	(-35, -27)
				Poverty line plus (917 euro/month)					
Pensions	65	0.03	-11	57	0.03	-10			
	(60, 70)	(0.02, 0.05)	(-12, -10)	(52, 61)	(0.03, 0.05)	(-11, -9)			
Pensions + wealth	76	0.02	-11	69	0.03	-10			
	(71, 82)	(0.02, 0.04)	(-13, -10)	(63, 75)	(0.02, 0.05)	(-12, -10)			
Pensions + wealth + housing	97	0.02	-13	91	0.03	-13	84	0.03	-13
	(90, 104)	(0.02, 0.03)	(-14, -12)	(85, 98)	(0.02, 0.04)	(-15, -13)	(78, 89)	(0.02, 0.04)	(-15, -12)
				Heterogeneous targets - adequate expenditures					
Pensions	4	0.46	-29	-4	0.54	-31			
	(0, 10)	(0.41, 0.51)	(-34, -26)	(-9, 1)	(0.49, 0.58)	(-35, -29)			
Pensions + wealth	18	0.34	-25	10	0.41	-27			
	(13, 25)	(0.29, 0.39)	(-30, -23)	(5, 17)	(0.35, 0.46)	(-31, -24)			
Pensions + wealth + housing	37	0.23	-25	31	0.27	-26	23	0.32	-27
	(30, 43)	(0.19, 0.28)	(-31, -21)	(25, 37)	(0.23, 0.32)	(-32, -22)	(17, 29)	(0.28, 0.37)	(-32, -24)
				70% of current income					
Pensions	21	0.30	-22	12	0.37	-23			
	(16, 26)	(0.26, 0.35)	(-24, -22)	(8, 16)	(0.33, 0.42)	(-25, -22)			
Pensions + wealth	33	0.23	-22	26	0.28	-23			
	(28, 39)	(0.20, 0.27)	(-24, -21)	(20, 31)	(0.25, 0.32)	(-24, -22)			
Pensions + wealth + housing	52	0.14	-21	45	0.18	-21	38	0.21	-21
	(46, 58)	(0.12, 0.17)	(-22, -20)	(40, 50)	(0.16, 0.21)	(-23, -21)	(34, 44)	(0.19, 0.25)	(-24, -22)

Pensions include public and occupational mandatory savings, as well as private pensions.
Wealth includes all discretionary savings that are not automatically annuitized, except for property.
90% confidence intervals in parentheses, calculated by parametric bootstrap (500 replications).

The third panel of Table 5.9 compares annuitized wealth with the *desired* level of expenditures. Looking first at the baseline simulations, we find that 46% of the population is expected to fall short of their ideal consumption level if they would rely exclusively on their pension entitlement (the median difference is only 4%). However, if we also include wealth other than real estate, the fraction that falls short is reduced to 34% and the median individual exceeds their desired level of expenditures by a comfortable 18%. Allowing households to spend down housing wealth makes for an even more favorable picture: less than a quarter of the sample would be unable to afford their consumption goal and the median difference between annuities and desired consumption is 37%.

Looking at the bottom panel of Table 5.9, we notice that adequate expenditures are higher in general than 70% of current household income. Comparing the simulation results in the bottom two panels, the fraction that falls short is 9-16 percentage points lower when using the replacement rate criterion compared to heterogenous targets. Similarly, the median difference between the annuity and adequate expenditures is about 15 percentage points larger.

The middle columns of Table 5.9 show that cuts in occupational pension benefits up to 20% have a limited influence on our conclusions: the fraction that cannot afford their minimal expenditures increases by no more than 8 percentage points while the amount by which the median individual exceeds the various consumption floors drops by 5-9 percentage points. Note that around the poverty line we find households without any private pensions entitlement: the fraction of households that cannot afford the poverty line remains unchanged when we cut pensions.

As shown in the rightmost columns of Table 5.9, a drop of 20% in the price of houses has a larger effect on savings sufficiency for homeowners. The proportion that falls short of the various consumption floors under the reduced housing prices scenario is close to that obtained if we disregard housing wealth altogether.

Appendix 5.D investigates the robustness of simulation results to self-reported question difficulty and to the forecasts made by Statistics Netherlands. In addition to the baseline results from Table 5.9, we also report simulations for which we set the understanding of the questions to the maximal level. Doing so reduces the median difference between annuity and minimal consumption by 3 percentage points and increases the fraction with insufficient savings by

1-2 percentage points. We find larger discrepancies for self-reported adequate expenditures: the fraction that does not have enough savings to finance their desired expenditures increases by 2-7 percentage points and median excess savings are reduced by 6-7 percentage points. Finally, Appendix 5.D also describes what happens to savings sufficiency if we set the age of all household heads to 50 in the annuity equations. The reason we would be interested in doing so is that pension entitlements are forecasted based on a continuation of the status quo: individuals are supposed to stay in the same job and earn the same wage until retirement. This simplistic forecasting scheme may not be appropriate for younger individuals. Simulations are similar when we set the age of all respondents to 50: median excess savings increase by 1-2 percentage points and the proportion of individuals who cannot afford to reach the various consumption levels decreases by the same amount. Hence, the simplifying assumption of a continuation of the present labor market situation up to retirement does not affect our conclusions.

Table 5.10 illustrates the impact of variation in resources and goals on self-assessed retirement readiness. The upper panel shows simulations based on the actual data, where we see that university graduates are more likely to fall short of their ambitions than are their less educated peers (47% fall short compared to 29% of those who haven't obtained a degree beyond secondary school). This difference does not, however, reflect poor preparation of the highly educated: when we set the education level of all respondents in the annuity equation to university, this cuts the incidence of insufficient preparation among those with no more than secondary school by a factor of 2 or more, depending on which assets we take into account. Instead, university graduates tend to set very high minimal expenditure levels: the third panel of Table 5.10 shows that if all respondents would be as ambitious as university graduates, the fraction that cannot afford their retirement would double for the lowest education categories.

Finally, Table 5.11 demonstrates the impact of different assumptions regarding the correction of occupational pensions for inflation. All preceding results are based on the assumption that nonpublic pensions will be indexed for 50% of inflation for the remaining lifetime of the panel members. The top panel of Table 5.11 reiterates those results and breaks them up by age brackets. For the annuity based on pensions we find that younger individuals fall short of their

Table 5.10: Simulated incidence of shortfalls w.r.t. minimal expenditures across education categories

		By education level					
	All respondents	Primary	Lower sec.	Higher sec.	Inter. voc.	Higher voc.	University
				Data			
Pensions	0.33	0.29	0.30	0.40	0.30	0.34	0.47
	[-31]	[-30]	[-29]	[-35]	[-29]	[-30]	[-40]
Pensions + wealth	0.26	0.22	0.23	0.34	0.25	0.25	0.35
	[-28]	[-27]	[-27]	[-31]	[-28]	[-27]	[-33]
Pensions + wealth + housing	0.17	0.17	0.15	0.22	0.18	0.16	0.27
	[-26]	[-27]	[-24]	[-28]	[-27]	[-25]	[-31]
				All annuities set to level of university graduates			
Pensions	0.24	0.14	0.16	0.28	0.21	0.29	0.47
	[-28]	[-24]	[-23]	[-30]	[-26]	[-29]	[-40]
Pensions + wealth	0.15	0.08	0.08	0.18	0.13	0.19	0.35
	[-26]	[-24]	[-21]	[-25]	[-25]	[-25]	[-33]
Pensions + wealth + housing	0.11	0.06	0.06	0.12	0.10	0.13	0.27
	[-24]	[-26]	[-21]	[-26]	[-22]	[-23]	[-31]
				Minimal expenditures set to level of university graduates			
Pensions	0.50	0.58	0.56	0.53	0.45	0.46	0.47
	[-39]	[-45]	[-40]	[-40]	[-35]	[-36]	[-40]
Pensions + wealth	0.41	0.49	0.47	0.44	0.38	0.35	0.35
	[-35]	[-39]	[-36]	[-36]	[-34]	[-32]	[-33]
Pensions + wealth + housing	0.29	0.39	0.32	0.31	0.29	0.23	0.27
	[-32]	[-37]	[-33]	[-31]	[-31]	[-28]	[-31]

Median negative percentage differences between annuities and expenditures in brackets.
Simulations use 50 draws of the error terms.

minimal expenditures slightly less often than do older people: the fraction that falls short increases with age from 27% for respondents aged 25-34 to 39% for respondents aged 65 or older. However, if we allow households instead to consume all their wealth, older individuals fall short less often (14% of 65+ and 22% of the youngest age group). Though pension funds usually index benefits, in principle the entitlements for occupational pensions are nominal. To safeguard the solvency of pension funds, pensions have not been corrected to keep up with rising prices since 2008. To illustrate the effect of such suspension of indexation, the bottom panel of Table 5.11 simulates shortfall relative to minimal expenditures under the scenario that occupational pensions are not indexed for inflation. The effect on the fraction that fails to meet their personal consumption floor is limited: the fraction that fall short increases with a modest 3-4 percentage points depending on the definition of wealth used. This increase in the incidence of inadequate annuities is concentrated in the younger cohorts: the fraction below age 35 that cannot afford minimal expenditures increases with 5-8 percentage points. For the oldest cohorts, on the other hand, suspension of indexation almost does not affect their financial position.

Table 5.11: Simulated incidence of shortfalls w.r.t. minimal expenditures across age groups

		By age bracket				
	All respondents	25-34	35-44	45-54	55-64	65+
		50% indexation				
Pensions	0.33	0.27	0.31	0.34	0.36	0.39
	[-31]	[-28]	[-30]	[-31]	[-33]	[-32]
Pensions + wealth	0.26	0.23	0.25	0.26	0.27	0.26
	[-28]	[-28]	[-28]	[-29]	[-29]	[-26]
Pensions + wealth + housing	0.17	0.22	0.19	0.17	0.16	0.14
	[-26]	[-28]	[-28]	[-27]	[-27]	[-24]
		No indexation				
Pensions	0.37	0.35	0.37	0.38	0.38	0.38
	[-32]	[-31]	[-32]	[-32]	[0.33]	[-32]
Pensions + wealth	0.29	0.30	0.30	0.29	0.28	0.25
	[-29]	[-30]	[-30]	[-29]	[-29]	[-27]
Pensions + wealth + housing	0.20	0.27	0.22	0.19	0.17	0.14
	[-28]	[-31]	[-29]	[-26]	[-27]	[-24]

Median negative percentage differences between annuities and expenditures in brackets. Simulations use 50 draws of the error terms.

5.8 Conclusion

Population aging together with the poor performance of financial markets during recent years have put pension systems around the world under severe pressure. As a result pensions have become less generous, shifting responsibility for maintaining an adequate standard of living during retirement to the individual. Against the backdrop of this changing environment, we investigate whether the Dutch can reasonably hope to accumulate sufficient resources to meet self-defined expenditure requirements for subsistence and for adequate living.

In contrast to previous research, we evaluate retirement readiness by comparing the expected financial situation at age 65 with expenditure thresholds that are specified by the respondents themselves, allowing needs to vary across the sample. We take these subjective expenditure needs from a survey that was distributed to the LISS-panel in January of 2008. This deviation from a one-size-fits-all yardstick for sufficiency of savings is found to be important: both minimal and adequate expenditures vary widely between households. The variation in expenditures is related to factors such as education and type of

employment, with higher educated and self-employed respondents indicating higher self-perceived expenditure needs during retirement compared to the poorly educated and salary workers. Our analysis disentangles the roles of goals and resources and identifies groups who are at risk of accumulating insufficient assets to retire comfortably.

Another important facet of the paper is the use of administrative data on various asset categories such as savings, investments, housing wealth and public and occupational pensions. We take into account "automatic" saving in public and occupational pensions, by using pension funds' best predictions of accrued entitlements at age 65 under continuation of the status quo. Such administrative data measure assets more precisely than would be possible using survey data alone. However, tax records necessarily miss saving vehicles that are subject to delayed taxation for those who are not yet drawing annuities. Therefore we rely on administrative data where possible and supplement with information from the LISS assets survey where necessary to construct a complete picture of the resources available to households.

We show that wealth, especially as accumulated through public and occupational pensions, suffices for a majority of respondents to meet and exceed their own minimal and adequate expenditures. By age 65, the median respondent is likely to be able to afford 25% higher expenditures compared to his/her own personal minimal level based on pensions alone. If the households also consume out of non-pension savings, the median excess of annuities over minimal consumption rises to 37% if we do not include net housing wealth and 57% if we do. However, the affluence of the sample as a whole hides a sizable minority of 17 percent that will be unable to afford their minimal expenditures, even if they continue to accumulate pensions until age 65 and if we include housing in our measure of wealth. Self-reported minimal consumption is high compared with the official poverty line of 917 euro per month: less than 5% of the adult population falls short of that yardstick.

Joint models of annuitized wealth and subjective expenditures show that homeowners and the highly educated accumulate relatively much wealth, in pensions and (non-)housing savings, while households in which members are self-employed are on average 17-30% poorer. Both minimal and desired expenditures are positively related to income and education, though we find that personal income matter much more for men than for women, whose

expenditure wishes are correlated mostly with household income rather than personal income. The net effect is that we find that the highly educated are more likely to fall short of their own goals, since their goals are much more ambitious. Indeed, once we control for ambitiousness the highly educated are found to be less likely to fall short. For desired retirement expenditures we find that income raises expenditure targets more than resources.

5.9 Acknowledgements

The authors thank Marcel Das, Edwin de Vet, Bettina Lamla, Annette Scherpenzeel, Maarten van Rooij, Arthur van Soest, Mathijn Wilkens, and participants of the MESS conference 2012, the PHF-SAVE conference 2013 on Household Finances, Saving and Inequality: An International Perspective, and the CPB Research Seminar 2013 at the Netherlands Bureau for Economic Policy Analysis for their help and useful comments.

More details on sample selection 5.A

Survey and item non-response 5.A.1

Though survey response to the retirement expenditures questionnaire is a satisfactory 83 percent, the response rate to the questions that actually elicit expenditures during retirement is only 54-62 percent. Such low rates of item response are typical for the questionnaire and probably due to the difficulty of the questions. Indeed, 33 percent of survey respondents agree that many questions are difficult for them to answer, even though only 13 percent find it very difficult to imagine how much money they would want during retirement. Next we describe the processes of survey and item non-response in detail.

First we estimate a univariate probit for response to the survey, comparing the 2,005 respondents with the 400 non-respondents.[14] Older individuals are more likely to respond: an increase in age of 10 years is associated with a 9 percentage point higher probability of answering at least 1 question. Moreover, respondents with children are slightly less likely to answer, the difference being 2 percentage points per child. Education matters too: those who have completed at least intermediate vocational training are 5-7 percentage points more likely to answer. Unfortunately, being self-employed is associated with a 9 percent point lower probability to respond to the questionnaire. The sample of actual survey respondents is older, better educated, has less children and is less likely to be self-employed than the potential sample.

We also analyze item non-response conditional on answering at least 1 question of the survey. We allow for dependence between non-response to different items in the same survey by estimating a trivariate probit.[15] Response to the question on minimal expenditures during retirement follows an inverted U-shaped pattern in age: respondents around the age of 46 are most likely to answer that question (response to the questions on adequate expenditures does not vary systematically with age). Also, household heads are 12 percentage points more likely to answer the minimal expenditures question, but equally likely as their spouses to answer the other questions. Homeowners are 7-9 percentage points more likely to provide an assessment of their adequate retire-

[14] Estimates available on request.
[15] Estimates available on request.

ment expenditures. Conditional on answering to the survey, individuals who have completed higher vocational training or university are 8-10 percentage points more likely to answer the difficult questions, though this difference disappears for adequate expenditures under a high interest rate. Perhaps because they find it easier to answer, retirees are 14 percentage points more likely to answer the minimal expenditures question (but answer similarly often to the adequate expenditures questions). Non-response to the different retirement expenditures questions is not independent: the correlations between the error terms of the equation for response to the minimal expenditures question and those for adequate expenditures are 0.32 and 0.33, s.e. 0.04, and the correlation between the error terms of the two measures of adequate expenditures is 0.85, s.e. 0.02.

The conclusion from the previous paragraphs is that neither survey nor item non-response occurs randomly across the potential sample. However, for our purpose of evaluating the retirement preparedness of the Dutch the vital question is whether these selection effects introduce endogeneity in equations that explain retirement expenditures. In order to test this, we collapse survey and item non-response into a single selection indicator per question and run 2-step Heckman selection models of the logs of the different measures of expenditures on the covariates listed in Table 5.1.[16] Estimation results from the level equations for expenditures and replacement rates are discussed in section 5.5.1, here we limit our discussion to sample selection. Our exclusion restrictions are measures of survey attitude taken from the 2008 personality questionnaire distributed to the LISS panel, supplemented with a dummy that indicates respondents who failed to answer to any of the yearly personality questionnaires in the 2008-2011 period. The explanatory power of those instruments in the selection equation is satisfactory: Wald tests for joint significance all convincingly reject the null at a significance level of 1 percent (test statistics are 227.83 for minimal expenditures and 132.66 and 147.86 for the measures of adequate expenditures against a critical value of 23.2 at a significance level of 1%). Despite the relevance of the instruments, we do not find any evidence for significant selection issues with respect to retirement expenditures or in-

[16]We disregard the correlations between non-response to different questions, because joint modelling of the selection processes would rely on joint normality assumptions. We opt for robustness and carry out estimation in 2-steps. Estimates are available on request.

come replacement rates (the inverse Mill's ratios are insignificant at 10% in all equations and remain insignificant when subsets of the instruments are considered). For all measures of expenditures during retirement the selection process is independent from expenditures, allowing us to model expenditures without correcting for sample selection through non-response.

Linking the LISS to administrative data 5.A.2

The match between the LISS survey and the tax records of assets and earnings was made during the first months of 2012. All individuals who participated in the LISS panel at that moment received an E-mail asking whether they objected against matching their surveys with administrative sources. Because of ethical considerations, respondents who had dropped out of the LISS and hence did not give their permission were not matched. Of the 2,405 potential respondents to the retirement expenditures questionnaire 1,292 individuals, or 54%, remained in the panel at the time of the merge and were asked for permission. A small group of 134 actively did not allow for their data to be matched, so that 1,158 survey records could potentially be combined with administrative data. Hence, the selection issue with respect to assets records is primarily one of panel attrition with few active objections against the combination of survey and administrative data.

First we describe how the sample of 1,292 respondents who received the request for the merge differs from the potential sample of 2,405 (based on their characteristics in 2008).[17] The tendency to remain in the panel follows an inverted U-shape in age, with a maximum at age 48. As was the case for survey response, we find that the self-employed are 9 percentage points more likely to drop out of the sample altogether. Retirees, on the other hand, are 15 percentage points less likely to drop out. The strongest predictor of remaining in the sample is owning a simPC: respondents who were provided with a simple computer to complete the online questionnaires are 21 percentage points less likely to leave the sample between 2008 and 2012. We should be careful, however, not to interpret this large difference as a causal effect, since

[17] Estimates available on request.

respondents who did not own a computer in 2008 are likely to differ from the other respondents in many other ways, some of which may be unobserved.

Only 10 percent of the respondents who remained in the panel objected against linking their survey records to administrative data. Comparing those objectors to all other respondents who were still in the panel by 2012, we find that age is the only predictor that is significant at 5%.[18] The tendency to object is non-linear in age, with a peak at 59.

Non-response in the retirement expenditures questionnaire, which determines whether we observe desired expenditures, and attrition from the sample, which drives whether we observe assets, are likely to be related to one another. Indeed, bivariate probits of successful merges and response to the relevant questions reveal that the correlations between the error terms are in the range 0.19-0.25 (with standard errors close to 0.032). However, once we condition on the survey attitude variables used as exclusion restrictions in the selection models, the error correlations are reduced to 0.06-0.08 (with standard errors around 0.035). Hence, we feel safe to model non-response and attrition separately and include only a single selection equation in the models.

[18]Estimates available on request.

Measurement error in subjective expenditures 5.B

We expect that giving an indication of the expenditures one needs or desires during retirement is a challenging task for respondents. The fact that only 1,300-1,500 of the 2,005 survey-respondents answer the particular questions on minimal and adequate expenditures suggests that those questions are difficult. In this appendix we use additional information from the LISS questionnaire to investigate the effect of question difficulty, and presumably measurement error, on the analysis.

Thinking about retirement 5.B.1

One reason why respondents may not be able to give a good indication of their expenditures during retirement is that they may not have thought about retirement yet. As mentioned in Section 5.5.1, our data include a self-assessment of the extent to which respondents have thought about retirement, so we can test whether those who have given retirement a lot of thought give different answers compared to those who have not. Table 5.12 shows that 71% of the sample has thought either "a little" or "hardly at all" about retirement, which may be a problem when answering questions on expenditures during retirement. Moreover, retirement is clearly a more salient concern to pre-retirees: 86% of the respondents aged 25-34 have not yet thought about retirement compared to 51% of those aged 55-64 (65 was the eligibility age for the public pension in 2008).

As reported in the text, we check for systematic differences in reported expenditures between respondents who have and have not thought about retirement by adding dummies corresponding to the categories in Table 5.12 to the equations for minimal and desired expenditures. We find no evidence for such differences.

Difficulty of the questions 5.B.2

In addition to the salience of retirement, we also want to control for the extent to which respondents understand the questions. Self-reported question difficulty and understanding allow us to investigate whether those who do not

Table 5.12: Descriptives of thinking about retirement

		By age				
	All respondents	25-34	35-44	45-54	55-64	65+
	Mean	Mean	Mean	Mean	Mean	Mean
Thought a lot about retirement	0.06	0.02	0.03	0.05	0.13	0.07
Thought some	0.23	0.12	0.19	0.24	0.37	0.28
Thought a little	0.52	0.54	0.57	0.55	0.41	0.46
Thought hardly at all	0.19	0.32	0.22	0.15	0.10	0.19
N	1,671	289	486	507	335	54

understand the questionnaire give systematically different answers. Moreover we correct for variation in understanding and find that our results are robust (see Tables 5.9 and 5.15). Table 5.13 summarizes the items that measure understanding and difficulty of the questions.

Table 5.13: Descriptives of self-reported question difficulty

		By age				
	All respondents	25-34	35-44	45-54	55-64	65+
	Mean	Mean	Mean	Mean	Mean	Mean
Many questions didn't make sense to me						
Definitely not = 1	0.09	0.07	0.08	0.10	0.13	0.07
2	0.10	0.11	0.08	0.10	0.11	0.11
3	0.18	0.14	0.17	0.18	0.20	0.21
4	0.30	0.25	0.32	0.30	0.29	0.29
Yes, definitely = 5	0.33	0.43	0.34	0.32	0.26	0.31
Many questions were too abstract for me						
Definitely not = 1	0.09	0.08	0.10	0.08	0.08	0.12
2	0.23	0.28	0.26	0.21	0.19	0.20
3	0.29	0.30	0.29	0.29	0.30	0.29
4	0.24	0.24	0.23	0.25	0.25	0.21
Yes, definitely = 5	0.15	0.10	0.13	0.17	0.17	0.17
I generally do not like to think about old-age provision						
Definitely not = 1	0.20	0.16	0.17	0.19	0.24	0.26
2	0.20	0.15	0.22	0.21	0.20	0.20
3	0.33	0.39	0.36	0.32	0.29	0.25
4	0.19	0.23	0.18	0.21	0.18	0.16
Yes, definitely = 5	0.08	0.07	0.06	0.08	0.09	0.13
I find it very difficult to imagine how much money I would want to have during retirement						
Definitely not = 1	0.17	0.22	0.19	0.15	0.13	0.18
2	0.20	0.25	0.26	0.22	0.11	0.17
3	0.28	0.27	0.28	0.29	0.29	0.23
4	0.22	0.17	0.20	0.23	0.27	0.22
Yes, definitely = 5	0.13	0.09	0.07	0.11	0.20	0.20
I like to take some responsibility for my old-age provision						
Definitely not = 1	0.16	0.16	0.19	0.16	0.12	0.16
2	0.21	0.28	0.21	0.25	0.15	0.16
3	0.31	0.36	0.36	0.28	0.32	0.24
4	0.19	0.14	0.15	0.19	0.23	0.22
Yes, definitely = 5	0.13	0.06	0.08	0.13	0.18	0.23
N	1,990	287	485	502	441	275

5.C Estimates of the selection equations

Table 5.14: Joint models of annuities and minimal retirement expenditures - selection equations

	Pensions		Pensions + Wealth		Pensions + Wealth + Housing	
Single	-0.105	(0.138)	-0.0240	(0.137)	-0.0242	(0.137)
Female × single	0.0540	(0.140)	0.0450	(0.140)	0.0446	(0.140)
Age HH head	-0.00454	(0.00377)	-0.00437	(0.00376)	-0.00437	(0.00376)
Any kids	-0.146	(0.125)	-0.135	(0.125)	-0.135	(0.125)
Number children	0.0584	(0.0558)	0.0537	(0.0557)	0.0535	(0.0557)
Homeowner	-0.0374	(0.0798)	-0.00914	(0.0797)	-0.00891	(0.0797)
Log HH income	-0.0918	(0.0683)	-0.0984	(0.0672)	-0.0989	(0.0675)
Inter. sec. ed.	-0.266*	(0.153)	-0.295*	(0.153)	-0.295*	(0.153)
Higher sec. ed.	-0.277	(0.179)	-0.263	(0.179)	-0.264	(0.179)
Int. vocational ed.	-0.207	(0.156)	-0.204	(0.156)	-0.205	(0.156)
Higher voc. ed.	-0.220	(0.154)	-0.229	(0.154)	-0.229	(0.154)
University	-0.284*	(0.172)	-0.327*	(0.172)	-0.327*	(0.172)
1 salary worker	0.284**	(0.126)	0.247**	(0.126)	0.247	(0.126)
All salary workers	-0.0669	(0.0919)	-0.0803	(0.0918)	-0.0799	(0.0918)
1 self employed	0.0401	(0.129)	0.0121	(0.129)	0.0116	(0.129)
All self employed	0.0848	(0.211)	0.0504	(0.211)	0.0501	(0.211)
1 retired	-0.0728	(0.152)	-0.0956	(0.152)	-0.0969	(0.152)
All retired	0.115	(0.147)	0.113	(0.147)	0.114	(0.147)
1 disabled	0.434***	(0.163)	0.449***	(0.163)	0.449***	(0.163)
All disabled	-0.783**	(0.352)	-0.812**	(0.353)	-0.813**	(0.353)
Separated/divorced	0.0584	(0.131)	-0.0179	(0.131)	-0.0175	(0.131)
Widow	-0.132	(0.200)	-0.188	(0.200)	-0.188	(0.200)
Never married	-0.152	(0.104)	-0.209**	(0.104)	-0.209**	(0.104)
Personality missing	-0.823***	(0.0803)	-0.810***	(0.0805)	-0.809***	(0.0805)
Constant	1.376**	(0.554)	1.398***	(0.546)	1.403***	(0.547)
Log likelihood	-2,072.156		-2,255.035		-2,322.148	
N	1,780		1,780		1,780	

Dependent variables are indicators equal to 1 if we observe the annuity.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 5.15: Robustness w.r.t. question difficulty and extrapolation of pension entitlements

	Simulations based on observed data			Everybody understands questions			Annuities for age 50		
	Median	Fraction <0	Med. shortfall	Median	Fraction <0	Med. shortfall	Median	Fraction <0	Med. shortfall
				Heterogenous targets - minimal expenditures					
Pensions	25	0.33	-31	22	0.35	-31	25	0.33	-31
	(17, 33)	(0.29, 0.39)	(-37, -30)	(6, 37)	(0.26, 0.46)	(-38, -28)	(17, 33)	(29, 39)	(-37, -30)
Pensions + wealth	37	0.26	-28	36	0.26	-27	37	0.26	-29
	(30, 46)	(0.22, 0.31)	(-35, -27)	(20, 51)	(0.19. 0.37)	(-34, -25)	(30, 46)	(0.22, 0.31)	(-35, -28)
Pensions + wealth + housing	57	0.17	-26	54	0.18	-26	57	0.17	-27
	(48, 65)	(0.15, 0.23)	(-33, -25)	(39, 70)	(0.13, 0.26)	(-32, -24)	(49, 65)	(0.15, 0.23)	(-33, -25)
				Poverty line plus (917 euro/month)					
Pensions	65	0.03	-11				65	0.03	-10
	(60, 70)	(0.02, 0.05)	(-12, -10)				(60, 70)	(0.02, 0.05)	(-12, -10)
Pensions + wealth	76	0.02	-11				77	0.02	-11
	(71, 82)	(0.02, 0.04)	(-13, -10)				(71, 83)	(0.02, 0.04)	(-13, -10)
Pensions + wealth + housing	97	0.02	-13				99	0.02	-12
	(90, 104)	(0.02, 0.03)	(-14, -12)				(91, 105)	(0.01, 0.03)	(-14, -11)
				Heterogeneous targets - adequate expenditures					
Pensions	4	0.46	-29	-3	0.53	-28	4	0.46	-30
	(0, 10)	(0.41, 0.51)	(-34, -26)	(-12, 5)	(0.45, 0.61)	(-34, -26)	(0, 10)	(0.41, 0.51)	(-34, -27)
Pensions + wealth	18	0.34	-25	12	0.38	-23	18	0.35	-26
	(13, 25)	(0.29, 0.39)	(-30, -23)	(3, 22)	(0.30, 0.47)	(-28, -22)	(11, 24)	(0.30, 0.41)	(-31, -23)
Pensions + wealth + housing	37	0.23	-25	30	0.25	-22	37	0.22	-25
	(30, 43)	(0.19, 0.28)	(-31, -21)	(21, 39)	(0.20, 0.33)	(-25, -20)	(32, 44)	(0.18, 0.27)	(-30, -22)
				70% of current income					
Pensions	21	0.30	-22				21	0.30	-23
	(16, 26)	(0.26, 0.35)	(-24, -22)				(16, 26)	(0.26, 0.35)	(-24, -22)
Pensions + wealth	33	0.23	-22				34	0.22	-22
	(28, 39)	(0.20, 0.27)	(-24, -21)				(28, 39)	(0.19, 0.26)	(-23, -21)
Pensions + wealth + housing	52	0.14	-21				53	0.13	-21
	(46, 58)	(0.12, 0.17)	(-22, -20)				(48, 59)	(0.11, 16)	(-22, -20)

Pensions include public and occupational mandatory savings, as well as private pensions.
Wealth includes all discretionary savings that are not automatically annuitized, except for property.
90% confidence intervals in parentheses, calculated by parametric bootstrap (500 replications).

6 Can Survey Participation Alter Household Financial behavior?

This chapter is based on Crossley et al. (2013).

6.1 Introduction

Much research in empirical economics and finance relies on the analysis of survey data collected during interviews with individuals and households. Increasingly, models that account for individual or household heterogeneity make use of panel data where people are surveyed repeatedly over time. A wide range of potential survey effects need to be considered to ensure that data collected from such exercises is valid and reliable. A fundamental question is whether confronting respondents with detailed questions about their financial circumstances can alter their later behavior. The extent to which measurement alters the behaviors it is measuring is a core question for many disciplines and one that has not been confronted in depth in economics and finance. We provide the strongest empirical evidence to date on this important question in a central domain of economic behavior.

There are a number of intuitive reasons why survey-induced behavioral change may occur. One is the notion of limited attention, which is common in behavioral finance (see DellaVigna 2009, for an overview of empirical evidence for limited attention in field experiments). Limited attention means that individuals tend to overlook some of the consequences of their decisions. If those unnoticed consequences materialize in the future, as do the benefits of saving today, this results in biases that are similar to those induced by

limited self-control (Karlan et al. 2012). However, in contrast to self-control problems, limited attention suggests that behavior can be corrected by focusing individuals' attention on the aspects they are missing. For instance, Karlan et al. (2012) show that reminders are an effective means to increase savings. In the context of developing countries, they find that reminding people who enrolled in goal-specific savings programs of the benefits of saving increases the amount invested in those accounts and the likelihood of attaining the corresponding goal. Surveys may also serve as shocks to attention. Stango and Zinman (2011) show that survey respondents in the US are less likely to incur overdraft fees, penalties paid to banks in case the balance of a deposit account turns negative, after answering non-informative survey questions that ask for their attitude towards those fees. They find that more general questions on spending control that do not directly concern overdraft fees have the same effect. This suggests that individuals think associatively when it comes to household finance and that relatively subtle reminders of certain aspects of (spending) decisions affect subsequent behavior.

Another literature in psychology proposes that there may be a "question-behavior" effect whereby asking respondents to predict future behavior results in decisions that are in line with those predictions (Dholakia 2010). One strand of literature has examined a "self-prophecy" effect, whereby committing to a normatively desirable action can increase the likelihood of that action occurring (Sherman 1980). For example, Spangenberg (1997) shows that asking people to predict their workout behavior induces them to visit the gym more often. Several papers have also shown that surveying people about risky behaviors can increase the propensity to engage in risk behaviors (Dholakia 2010, Fitzsimmons and Moore 2008, Fitzsimmons and Shiv 2001). This literature demonstrates that asking people questions can influence specific behaviors particularly when the question relates to defined time periods. There have even been recommendations in the public health literature to use self-recording of behavior as a potential behavioral change intervention (Michie et al. 2011a,b).

However, survey measurement could also influence reporting of behavior with respondents either becoming fatigued from regularly reporting information and therefore less accurate or becoming more sensitive to perceived appropriate behavior. A body of work on food diary surveys has demonstrated

that respondents tend to report fewer episodes of food consumption toward the end of 1-week diary studies (Biltoft-Jensen et al. 2009, Lillegaard et al. 2007).

Furthermore, it is important to disentangle effects that arise from filling out the survey and effects that arise due to differential attrition. Panel conditioning has been discussed in the survey and economics literature for several years. Several studies have examined this potential effect in domains such as subjective well-being (Van Landeghem 2012); marital satisfaction (Glenn 1998) and a number of other areas. However, as pointed out by Das et al. (2011), disentangling panel conditioning from unobserved factors influencing attrition is a complex task. Das et al. (2011) find that, controlling for unobserved attrition factors, that there are significant panel conditioning effects on knowledge questions but not in other types of questions.

An important recent paper by Zwane et al. (2011) provides the most compelling evidence to date that surveys can directly impact on objectively observed economic behavior. Their study examined the effect of randomly assigning people to extra survey monitoring in the context of randomized controlled trials conducted across five developing country applications. A significant feature of their study was the ability to monitor objective outcomes, thus ensuring that any survey effects were indeed due to behavioral change rather than changes in reporting style. Strikingly, while they find no effect in two of the experiments, randomly assigning respondents to extra surveys significantly increased the probability of water treatment product usage, medical insurance usage and biased estimates of the health benefits of improved water source quality. As well as being statistically significant, the magnitude of the survey effect on water quality is large enough to obscure the effect of improved source quality on the incidence of diarrhea. Such effects demonstrate that measurement effects on behavior operate distinctively from effects of actually being observed.

We make a number of contributions in this paper. Firstly, we provide the first experimental estimates of the effect of being randomly assigned to a detailed financial survey on later financial behavior in a large representative population survey. In particular, we examine the effect of being asked questions about retirement provision on later savings behavior. Our methodology provides for an almost perfectly clean experimental estimate of the effect and we employ a number of recently developed techniques for dealing with treatment non-compliance to deal with any potential threats to validity.

We test the main hypothesis using the Dutch LISS panel, a household survey that has taken place in the Netherlands since 2007. LISS is particularly useful as it is conducted on a monthly basis assessing a range of financial, health and social outcomes. Crucially for this study, respondents are randomly assigned to be eligible for some of the modules. Thus, we have a strong measure of intention-to-treat when it comes to our main treatment. The treatment itself is a lengthy survey on retirement expectations taken by respondents in January 2008. The survey neither provides any information on the Dutch pension system in general nor on the pension entitlements of respondents. This was the first randomized module introduced on the LISS panel. Our main outcome variable is constructed through linkage, with respondent permission, to the Dutch national tax record system. This records savings and debt across different asset classes. We are thus able to construct a very robust measure of saving between 2007 and 2009 and examine whether the survey randomization influenced real behavior. Thus, as well as providing experimental evidence on a large representative sample, we also can be sure that any experimental effect is affecting behavior as opposed to survey reporting. Furthermore, the nature of the sample allows for exploration of heterogeneous treatment effects. Savings behavior is crucial to economics and finance and providing experimental evidence in such a central domain is particularly relevant.

Our findings demonstrate a very strong, robust and negative effect of the treatment on savings rates in the sample. We demonstrate this through the use of a quantile IV procedure that uses the survey offer as an intention-to-treat variable. There is no conceivable sample selection effect that drives our results. Attrition from the survey and permission to grant access to tax records were both unaffected by the randomization. Furthermore, a falsification test examining the effect of the survey treatment on pre-survey savings behavior yields completely null results. The Dutch retirement system is among the highest ranked in the world in terms of covering post-retirement income, with state and occupational pensions alone covering a median of 70 per cent of highest income post-retirement (Bovenberg and Meijdam 2001). Therefore, in that institutional context surveying leads mainly to dissaving. To the extent that there are heterogeneous treatment effects, the treatment largely affects respondents who are closer to retirement and better educated. Such heterogeneity is plausible given that older and better educated households have the highest

pension entitlements (De Bresser and Knoef 2013). It should be noted that these are not Hawthorne effects, whereby respondents might change their behavior if aware directly they were being monitored. There is no incentive for respondents to change their behavior in terms of relationships between the participants and the survey agency. Therefore, what we are measuring is a pure survey measurement effect that is particularly concerning for panel studies of financial behavior.

The overall reduction in savings and the effect heterogeneity along the lines of age and education are consistent with limited attention. According to that interpretation, the survey on pensions and retirement increased the salience of those subjects, reminding individuals of issues in the future that they tend to overlook. After reflecting on their expenditure needs in retirement, older and highly educated households concluded that they can afford to save less while the young and poorly educated marginally increased their savings. This reflection was aided by the availability of information on individual-level pension entitlements provided by Universal Pensions Overviews (UPOs), which financial institutions were obliged to provide to all pension holders from 2008 onward. Those UPOs give members of occupational and private pension funds standardized information on current entitlements and projects for age 65 (the statutory retirement age for public pensions during the period covered by the sample).

Our findings are the most robust evidence to date of survey effects on behavior. One implication is that targeted informational interventions may work in the field of household finance as they do in other fields (examples of information affecting household finance are given in Stango and Zinman 2011, and Karlan et al. 2012). Methodologically, survey effects need to be given far greater consideration in terms of potential biases they introduce to models that require the use of panel data. The results confirm the notion by Zwane et al. (2011) that a survey design which infrequently fields surveys to a large panel may be preferable to one that surveys a smaller sample intensively. Moreover, our research design illustrates the use of randomization of survey modules as a credible identification strategy for the estimation of causal effects. Such randomized modules can be used to study the updating of expectations and the effectiveness of various types of financial education and their effects on financial behavior. In general, randomization in panel surveys provides a

relatively cost efficient way to evaluate interventions and test theories, outside the confines of the lab.

The rest of this paper is structured as follows. Section 6.2 outlines the research design in the study. Section 6.3 outlines the LISS panel data and administrative data being used. Section 6.4 describes in details the experimental results, robustness checks and falsification tests. Section 6.5 concludes with implications for future research.

6.2 Research design

6.2.1 Overview

The aim of this paper is to identify the effect of survey participation on household savings. We use survey data from the LISS panel, which is a representative random sample from the Dutch population that was initiated during the autumn of 2007.

The LISS panel is administered by CentERdata, a survey research institute affiliated with Tilburg University, and follows close to 8,000 individuals from 5,000 households. Surveys are distributed over the internet every week. Though the Netherlands has a high rate of internet access, more than 80% of Dutch households are connected, CentERdata safeguards representativeness by providing selected households with an internet subscription and a simple computer when necessary.

In addition to core surveys on subjects such as demographics, income and assets, researchers have the possibility to design their own questionnaires. Those customized modules are usually distributed only once and, to keep costs down, are often limited to a random subsample of the eligible sample. This distribution of modules to random subsets of the LISS panel generates exogenous variation in survey participation, which we exploit to estimate the effect of participation on household financial behavior.

The treatment 6.2.2

We define treatment as participation in the survey entitled "What is an adequate old age income?". This questionnaire was constructed by Johannes Binswager and Daniel Schunk and it was the first randomized module to be fielded in the LISS panel in January 2008. It consists of around 60 items that primarily concern minimal and desired expenditure levels in retirement, the tradeoff between current and future consumption and risk preferences with respect to income after retirement. Moreover, the questionnaire asks about respondents' willingness to cut back on housing expenditures and the extent to which they have thought about retirement. The questionnaire did not provide respondents with any information about the Dutch system of retirement income provision in general or the respondent's personal entitlements in particular. For our purpose of cleanly identifying a survey effect, it is important to note that the survey did not include any questions on predicted or intended savings. Therefore, our design rules out question behavior: respondents are not asked to predict their own behavior.

Eligibility for the survey was limited to all LISS members that were 25 years or older, who had a net household income of at least 800 euro per month and who were either the head of the household or his/her partner (children or other household members were excluded from participation). This lead to a total eligible sample of 5,435 individuals, 2,755 of which were selected at random and were offered the survey. The take-up rate among those that received the offer was 74%. See Binswanger and Schunk (2012) for more information on the questionnaire and an analysis of the answers it elicited.

Outcome measures 6.2.3

We investigate the effect of survey participation on household savings and on satisfaction with the state of the Dutch economy. Though the LISS data include elaborate biannual surveys on assets and debt, we prefer to use administrative wealth records for two reasons. Firstly, there is the general concern about the quality of self-reported survey data on assets (Bound et al. 2001). Exploratory analysis reveals that the self-reports of wealth in the LISS data are no exception. Comparing administrative records with self reports from the

same households, we find that the reported ownership of asset classes such as stocks and bonds is approximately 10 percentage points lower in the LISS assets module. Furthermore, conditional on ownership panel members tend to understate the value of their assets. Therefore, measurement error is an important reason to prefer tax-derived administrative records over self-reports when it comes to savings. Secondly, we want to rule out the confounding influence of differences in reporting behavior between those who were and were not offered the retirement expenditures survey. If we would find an effect of survey participation on self-reported savings, one could argue that the survey changed reporting styles rather than behavior. Deriving our outcome measures from administrative data mitigates that concern.

Informed consent for the match of LISS data with administrative records was elicited in September of 2011. Unfortunately, panel attrition limits the number of households for which we could obtain a match: out of the 3,125 households that contain at least 1 member that was eligible for the retirement expenditures survey, we could match only 1,602. De Bresser and Knoef (2013) show that this loss of data is mostly due to panel attrition rather than objections: only 10% of the respondents to the retirement expenditures survey that were still in the LISS panel in 2011 objected against the match.

The administrative assets data are taken from the Complete Asset Data of the Netherlands-dataset (Integraal Vermogensbestand, CAD), which was constructed by Statistics Netherlands. The CAD contains a detailed decomposition of household-level wealth for the entire Dutch population. The categories of assets that we observe are checking and saving accounts, bonds, stocks, property, other real estate, business capital and other tangibles. For debt the CAD distinguishes between mortgage and other debt. It measures assets on the first of January for the years 2007, 2008 and 2009 (data for more recent years are not yet released at the time of writing). Available records thus allow us to compute yearly savings during 2007 and 2008 as the differences between wealth stocks in consecutive years. We compute these wealth stocks net of the value of the primary residence, because we want to focus on pure savings and housing has an important consumption component.

The CAD is based on tax records, which are supplemented with information from banks. Though the records provide a measure of assets that is likely to be more accurate than survey data, the fact that they are mostly derived from

taxes means they are not complete. We miss savings held in small accounts, because banks are not obliged to report accounts with a balance less than 500 euro or 15 euros in interest payments. We also do not observe debt for households without capital income, which means that we miss most short-term debt. Finally, we miss savings held in tax-exempt 3rd pillar pensions. Such accounts are taxed only during the payout phase and are therefore invisible in tax records up to retirement. Given that the treatment-survey concerns expenditures after retirement, one may expect the effect of participation on savings to be particularly strong for those savings vehicles aimed at generating additional income after retirement. We cover these third pillar accounts by means of specific items from the LISS assets module that ask for ownership and balances of such accounts. Because of the high likelihood of substantial measurement error, we do not add the self-reported private pensions to the tax records. Instead, we analyze them separately and report the findings in section 6.4.5.

In our analysis of savings we look both at *levels*, in euro per year, and *rates*, which are levels divided by yearly disposable income. The data on the yearly disposable income of households are also taken from tax records. We use the Complete Household Income Data of the Netherlands-dataset (Integraal Huishoudens Inkomstenbestand, CHID), assembled by Statistics Netherlands. The measure for primary income in the CHID is quite complete: in addition to labor income it includes income from entrepreneurship and from assets (interest payments and imputed rent for homeowners). Disposable income is defined as primary income plus government transfers that the household received minus the transfers and taxes paid by the household. The administrative income measure that we use is likely to be more accurate than survey measures of income, since information about the various income streams is provided electronically by employers and financial institutions to the tax authority. Hence, our income data is unlikely to suffer from reporting errors.

Though our focus is on the effect of survey participation on household savings, we also look at the impact of the survey on subjective outcomes that may be relevant to savings. The LISS data provide a rich set of relevant outcome measures. In particular, the income module that was fielded during the summer of 2008 asked respondents about their satisfaction with the economic situation

in the Netherlands. Such subjective perceptions may drive savings and therefore we also investigate whether they were affected by the survey.

6.2.4 Institutional context

In order to understand how a survey about expenditures in retirement might affect savings, we need to describe briefly the institutional context of Dutch pensions. The Dutch system of income provision during retirement is easily understood in terms of 4 categories or "pillars". The first pillar is that of the public pension, which provides everybody who lived in the Netherlands between the ages of 15 and 65 with a subsistence income. Coverage of the public pension is close to universal, since uninterrupted residence in the country is the only criterion (benefits are cut by two percent for each year spent abroad).[1] The level of the public pension is set in reference to the minimum wage. Since public pensions only provide a minimum income, almost all employees accumulate additional entitlements in occupational pensions (the second pillar). Such arrangements cover 90% of all employees and are usually organized at the level of the sector or of the company (Bovenberg and Meijdam 2001). Participation in the first two pillars is mandatory and together they replace 70 percent of gross last earned income on average, which translates to replacement rates net of taxes above 80 percent (Bovenberg and Meijdam 2001, Kapeyn and De Vos 2008). The third pillar contains all private savings vehicles that are aimed specifically at retirement, such as life annuities. Such voluntary arrangements are especially important for individuals that are not included in occupational pensions, such as the self employed. Finally, all other forms of wealth that can be drawn down to generate additional income after retirement, such as savings accounts, investments and real estate, make up the fourth pillar. As explained in the previous section, our assets data allow us to look at the effect of survey participation on the accumulation of assets in this fourth pillar.

It is important to stress that the Dutch pension system in 2008 was characterized by arrangements that were almost universal and provided extremely

[1] Technically, one is covered by the public pension if one's income is subject to Dutch income taxes. Residence abroad does not affect the accumulation of entitlements as long as your income is taxed within the Dutch system.

generous income replacement. In this institutional environment it is not surprising that the first two pillars, public old age pensions and occupational pensions, together provided 95% of income in retirement (Kapeyn and De Vos 2008). The final 5% was accounted for by private pension products and other savings.

Starting from 2008, individuals can find detailed information on their own pension entitlement in occupational and private funds in their Uniform Pension Overview (UPO). These UPOs provide all members of pension funds, both in the second and third pillar, with yearly updates on their current entitlements and projected future entitlements at age 65. UPOs are mandatory for all financial institutions in the Netherlands since january 1st of 2008.

Threats to validity 6.2.5

Our analysis faces two threats to internal validity. The first problem is the issue of incomplete compliance with the treatment: not everybody who was offered the survey responded. We apply two remedies. First, we do an intention-to-treat (ITT) analysis that compares those who did receive the offer with those who did not (instead of comparing those who were treated with those who were not). Second, we do instrumental variables (IV) analyses in which we use the random offer of treatment as an instrument for being treated. Both methods allow us to obtain estimates of treatment effects that are not affected by endogenous sample selection as a result of non-response, since they rely on exogenous variation in the survey-offer. In addition to IV regressions for the conditional mean of the savings distribution, we also estimate unconditional decile treatment effects in order to establish the robustness of our results. We use the estimator proposed in Frölich and Melly (2013) (for Stata code implementing this estimator, see Frölich and Melly (2010)). Our estimates can be interpreted as local average treatment effects, since our research design imposes the monotonicity requirement for the effect of the instrument on the likelihood of being treated explained in Angrist et al. (1996): nobody who wasn't offered the survey could participate in it.

The second threat to internal validity is that of selection into outcome measurement due to the substantial loss of observations when we match LISS observations with administrative records. As mentioned above, we could only link 1,602 out of 3,125 eligible households in the LISS panel to administrative

records because of attrition in the period between the survey (January 2008) and the match (September 2011). Therefore, it is important that we verify that sample selection is not related to the offer of the retirement expenditures survey. We check this by testing for mean independence of the instrument from sample selection.

6.3 Data

6.3.1 Matching LISS and administrative data

The basic unit of our analysis of savings is the household, since we measure both wealth and income at the household level. We classify a household as being offered the survey if at least one household member that was eligible for the survey received the offer of participation. Likewise, we classify all households as treated in which at least one member that was offered the survey actually filled it out. The construction of our estimation sample starts with 3,125 households that contain at least one member that was eligible for the retirement expenditures survey according to the criteria mentioned in section 6.2.2. After matching the LISS respondents to administrative data, we obtain wealth records for 1,429, 1,437 and 1,449 households in the years 2007-2009 respectively. We drop those households for which all eligible members were retired in 2008, reducing the sample to 1,275 households. Finally, we trim all households for which 2008 savings rates relative to after-tax household income were larger than 50% in absolute value, leaving us with an estimation sample of 999 households.[2]

Table 6.1 presents descriptive statistics for the full sample and for the estimation sample, separately for couples and singles (note that the full sample consists of 2,816 rather than 3,125 households, because we exclude households for which all eligible members were retired in 2008). For couples, defined as households in which two partners live together irrespective of their marital status, individual-specific attributes are reported for the head of the household. For instance, about 12% of couples have a female head of the household, while

[2]We also tried trimming the sample at savings rates equal to 75% and 100% of net household income and found that our results are robust. Estimates are reported in Appendix 6.B.

around 60% of the singles are females. The mean age is 47 for couples and 45 for singles and couples have more children on average than do singles (1.15 compared with 0.40). Homeownership is prevalent among couples but not among singles: 82% of the former own their home while only 45% of the latter do. Among those living with a partner, over 80% are married compared to 5% among people living without a partner. Around half of the singles were never married, while a little over a third are widowed. Three education levels stand out that account for about 25% of the sample each: intermediate secondary education, intermediate vocational and higher vocational training. Household heads and singles alike are mostly engaged in some form of employment (75% of the former and 70% of the latter are).

Table 6.1: Descriptive statistics

	Couples				Singles			
	Full sample (LISS)		Estimation sample		Full sample (LISS)		Estimation sample	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Female	0.12	(0.32)	0.11	(0.32)	0.58	(0.49)	0.62	(0.49)
Age	47.4	(11.8)	46.9	(11.5)	45.4	(12.1)	44.6	(11.6)
Children	1.16	(1.15)	1.15	(1.16)	0.41	(0.81)	0.36	(0.78)
Homeowner	0.83	(0.38)	0.81	(0.40)	0.49	0.50)	0.42	(0.50)
Marital status								
Married	0.81	(0.39)	0.83	(0.38)	0.06	(0.23)	0.04	(0.19)
Separated/divorced	0.05	(0.23)	0.04	(0.20)	0.33	(0.47)	0.38	(0.49)
Widowed	0.002	(0.04)	0.001	(0.04)	0.10	(0.30)	0.07	(0.25)
Never married	0.13	(0.34)	0.11	(0.32)	0.51	(0.50)	0.52	(0.50)
Education								
Primary	0.08	(0.28)	0.08	(0.28)	0.09	(0.28)	0.09	(0.29)
Int. Secondary	0.23	(0.42)	0.25	(0.43)	0.24	(0.43)	0.26	(0.44)
Higher secondary	0.07	(0.26)	0.08	(0.28)	0.08	(0.27)	0.07	(0.25)
Int. vocational	0.25	(0.43)	0.25	(0.43)	0.23	(0.42)	0.24	(0.43)
Higher vocational	0.25	(0.43)	0.25	(0.43)	0.25	(0.43)	0.28	(0.45)
University	0.11	(0.32)	0.09	(0.29)	0.10	(0.30)	0.06	(0.25)
Most imporant activity								
Employed	0.72	(0.45)	0.75	(0.43)	0.69	(0.46)	0.70	(0.46)
Self employed	0.11	(0.31)	0.07	(0.26)	0.09	(0.29)	0.08	(0.28)
HH work	0.01	(0.11)	0.01	(0.11)	0.05	(0.21)	0.06	(0.23
Retired	0.11	(0.31)	0.11	(0.32)	0.00	(0.00)	0.00	(0.00)
Disabled	0.03	(0.16)	0.03	(0.17)	0.07	(0.26)	0.07	(0.25)
Other	0.03	(0.16)	0.02	(0.15)	0.10	(0.30)	0.09	(0.29)
N	2167 (77.0%)		768 (76.9%)		649 (23.0%)		231 (23.1%)	

For couples all individual-specific variables refer to the head of the household.

In addition to describing the complete LISS data, Table 6.1 allows one to compare the characteristics of all eligible LISS households with those of the

final estimation sample. Both for couples and singles the samples are found to be similar.

One difference between the full sample and the estimation sample that is not in Table 6.1 is that of compliance to the survey offer. As mentioned in Section 6.2.2, 74% of the individuals in the complete sample who were offered the survey participated. Household-compliance is around 5 percentage points higher: among those households for which at least one eligible member was offered the survey at least one member filled it out in 79% of the cases. In the estimation sample the corresponding compliance rates are 82% for individuals and 87% for households, which is 8 percentage points higher than in the complete sample. It is not surprising that compliance is related to being observed in our final dataset, since non-compliers to the survey offer are more likely to attrit from the LISS-panel altogether. As a result, non-compliers were less likely to give their permission for the match to administrative records and are lost from our estimation sample. However, this does not compromise our research design, so long as the instrument is orthogonal to this selection process. In Section 6.4.1 we show that the instrument is not correlated with selection into the estimation sample. Hence, the comparison of households based on the random survey offer is as valid there as it would be in the complete sample.

6.3.2 Descriptive statistics

Table 6.2 describes our administrative assets records for the years 2007, 2008 and 2009 (all in 2008 euros). The single most important category of assets is that of the primary residence, with an average value of around 200,000 euro. Savings accounts follow at great distance as the second most important type both in terms of mean (27,000 euro) and median (13,000 euro) value. Real estate other than the primary residence is also important, but only for a small minority: the mean value is around 7,000 euro though only 8% of the sample has any non-residential real estate. The mean value of risky assets, stocks and bonds, drops from 7,210 euro in 2007 to 4,857 euro in 2009 (median holdings are zero in all years). Business wealth and other wealth are the least important categories of assets with a mean value below 1,500 euro in all years.

As mentioned in section 6.2.3, we observe both mortgage and non-mortgage debt. On average households have about 105,000-110,000 euro in mortgage

Table 6.2: Descriptives of assets and debt

	2007			2008			2009		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Assets									
Saving accounts	25,551	12,799	40,870	26,728	12,728	42,568	28,008	13,196	44,963
Risky assets	7,210	0	23,974	6,627	0	22,560	4,857	0	17,468
Property	196,713	205,571	162,020	201,325	212,589	161,463	199,616	212,181	155,063
Real estate	9,808	0	53,750	6,906	0	41,722	7,689	0	44,442
Business	1,202	0	12,588	1,212	0	13,842	1,459	0	15,485
Other	861	0	9,914	959	0	10,871	1,014	0	10,779
Debt									
Mortgage	105,119	86,091	105,553	104,079	84,827	103,787	108,243	91,412	106,892
Non-mortgage debt	2,375	0	13,796	1,992	0	12,621	2,432	0	18,823
Net worth	133,852	78,760	172,072	137,687	81,551	170,305	131,968	77,021	167,938
Net housing wealth	91,594	40,984	131,904	97,246	47,128	128,481	91,374	43,949	124,627
Net worth excl. housing	42,258	16,292	82,970	40,440	15,643	76,797	40,594	15,232	77,325
N		983			999			999	

All assets are reported in 2008 euros.

debt and around 2,000-2,500 euro of non-mortgage debt. Non-mortgage debt is concentrated in a small minority of 6% of the sample, among which the mean non-mortgage debt is around 20,000 euro.

Taking assets and debt together, the mean net worth of the households in the sample is around 135,000 euro. Unsurprisingly, net worth is concentrated in the primary residence, which has a mean value net of mortgage of around 95,000 euro. Because of the consumption value of housing, we compute savings based on the remaining 40,000 euro of non-housing savings.

Table 6.3 presents summary statistics of net income and for the outcome variables for savings used in the analysis. Mean household income is 38,165 euro in 2008 and the median is 35,699 euro, both of which are slightly higher than the average of 33,100 euro for the Dutch population at large (Statistics Netherlands 2012). 2008 non-housing savings, computed as the difference between the non-housing wealth stocks of January 1st 2008 and 2009, has a sample mean of 154 euro and a median of 2 euro, showing that the distribution of savings is centered around zero. There is, however, considerable variation in savings: the standard deviation is 9,411 euro. We compute savings rates as the level of savings divided by after-tax income. The distribution of savings rates is centered around zero, but there is considerable variation: the standard deviation of the savings rate is 19 percentage points.

Table 6.3: Descriptive statistics of outcomes

			Percentiles				
	Mean	Std. dev.	0.05	0.25	0.5	0.75	0.95
HH income[a]	38,165	17,649	16,289	27,056	35,699	46,107	67,474
Non-housing savings[b]							
Levels (2008 euros)	154	9,411	-13,632	-3,221	2	3,084	14,860
Savings rates	-0.01	0.19	-0.40	-0.09	0.00	0.09	0.33
N				999			

[a] HH income net of taxes.
[b] Savings corrected for inflation and net of property value and mortgages.

6.4 Results

6.4.1 Validity of the instrument

As explained in section 6.3.1, we lose about half of our sample when we match LISS records with administrative data. This loss of data compromises the internal validity of our empirical strategy if attrition from the LISS panel is related to our instrument, the offer of the retirement expenditures survey. Therefore, it is important to establish that the offer of the survey is not related to sample selection. Table 6.4 shows estimates of a linear model that uses our instrument, called "offer", to explain an indicator of sample retention. We find that sample selection is not correlated with the offer of the retirement expenditures survey, so the substantial loss of data that comes from matching survey participants to administrative data does not affect our identification.

In the regression in Table 6.4 and in all other models reported below we control for the presence of multiple household members. This is necessary, because the randomization of the offer was done at the level of the individual while the outcome variables we analyze are measured at the household level. We classify a household as being offered the survey if at least one eligible member received the offer, so by construction households with multiple eligible members are more likely to receive the offer. Conditional on the number of eligible household members, however, randomization across individuals ensures that the offer is random at the household level. We checked whether the randomization was successful by regressing our instrument, an indicator for

Table 6.4: Exogeneity of the instrument w.r.t. sample selection

	Dependent variable: indicator for estimation sample
Offered	-0.0209
	(0.0201)
Multiple eligibles	0.00700
	(0.0215)
Constant	0.364***
	(0.0207)
Number of selected HHs	999
N	2,816

Robust standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

being offered the survey, on all socio-demographic variables listed in Table 6.1 (controlling for the presence of multiple household members). All the variables from Table 6.1 are jointly insignificant, with a P-value equal to 0.901. Hence, there is no evidence to suggest that the randomization failed.

Having established that the instrument is uncorrelated with sample selection and that the randomization succeeded in creating a valid instrument, it remains to be shown that the instrument is relevant for our treatment. The first-stage regression shows that the instrument is highly relevant: the F-statistic for the coefficient of the instrument in a model that controls for the presence of multiple eligibles is 4,818. Complete estimation results for the first stage are given in Appendix 6.A.

Main results on saving 6.4.2

Figure 6.1 presents box-and-whiskers diagrams of savings levels (left panel) and savings rates (right panel) by offer-status, separately for couples and singles. This comparison of savings between households that were offered the survey with savings of those that were not constitutes an ITT-analysis, which is valid because of random assignment to the groups. The left figure shows that the median level of savings in households with multiple eligibles is similar regardless of being offered the survey, but that both the 25th and the 75th percentiles are lower for households that did receive the offer. The dispersion

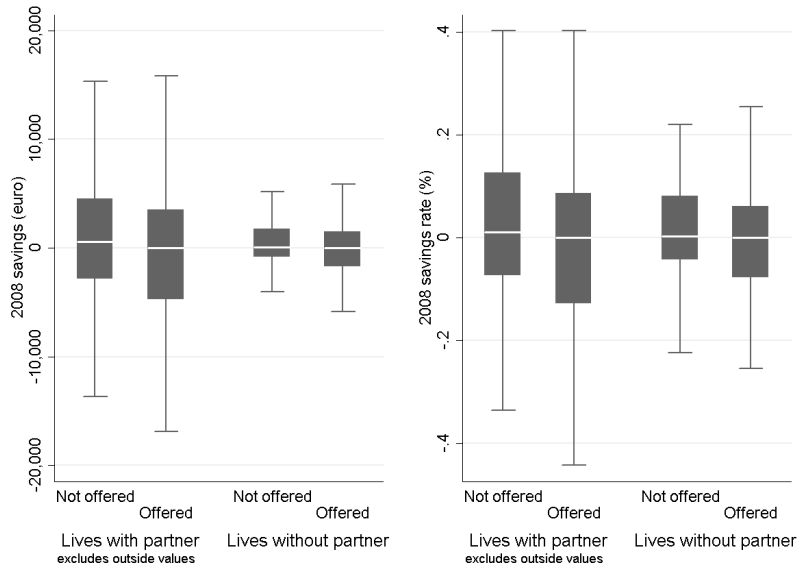


Figure 6.1: Graphical intention-to-treat analysis.

for savings is much less among singles and the differences between offered and non-offered households are smaller. In the right panel of Figure 6.1 we find the same patterns for savings rates: the median savings rate is similar for all households but both the 25th and 75th percentiles are substantially lower for those households that were offered the retirement expenditures survey. In contrast to the levels, for savings rates this difference is present both for couples and singles. Our ITT analysis suggests that the random offer of the survey reduced household savings on both sides of the median.

Table 6.5 presents our main estimation results for the effect of participation in the retirement expenditures survey on savings. The top panel uses 2008 non-housing savings as outcome measure, while the bottom panel explains the 2008 savings rate (non-housing savings divided by household income). The leftmost column shows the estimated coefficients and accompanying standard errors for the treatment dummy in 2SLS models where we instrument survey participation with the random offer of the survey. Participation in the survey caused households to save 1,683 euro less on average during 2008. This is a large effect, especially considering the sample average of 154 euro and the standard deviation of 9,411. When we express savings relative to household income, we also find a significant and negative effect. Survey participation

caused households to save 3.5 percentage points less on average, compared with a sample average of -1% and standard deviation of 19 percentage points.

Table 6.5: The effect of survey participation on savings

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
	Dependent variable: 2008 non-housing savings (thousands of euros)							
Treated	-1.683**	-1.792	-1.193**	-0.644	-0.474	-0.955**	-1.085**	-0.784
	(0.764)	(1.119)	(0.552)	(0.458)	(0.438)	(0.461)	(0.530)	(0.709)
Sample statistics	0.154	-4.935	-2.061	-0.583	0.002	1.060	2.245	4.393
Proportion compliers				0.875				
N				999				
	Dependent variable: 2008 non-housing savings rate (1 = 100%)							
Treated	-0.0351**	-0.0519*	-0.0337**	-0.0224*	-0.00922	-0.0352**	-0.0247	-0.0317
	(0.0144)	(0.0286)	(0.0166)	(0.0135)	(0.0130)	(0.0141)	(0.0153)	(0.0208)
Sample statistics	-0.01	-0.14	-0.07	-0.02	0.00	0.03	0.07	0.13
Proportion compliers				0.875				
N				999				

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

In order to get a better feel for the nature of the effect and its robustness, we check which parts of the savings distribution were affected most strongly by means of IV quantile models. We report estimates for the second up to the eight decile in the remaining columns of Table 6.5. For the level of savings, we find significant and large effects for the third, sixth and seventh deciles. The estimated coefficients for the other deciles are also all negative. For the non-housing savings rate, we find strongly significant effects at the third and sixth decile as well as marginally significant effects at the second and fourth deciles. These estimates show that large parts of the savings distribution were shifted by the survey, with similar effect sizes below and above the median.[3]

The finding that participation in a survey about retirement expenditures reduced savings may seem counterintuitive. One might expect that such survey induces respondents to think about their consumption needs during retirement, which would underline the importance of building a nest egg. However, we

[3]Assets held in bank accounts and risky investments are provided directly by banks to the tax authority and as a result they are probably measured most accurately. Therefore, we also tried yearly savings in those categories as alternative outcome variables. Appendix 6.C presents the estimates and shows that our results are robust, especially for the savings rates. Moreover, we find no evidence that the survey led people to change their ownership status with respect to risky assets.

believe that the finding makes sense in the context of the Dutch pension system. Remember that in the Netherlands 95% of the entitlement to income during retirement is accumulated in public and occupational pension accounts. Savings in those accounts are mandatory for those who are covered and together they provide generous income replacement relative to the final earned wage. Hence, it is reasonable that respondents give the issue more thought and conclude that there is really no need for the accumulation of additional resources to finance their desired expenditures. De Bresser and Knoef (2013) show that around over 60% of the Dutch accumulate enough resources in non-voluntary pensions alone to meet their minimal level of expenditures (the median household exceeds their consumption floor by 23%). In such an institutional context it is not surprising that thinking more about one's consumption needs later in life causes a decline in savings.

6.4.3 Falsification tests

Our identification is based on the randomized distribution of the retirement expenditures survey to a subset of the eligible panel members. This reliance on an actual randomization allows us to cleanly measure the causal effect of interest. In order to establish the credibility of our identification strategy, we ran the same models reported in Table 6.5, but now controlling for all covariates described in Table 6.1. All estimated effects are robust to changing the specification.[4]

As an additional check, we run the same models on 2007 savings. The estimation results are shown in Table 6.6. We find no evidence for any systematic difference in savings behavior prior to the time of the survey, neither in terms of the average level of savings nor any of the deciles. Note that, in contrast to Table 6.5, the coefficients of the various deciles are not even of the same sign.

[4]Estimates available on request.

Table 6.6: Falsification tests

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
		Dependent variable: 2007 non-housing savings (thousands of euros)						
Treated	-0.406	-0.582	0.130	0.158	-0.090	-0.393	-0.684	-1.792
	(0.749)	(0.889)	(0.460)	(0.414)	(0.417)	(0.478)	(0.832)	(1.251)
Proportion compliers					0.866			
N					1,014			
		Dependent variable: 2007 non-housing savings rate (1 = 100%)						
Treated	-0.0136	-0.00961	-0.00279	0.00197	-0.00198	-0.00296	-0.0211	-0.0257
	(0.0147)	(0.0249)	(0.0133)	(0.0121)	(0.0122)	(0.0136)	(0.0218)	(0.0298)
Proportion compliers					0.866			
N					1,014			

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Effect heterogeneity 6.4.4

So far we have shown that participation in the retirement expenditures survey reduced average household non-housing saving, both in absolute terms and relative to household income. Next we investigate effect heterogeneity. One approach would be to run IV analyses on subsamples, but many variables that could be used for interesting splits of the sample are correlated. Examples are income and education and income and age. Therefore, we prefer a regression approach, where we regress savings on an indicator equal to one if a household was offered the survey (our instrument in the preceding analysis); all covariates shown in Table 6.1; and interaction terms of covariates with the survey offer. Hence, we can interpret the results in this section as (heterogeneous) intention-to-treat effects. We investigate heterogeneity along the lines of income, education and age. Note that, for reasons of sample size, the specification does not contain dummies for all cells shown in Table 6.7, but only interactions of the separate variables with the treatment indicator. The upper panel of Table 6.7 displays coefficient estimates for the main effect and interaction terms of the model with the level of savings as the dependent variable. According to these estimates, the offer of the survey increased average savings of young, income-poor households that are poorly educated by 2,910 euro per year. We find strong evidence for differential effects along the lines of age and education: households with more highly educated or older heads

reduced their savings more after filling out the survey. The lower panel of Table 6.7 shows the differences in savings between offered and non-offered households for subsamples defined along age, education and income categories. Offered households with poorly educated heads who are younger than 40 actually saved 2,900-3,000 euro more than non-offered households with the same education and age. Households in the highest education category saved less regardless of the age of the head and their household income, but we find the strongest effects for older households: the difference is 2,400 euro for households below 40; 3,500-3,600 for age 40-54; and 6,800-7,000 for households aged 55 or older.

The magnitudes of the intention-to-treat effects on absolute savings are similar for households above and below median income, the interaction term of being offered the survey with income is not significant, so we should find smaller effects for high-income households if we divide savings by household income. Table 6.8 shows exactly that pattern: highly educated households with incomes below the overall median reduced savings by 8-15 percentage points depending on age, while high income households only reduced savings by 3-10 percentage points.

In their analysis of the differences in pension entitlements between socio-economic groups, De Bresser and Knoef (2013) show that education and age are positively correlated with forecasted (occupational) pension entitlements at age 65. Households in which at least one partner has finished a university degree have 37% higher entitlements on average compared to households in which neither spouse finished secondary school. Our analysis of heterogeneous effects indicates that those groups that can look forward to the most generous pensions, namely older and more highly educated households, cut their savings the most after participating in the survey on retirement expenditures. The pension-poor young and poorly educated increased their savings marginally as a result of the survey. These patterns suggest that the survey made respondents reflect on their preparedness for retirement, after which some groups concluded that they could afford to save a little less while others reached the opposite conclusion.

In addition to the effect heterogeneity reported in Tables 6.7 and 6.8, we also checked whether the intention-to-treat effect of the survey offer differs depending on whether the individual(-s) who received the offer is a husband;

Table 6.7: Heterogeneous intention-to-treat effects – level of savings

	Dependent variable: 2008 savings (thousands of euros)
Offered	2.910**
	(1.321)
Offered × HH inc. high	0.160
	(1.225)
Offered × educ. middle	-1.027
	(1.195)
Offered × educ. high	-5.315***
	(1.475)
Offered × age 40-54	-1.243
	(1.360)
Offered × age 55+	-4.624***
	(1.779)
Controls	Yes
R-squared	0.0749
N	999

	Heterogeneous effects					
	Income below median			Income above median		
	Education			Education		
	Low	Middle	High	Low	Middle	High
Age <40	2.910**	1.883*	-2.405**	3.070**	2.043	-2.245
	(1.321)	(1.056)	(1.157)	(1.535)	(1.339)	(1.601)
Age 40-54	1.667	0.640	-3.648***	1.827	0.800	-3.488**
	(1.388)	(1.252)	(1.346)	(1.397)	(1.288)	(1.565)
Age 55+	-1.714	-2.741*	-7.029***	-1.554	-2.581	-6.869***
	(1.149)	(1.538)	(1.813)	(1.755)	(2.047)	(2.379)

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 6.8: Heterogeneous intention-to-treat effects – savings rate

	Dependent variable: 2008 savings rate (1 = 100%)
Offered	0.0213
	(0.0300)
Offered × HH inc. high	0.0482*
	(0.0268)
Offered × educ. middle	-0.0150
	(0.0319)
Offered × educ. high	-0.100***
	(0.0319)
Offered × age 40-54	-0.0200
	(0.0294)
Offered × age 55+	-0.0706**
	(0.0330)
Controls	Yes
R-squared	0.0574
N	999

	Heterogeneous effects					
	Income below median			Income above median		
	Education			Education		
	Low	Middle	High	Low	Middle	High
Age <40	0.0213	0.00626	-0.0789***	0.0695*	0.0545*	-0.0307
	(0.0300)	(0.0260)	(0.0290)	(0.0377)	(0.0304)	(0.0298)
Age 40-54	0.00129	-0.0138	-0.0989***	0.0495	0.0344	-0.0507*
	(0.0300)	(0.0308)	(0.0330)	(0.0319)	(0.0280)	(0.0270)
Age 55+	-0.0493*	-0.0643*	-0.150***	-0.00111	-0.0162	-0.101***
	(0.0280)	(0.0348)	(0.0342)	(0.0349)	(0.0371)	(0.0336)

Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

a wife; both husband and wife; a single male; or a single female. However, we find no evidence for different intention-to-treat effects depending on whether men or women were offered the survey. Moreover, we investigated whether the intensity of the effect was different for respondents that took longer than the median time to answer the questions, but found no evidence to suggest that completion time affected the effect size. Estimates are available on request.

Evidence from survey data 6.4.5

Though our use of administrative data to reduce measurement error adds to the reliability of the analysis, it also limits our scope. As explained in section 6.2.3, we do not observe investments in private pension schemes since those funds are tax-exempt during the accumulation phase. However, one might expect that the negative effect of the survey on household non-housing savings is indicative of a re-allocation of assets to accounts that are linked explicitly to retirement. In order to investigate that possibility, we used the LISS assets module to look specifically at assets invested in private pensions. We looked for an effect on ownership, changes in ownership, balance conditional on ownership and unconditional balances and carried out the analysis at the level of the individual and of the household. However, we do not find any evidence for a change in savings in voluntary pension accounts.[5] Unfortunately, we cannot rule out the possibility of measurement error obscuring an actual effect.

We also used self-reports in the LISS to investigate whether participation in the retirement expenditures survey altered the subjective outlook of respondents in ways that are relevant to savings. We find some evidence that survey participation made respondents more satisfied with the economic situation in the Netherlands. However, that effect is not robust to limiting the sample to those individuals that could be matched with administrative records.

Conclusion 6.5

In this paper we show that participating in a non-informative survey on expenditures after retirement led Dutch households to save significantly less. Our

[5] Estimates available on request.

analysis uses administrative wealth data to calculate clean measures of savings that are not contaminated by the reporting styles of survey respondents. Participation in the survey is instrumented by randomized assignment to treatment conditions, so our estimates are unaffected by endogenous compliance. Estimated effects are large: the survey caused households to save 1,700 euro, or 3.5 percentage points relative to disposable income, less during 2008 (sample means: 154 euro and -1%). These effects are robust to various trimming rules and decile IV models show that they occur on both sides of the median of the distribution of savings. Furthermore, falsification checks reveal no effect on savings prior to the survey, supporting the validity of our identification strategy. We find evidence for heterogenous effects, with the strongest impact among highly educated and older households.

The decline in savings after participating in a survey on consumption during retirement makes sense in the particular institutional context of the Netherlands in 2008, which was characterized by mandatory pensions that were generous, replacing 70% of final income on average, and covered nearly the entire population. Furthermore, De Bresser and Knoef (2013) show that older and more educated households can look forward to higher (occupational) pensions, which fits with the effect heterogeneity we document.

In terms of the mechanism behind these survey effects, limited attention seems to provide a coherent interpretation for our results. The heterogeneity in treatment effects we document is consistent with the survey acting as an attention shock. After being reminded of the tradeoff between current and future consumption, most individuals conclude that their pensions alone will provide them with an adequate income in retirement, resulting in a negative effect on average non-pension savings. However, some groups, namely the young and poorly educated, foresee that they may not be so lucky and do not cut back on savings (or even save a little more). In that sense, our results are comparable to those of Stango and Zinman (2011), who show that financial behavior can be influenced even by questions that do not directly concern that behavior.

Though we do not expect to find similar effects in different institutional contexts, the general point of household financial behavior being affected by participation in household surveys is of considerable importance to empirical economists. It may lead panels that are representative in terms of demographics

to behave in non-representative ways. As a result the external validity of any study based on that data would be compromised. If future research confirms that financial decisions of households are susceptible to survey effects, that would be an important reason for researchers to prefer administrative data whenever available. Also, as noted by Zwane et al. (2011), the presence of survey effects suggests that large panels that participate in surveys infrequently may be a better way to get statistical power than small samples that fill out questionnaires on a weekly basis.

Our use of a randomized survey module to identify a causal effect highlights the potential of household panels as laboratories for economic experiments. Randomized questionnaires can be used to investigate how individuals update their expectations and how financial education affects perceptions and behavior. Researchers can use those surveys to identify causal effects of interventions and test economic theories.

One limitation of the present study is the fact that our wealth data only allow us to compute savings during 2007 and 2008. Therefore, we do not know how long-lived the survey effect is. Naturally we will investigate the durability of the effects when more data becomes available. Documenting the range of behaviors that are affected by surveys and the subpopulations that are most sensitive are fruitful areas for future research. Our evidence suggests that economists should follow psychologists and other social scientists and acknowledge the relevance of survey effects, even if that means that our favorite method of data collection is simultaneously more versatile and less innocent than we would like to believe.

Acknowledgements 6.6

The authors thank Arthur van Soest, Frederic Vermeulen, Marcel Das, Pierre-Car Michaud, Rob Alessie, Martin Salm and the participants in seminars at Tilburg University, the University of Groningen, the University of Stirling, Mannheim University, Ludwig Maximilian University of Munich and Statistics Canada for their useful comments.

6.A First stage

Table 6.9: First stage

	Dependent variable: HH treated
HH offered	0.879***
	(0.0127)
Multiple eligibles	-0.0376**
	(0.0159)
Constant	0.0231**
	(0.00988)
R squared	0.688
F(1, n-(k+1))	4,818.37***
N	999

Robust standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Estimates under different trimming rules 6.B

Table 6.10: Robustness checks with different trimming rules

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
		Sample trimmed at savings rates between -100% and +100%						
		Dependent variable: 2008 non-housing savings (thousands of euros)						
Treated	-1.862*	-1.282	-1.498*	-1.017*	-0.124	-0.913*	-1.085*	-0.510
	(1.085)	(1.440)	(0.809)	(0.544)	(0.507)	(0.495)	(0.575)	(0.829)
Proportion compliers				0.873				
N				1,124				
		Dependent variable: 2008 non-housing savings rate (1 = 100%)						
Treated	-0.0343	-0.0803*	-0.0464**	-0.0284*	-0.00260	-0.0281*	-0.0243	-0.0220
	(0.0216)	(0.0418)	(0.0217)	(0.0157)	(0.0147)	(0.0149)	(0.0162)	(0.0234)
Proportion compliers				0.873				
N				1,124				
		Sample trimmed at savings rates between -75% and +75%						
		Dependent variable: 2008 non-housing savings (thousands of euros)						
Treated	-1.900**	-0.828	-1.306*	-0.828	-0.0301	-0.975**	-1.284**	-0.771
	(0.868)	(1.378)	(0.722)	(0.510)	(0.483)	(0.481)	(0.581)	(0.804)
Proportion compliers				0.88				
N				1,080				
		Dependent variable: 2008 non-housing savings rate (1 = 100%)						
Treated	-0.0356*	-0.0615*	-0.0399**	-0.0270*	-0.00139	-0.0332**	-0.0277*	-0.0273
	(0.0182)	(0.0364)	(0.0200)	(0.0151)	(0.0141)	(0.0147)	(0.161)	(0.0226)
Proportion compliers				0.88				
N				1,080				

[a] For decile models we report unconditional treatment effects.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

6.C Financial savings (savings accounts and risky assets)

In this appendix we present the results of an analogous analysis to that in the main text, which uses a different outcome variable. Here we analyze yearly financial savings, defined as the yearly differences of the sum of assets in savings accounts and risky assets. Information on savings accounts and risky assets holdings is provided directly by banks to the tax authority, whereas assets in other categories are reported by households. Though misreporting those other assets is a criminal offense, direct provision of account information by banks means that financial assets are probably measured even more precisely. Therefore, we report complementary results in Table 6.11 to substantiate our main findings.

Table 6.11: Alternative outcome variable: financial savings (savings accounts and risky assets)

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
				2008 financial savings[b]				
				Dependent variable: financial savings (thousands of euros)				
Treated	-1.055	-1.818	-0.836*	-0.528	-0.0768	-0.852*	-0.833	-0.751
	(0.646)	(1.119)	(0.508)	(0.430)	(0.419)	(0.439)	(0.521)	(0.677)
Proportion compliers				0.865				
N				1,043				
				Dependent variable: financial savings rate (1 = 100%)				
Treated	-0.0248*	-0.0534**	-0.0308**	-0.0190	-0.00207	-0.0269**	-0.0217	-0.0209
	(0.0139)	(0.0273)	(0.0158)	(0.0128)	(0.0125)	(0.0129)	(0.0142)	(0.0177)
Proportion compliers				0.865				
N				1,043				
				Falsification check: 2007 financial savings[b]				
				Dependent variable: financial savings (thousands of euros)				
Treated	-0.040	-1.305*	-0.234	0.000	-0.0236	-0.198	-0.797	-1.005
	(0.639)	(0.738)	(0.416)	(0.373)	(0.378)	(0.431)	(0.767)	(1.081)
Proportion compliers				0.862				
N				1,066				
				Dependent variable: financial savings rate (1 = 100%)				
Treated	-0.0110	-0.0254	-0.0107	0.00135	9.92e-06	-0.00138	-0.0164	-0.0283
	(0.0137)	(0.0222)	(0.0124)	(0.0109)	(0.0109)	(0.0121)	(0.0192)	(0.0261)
Proportion compliers				0.862				
N				1,066				

[a] For decile models we report unconditional treatment effects.
[b] Savings are trimmed at -50% and +50% of disposable household income.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 6.11 contains estimates for 2008 financial savings (top panel) and the corresponding falsification check using 2007 financial savings (bottom panel).

Analogously to our analysis of non-housing savings in the main text, we trim households with financial savings below -50% or above 50% of disposable household income. When we look at the *level* of financial savings, we find smaller, less significant results than for non-housing savings: only the third and sixth decile show marginally significant effects of survey participation around -800/-850 euro per year. However, for the financial savings *rate*, defined as yearly savings divided by yearly household disposable income, there is stronger evidence of a survey effect. The estimated mean effect is marginally significant and equal to -2.48 percentage points. Moreover, the decile treatment effects indicate that this survey effect is most pronounced in the bottom of the savings rates distribution: the second and third deciles are reduced by 3-5 percentage points. The bottom panel of Table 6.11 shows that financial savings, like non-housing savings, pass the falsification test: the estimated treatment effects prior to the treatment are generally small compared to those after the treatment and not significantly different from zero.

We also checked whether there is any evidence that survey participation caused respondents to reallocate their portfolio towards or away from risky assets, but we find no evidence to support this possibility at the extensive margin. Instrumental variables models show that there is no significant effect of survey participation on the likelihood of owning risky assets during any of the years 2007-2009. Because it is impossible to verify to what extent changes in the value of the stock of risky assets result from the decision to buy or sell rather than from variation in stock and bond prices, we ran the analysis on changes in the balance of bank accounts. Table 6.12 contains the corresponding estimates and shows that the result of a negative effect of the survey on savings is corroborated by the quantile models ($p = 0.103$ in the model for the mean savings rate). However, the part of the distribution of savings that is affected is different compared to the combined bank accounts/risky assets outcome analyzed above. When we combined assets from those categories, we found the largest effects below the median, at the third and fourth decile. If we focus exclusively on funds in bank accounts, the largest effects occur at the median and sixth and seventh deciles.

Table 6.12: Alternative outcome variable: savings in bank accounts (without risky assets)

		Deciles[a]						
	Mean	0.20	0.30	0.40	0.50	0.60	0.70	0.80
				2008 savings in bank accounts[b]				
				Dependent variable: savings in bank accounts (thousands of euros)				
Treated	-0.840	-0.412	-0.712	-0.429	-0.774*	-0.939**	-0.989*	-0.218
	(0.635)	(1.131)	(0.479)	(0.401)	(0.398)	(0.427)	(0.506)	(0.689)
Proportion compliers					0.866			
N					1,078			
				Dependent variable: savings rate for bank accounts (1 = 100%)				
Treated	-0.0219	7.81e-06	-0.0220	-0.0149	-0.0253**	-0.0311**	-0.0227*	-0.0199
	(0.0134)	(0.0272)	(0.0142)	(0.0118)	(0.0116)	(0.0122)	(0.0133)	(0.0174)
Proportion compliers					0.866			
N					1,078			
				Falsification check: 2007 savings in bank accounts[b]				
				Dependent variable: savings in bank accounts (thousands of euros)				
Treated	-0.00686	-0.533	0.138	0.263	0.115	1.76e-04	-0.156	-1.047
	(0.673)	(0.699)	(0.389)	(0.351)	(0.356)	(0.410)	(0.716)	(1.094)
Proportion compliers					0.857			
N					1,076			
				Dependent variable: savings rate for bank accounts (1 = 100%)				
Treated	-0.00271	0.0116	0.00292	0.00836	0.00264	0.00401	-0.00831	-0.0174
	(0.0136)	(0.0230)	(0.0113)	(0.0102)	(0.0104)	(0.0117)	(0.0183)	(0.0283)
Proportion compliers					0.857			
N					1,076			

[a] For decile models we report unconditional treatment effects.
[b] Savings are trimmed at -50% and +50% of disposable household income.
We control for the presence of multiple eligibles.
Standard errors in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Bibliography

AFM (2010): Geef Nederlanders pensioeninzicht: werken aan vertrouwen door dichten van de verwachtingskloof. Autoriteit Financiële Markten, Amsterdam. Cited on pages 21 and 57.

Alessie, R., A. Kapteyn, and F. Klijn (1997a): Mandatory pensions and personal savings in the Netherlands. *De Economist*, 145(3):291–324. Cited on page 157.

Alessie, R., A. Lusardi, and T. Aldershof (1997b): Income and wealth over the life-cycle: evidence from panel data. *Review of Income and Wealth*, 43(1):1–32. Cited on page 157.

Andrews, D. (1988): Chi-square diagnostic tests for econometric models. *Journal of Econometrics*, 37(1):135–156. Cited on pages 74 and 105.

Angrist, J., G. Imbens, and D. Rubin (1996): Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455. Cited on page 215.

Baetschmann, G., K. Staub, and R. Winkelmann (2011): Consistent estimation of the fixed effects ordered logit model. Working paper 4, Department of Economics, University of Zurich. Cited on page 22.

Banks, J., R. Blundell, and S. Tanner (1998): Is there a retirement-savings puzzle? *American Economic Review*, 88(4):769–788. Cited on page 156.

Bellemare, C., L. Bissonnette, and S. Kröger (2012): Flexible approximation of subjective expectations using probability questions. *Journal of Business and Economic Statistics*, 30(1):125–131. Cited on pages 12, 20, 45, 117, and 138.

Beresteanu, A. and F. Molinari (2008): Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814. Cited on page 114.

Beresteanu, A., F. Molinari, and D. Steeg Morris (2010): Asymptotics for partially identified models in STATA. Stata code for BLP on interval data. Cited on page 132.

Beshears, J., J. Choi, D. Laibson, and B. Madrian (2009): The importance of default options in retirement savings outcomes: evidence from the United States. In J. Brown, J. Liebman, and D. A. Wise, editors, *Social Security Policy in a Changing Environment*, pages 167–195. Univ. of Chicago Press, Chicago. Cited on page 156.

Biltoft-Jensen, A., J. Matthiesen, L. Rasmussen, S. Fagt, M. Groth, and O. Hels (2009): Validation of the Danish 7-day pre-coded food diary among adults: energy intake v. energy expenditure and recording length. *British Journal of Nutrition*, 102:1838–1846. Cited on page 207.

Binswanger, J. and D. Schunk (2012): What is an adequate standard of living during retirement? *Journal of Pension Economics and Finance*, 11(2):203–222. Cited on pages 159, 165, 175, and 211.

Binswanger, J., D. Schunk, and V. Toepoel (2013): Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, page nft030. Cited on page 159.

Bissonnette, L. and J. De Bresser (2013): Eliciting subjective survival curves: lessons from partial Identification. Mimeo, Tilburg University. Cited on page 109.

Bissonnette, L., M. Hurd, and P.-C. Michaud (2011): Individual survival curves combining subjective and actual mortality risks. Mimeo, Tilburg University. Cited on page 113.

Bottazzi, R., T. Jappelli, and M. Padula (2006): Retirement expectations, pension reforms, and their impact on private wealth accumulation. *Journal of Public Economics*, 90(12):2187–2212. Cited on page 12.

Bound, J., C. Brown, and N. Mathiowetz (2001): Measurement error in survey data. In J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 5, pages 3705–3843. Elsevier, North-Holland, Amsterdam. Cited on pages 168 and 211.

Bovenberg, A. and L. Meijdam (2001): The Dutch pension system. In A. H. Borsch-Supan and M. Miegel, editors, *Pension Reform in Six Countries*, pages 39–67. Springer, New York. Cited on pages 8, 16, 54, 155, 165, 208, and 214.

Brooks, C. and J. Manza (2007): *Why Welfare States Persist: the Importance of Public Opinion in Democracies*. University of Chicago Press, Chicago. Cited on pages 2 and 11.

Brugiavini, A. and M. Padula (2001): Too much for retirement? Saving in Italy. *Research in Economics*, 55(1):39–60. Cited on page 157.

Bruine de Bruin, W., B. Fischhoff, S. Millstein, and B. Halpern-Felsher (2000): Verbal and numerical expressions of probability: "It's a fifty-fifty chance". *Organizational Behavior and Human Decision Processes*, 81(1):115–131. Cited on pages 3 and 54.

Cremer, H. and P. Pestieau (2000): Reforming our pension system: is it a demographic, financial or political problem? *European Economic Review*, 44(4-6):974–983. Cited on pages 2 and 11.

Crossley, T., J. De Bresser, L. Delaney, and J. Winter (2013): Can survey participation alter household financial behavior? Mimeo, Tilburg University. Cited on page 205.

Das, M. and B. Donkers (1999): How certain are Dutch households about future income? An empirical analysis. *Review of Income and Wealth*, 45(3):325–339. Cited on page 15.

Das, M., V. Toepoel, and A. Van Soest (2011): Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods & Research*, 40(1):32–56. Cited on page 207.

Das, M. and A. Van Soest (1999): A panel data model for subjective information on household income growth. *Journal of Economic Behavior and Organization*, 40(4):409–426. Cited on page 22.

De Bresser, J. and M. Knoef (2013): Can the Dutch meet their own retirement expenditure goals? Mimeo, Tilburg University. Cited on pages 153, 209, 212, 224, 226, and 230.

De Bresser, J. and A. Van Soest (2013a): Retirement expectations and satisfaction with retirement provisions. Forthcoming in *Review of Income and Wealth*. Cited on pages 11, 65, 112, and 128.

——— (2013b): Survey response in probabilistic questions and its impact on inference. Forthcoming in *Journal of Economic Behavior and Organization*. Cited on pages 53, 112, and 120.

Delavande, A. and S. Rohwedder (2011): Individuals' uncertainty about future social security benefits and portfolio choice. *Journal of Applied Econometrics*, 26(3):498–519. Cited on page 56.

DellaVigna, S. (2009): Psychology and economics: evidence from the field. *Journal of Economic Literature*, 47(2):315–372. Cited on pages 7 and 205.

Dholakia, U. (2010): A critical review of question-behavior effect research. *Review of Marketing Research*, 7(7):145–197. Cited on page 206.

Dijkhuizen, C., J. Bartel, C. Caminada, N. Gubbels, P. Kavelaars, B. Kuipers, and C. Smits (2013): Naar een activerender belastingstelsel. Committee on income tax and allowances, final report. Cited on page 168.

Dominitz, J. (1996): A comparison of subjective expectations elicitation methods in the health and retirement study, the panel of income dynamics, and the survey of economic expectations. HRS/AHEAD Working Paper: 96-043. Cited on page 14.

——— (1998): Earnings expectations, revisions and realizations. *Review of Economics and Statistics*, 80(3):374–388. Cited on pages 3, 14, 15, 56, 110, and 138.

Dominitz, J. and C. Manski (1997): Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92(439):855–867. Cited on pages 12, 14, 20, 43, 45, 54, 56, 58, 65, 66, 110, 116, and 138.

——— (2006): Measuring pension benefit expectations probabilistically. *Labour*, 20(2):201–236. Cited on pages 14, 53, 56, 57, and 110.

Engen, E., W. Gale, C. Uccello, C. Carroll, and D. Laibson (1999): The adequacy of household saving. *Brookings Papers on Economic Activity*, 1999(2):65–187. Cited on page 156.

Ferrer-i-Carbonell, A. and P. Frijters (2004): How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, 114(497):641–659. Cited on page 21.

Fitzsimmons, G. and S. Moore (2008): Should we ask our children about sex, drugs and rock & roll? Potentially harmful effects of asking questions about risky behaviors. *Journal of Consumer Psychology*, 18(2):82–95. Cited on page 206.

Fitzsimmons, G. and B. Shiv (2001): Nonconscious and contaminative effects of hypothetical questions on subsequent decision making. *Journal of Consumer Research*, 28(2):224–238. Cited on page 206.

Frölich, M. and B. Melly (2010): Estimation of quantile treatment effects with Stata. *The Stata Journal*, 10(3):423–457. Cited on page 215.

——— (2013): Unconditional quantile treatment effects under endogeneity. *Journal of Business and Economic Statistics*, just-accepted. Cited on page 215.

Gan, L., M. Hurd, and D. McFadden (2005): Individual subjective survival curves. In D. A. Wise, editor, *Analyses in the Economics of Aging*. University of Chicago Press, Chicago. Cited on page 135.

Glenn, N. (1998): The course of marital success and failure in five American 10-year marriage cohorts. *Journal of Marriage and Family*, 60(3):569–576. Cited on page 207.

Gustman, A. and T. Steinmeier (2005): Imperfect knowledge of social security and pensions. *Industrial Relations: A Journal of Economy and Society*, 44(2):373–397. Cited on page 57.

Haveman, R., K. Holden, A. Romanov, and B. Wolfe (2007): Assessing the maintenance of savings sufficiency over the first decade of retirement. *International Tax and Public Finance*, 14(4):481–502. Cited on pages 157, 158, and 175.

Hudomiet, P., G. Kézdi, and R. Willis (2011): Stock market crash and expectations of American households. *Journal of Applied Econometrics*, 26(3):393–415. Cited on page 57.

Hurd, M. (2009): Subjective probabilities in household surveys. *Annual Review of Economics*, 1:543–562. Cited on pages 56 and 113.

Hurd, M. and K. McGarry (1995): Evaluation of the subjective probabilities of survival in the Health and Retirement Study. *Journal of Human Resources*, 30(Special Issue on the Health and Retirement Study: Data Quality and Early Results):S268–S292. Cited on page 113.

——— (2002): The predictive validity of subjective probabilities of survival. *Economic Journal*, 112(482):966–985. Cited on pages 14, 15, and 113.

Hurd, M., J. Smith, and J. Zissimopoulos (2004): The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, 19(6):761–775. Cited on page 109.

Hyde, M., J. Dixon, and G. Drover (2007): Assessing the capacity of pension institutions to build and sustain trust: a multidimensional conceptual framework. *Journal of Social Policy*, 36(3):457–475. Cited on page 18.

Imbens, G. and C. Manski (2004): Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857. Cited on pages 114 and 132.

Johansson, F. and A. Klevmarken (2007): Comparing register and survey wealth data. Mimeo, Department of Economics, Uppsala University. Cited on page 168.

Juster, F. and R. Suzman (1995): An overview of the health and retirement study. *Journal of Human Resources*, 30(Special Issue on the Health and Retirement Study: Data Quality and Early Results):S7–S56. Cited on page 113.

Kapeyn, A. and K. De Vos (2008): Social security and retirement in the Netherlands. In J. Gruber and D. A. Wise, editors, *Social Security and Retirement Around the World*, pages 269–304. University of Chicago Press. Cited on pages 214 and 215.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2012): Getting to the top of mind: How reminders increase saving. No. w16205. National Bureau of Economic Research. Cited on pages 206 and 209.

Kleinjans, C. and A. Van Soest (2013): Non-response and focal point answers to subjective probability questions. Forthcoming in *Journal of Applied Econometrics*. Cited on pages 3, 54, 57, 65, 67, 70, and 75.

Knoef, M., J. Been, R. Alessie, K. Caminada, K. Goudswaard, and A. Kalwij (2013): Measuring retirement savings adequacy; a first multi-pillar approach in the Netherlands. Netspar Design Paper, forthcoming. Cited on pages 157 and 173.

Krueger, A. and D. Schkade (2008): The reliability of subjective well-being measures. *Journal of Public Economics*, 92(8-9):1833–1845. Cited on page 15.

Kutlu, V. and A. Kalwij (2012): Individuals' survival expectations and actual mortality. Mimeo, Utrecht University. Cited on pages 6, 114, 128, 132, 135, 138, and 139.

Lillegaard, I., E. Løken, and L. Andersen (2007): Relative validation of a precoded food diary among children, under-reporting varies with reporting day and time of the day. *European Journal of Clinical Nutrition*, 61(1):61–68. Cited on page 207.

Luchak, A. and I. Gellatly (2002): How pension accrual affects job satisfaction. *Journal of Labor Research*, 23(1):145–162. Cited on pages 2, 11, and 15.

Lynch, J. and M. Myrskylä (2009): Always the third rail - pension income and policy preferences in European democracies. *Comparative Political Studies*, 42(8):1068–1097. Cited on pages 2, 12, and 15.

Manski, C. (2002): *Partial Identification of Probability Distributions*. Springer, New York. Cited on pages 14 and 56.

——— (2004): Measuring expectations. *Econometrica*, 75(5):1329–1376. Cited on pages 12, 14, 36, 53, 56, 88, and 110.

Manski, C. and F. Molinari (2010): Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics*, 28(2):219–231. Cited on pages 112, 114, 120, 121, 125, and 139.

Michie, S., S. Churchill, and R. West (2011a): Identifying evidence-based competences required to deliver behavioural support for smoking cessation. *Annals of Behavioral Medicine*, 41(1):59–70. Cited on page 206.

Michie, S., N. Hyder, A. Walia, and R. West (2011b): Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. *Addictive Behaviors*, 36(4):315–319. Cited on page 206.

Mitchell, O. (1988): Worker knowledge of pension provisions. *Journal of Labor Economics*, 6(1):21–39. Cited on page 57.

Mitchell, O. and J. Moore (1998): Can Americans afford to retire? New evidence on retirement saving adequacy. *Journal of Risk and Insurance*, 65(3):371–400. Cited on page 156.

O'Donnell, O. and P. Tinios (2003): The politics of pension reform: lessons from public attitudes in Greece. *Political Studies*, 51(2):262–281. Cited on pages 2, 11, and 12.

Peracchi, F. and V. Perotti (2011): Subjective survival probabilities and life tables: evidence from Europe. EIEF working paper 16/10. Cited on pages 6 and 111.

Perozek, M. (2008): Using subjective expectations to forecast longevity: do survey respondents know something we don't know? *Demography*, 45(1):95–113. Cited on pages 5, 6, 110, 111, 113, 116, 128, 135, 138, and 139.

Podsakoff, P., S. MacKenzie, J. Lee, and N. Podsakoff (2003): Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5):879–903. Cited on page 22.

Scholz, J., A. Seshadri, and S. Khitatrakun (2006): Are Americans saving optimally for retirement? *Journal of Political Economy*, 114(4):607–643. Cited on page 157.

Sherman, S. (1980): On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39(2):211–222. Cited on page 206.

Siermann, C., P. Van Teeffelen, and L. Urlings (2004): Equivalentiefactoren 1995-2000: Methode en belangrijkste uitkomsten. Statistics Netherlands, Sociaal-economische trends, 3. Cited on page 165.

Skinner, J. (2007): Are you sure you're saving enough for retirement? *Journal of Economic Perspectives*, 21(3):59–80. Cited on page 156.

Smith, V. K., J. D. Taylor, and F. Sloan (2001): Longevity expectations and death: can people predict their own demise? *American Economic Review*, 91(4):1126–1134. Cited on page 114.

Soltow, L. and J. Van Zanden (1998): *Income and Wealth Inequality in The Netherlands 16th-20th Century*. Het Spinhuis, Amsterdam. Cited on page 65.

Spangenberg, E. (1997): Increasing health club attendance through self-prophecy. *Marketing Letters*, 8(1):23–31. Cited on page 206.

Stango, V. and J. Zinman (2011): Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. No. w17028. National Bureau of Economic Research. Cited on pages 7, 206, 209, and 230.

Statistics Netherlands, T. (2012): Welvaart in Nederland - inkomen, vermogen en bestedingen van huishoudens en personen. Cited on page 219.

Tamer, E. (2010): Partial identification in econometrics. *Annual Review of Economics*, 2(1):167–195. Cited on page 114.

Train, K. (2003): *Discrete Choice Methods With Simulation*. Cambridge University Press, Cambridge. Cited on pages 71 and 93.

Van der Laan, J. (2009): Representativity of the LISS panel. Statistics Netherlands Discussion paper 09041. Cited on page 159.

Van Groezen, B., H. Kiiver, and B. Unger (2009): Explaining Europeans' preferences for pension provision. *European Journal of Political Economy*, 25(2):237–249. Cited on page 15.

Van Landeghem, B. (2012): Panel conditioning and self-reported satisfaction: evidence from international panel data and repeated cross-sections. SOEPpapers on Multidisciplinary Panel Data Research. Cited on page 207.

Van Praag, B., P. Frijters, and A. Ferrer-i-Carbonell (2003): The anatomy of subjective well-being. *Journal of Economic Behavior and Organization*, 51(1):29–49. Cited on pages 2, 11, 15, and 18.

Van Santen, P., R. Alessie, and A. Kalwij (2012): Probabilistic survey questions and incorrect answers: retirement income replacement rates. *Journal of Economic Behavior and Organization*, 82(1):267–280. Cited on pages 57, 61, 71, 84, 85, and 94.

Venti, S. and D. Wise (1997): The wealth of cohorts: retirement saving and the changing assets of older Americans. In S. Schieber and J. Shoven, editors, *Public Policy Towards Pensions*. Twentieth Century Fund and MIT Press. Cited on page 156.

Yakoboski, P. and J. Dickemper (1997): Increased saving but little planning: results of the 1997 retirement confidence survey. *EBRI Issue Brief*, 191:1–23. Cited on page 157.

Zafar, B. (2011): Can subjective expectations data be used in choice models? Evidence on cognitive biases. *Journal of Applied Econometrics*, 26(3):520–544. Cited on page 56.

Zwane, A., J. Zinman, E. V. Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Gine, E. Duflo, F. Devoto, B. Crepon, and A. Banerjee (2011): Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences*, 108(1):1821–1826. Cited on pages 207, 209, and 231.