Network for Studies on Pensions, Aging and Retirement

Netspar

Netspar DISCUSSION PAPERS

Peter van Santen, Rob Alessie and Adriaan Kalwij

Probabilistic Survey Questions and Incorrect Answers

Retirement Income Replacement Rates

Discussion Paper 02/2011-011

Probabilistic Survey Questions and Incorrect Answers: Retirement Income Replacement Rates

Peter van Santen[*], Rob Alessie[†], Adriaan Kalwij[‡]

February 21, 2011

Abstract

We use probabilistic survey questions to analyze the determinants of individuals' expected value of and the uncertainty in the retirement income replacement rate. We find that the expected replacement rate is U-shaped in age and is substantially lower for higher-educated individuals. Uncertainty in pension income decreases with age, but increases with education. An important aim of this paper is to investigate if the use of probabilistic survey questions yields an endogenous sample selection by removing individuals that provide answers incompatible with probability theory. We find that not accounting for endogenous sample selection biases the results toward more pessimistic expectations and excess uncertainty in the replacement rate. These biases are most prevalent for less-educated individuals.

Keywords: Probabilistic survey questions, Retirement income replacement rate, Uncertainty, Sample selection bias
JEL: H55, D91, E21

[*]Faculty of Economics and Business, University of Groningen, and Netspar. Correspondence: PO Box 800, 9700 AE Groningen, the Netherlands, tel. 00-31-50-363-6328, fax 00-31-50-363-7337, p.c.van.santen@rug.nl.

[†]Faculty of Economics and Business, University of Groningen, Netspar, and Tinbergen Institute, r.j.m.alessie@rug.nl.

[‡]Utrecht University School of Economics, and Netspar, a.s.kalwij@uu.nl.

Data provision by Corrie Vis (CentERData) and Arthur van Soest (Tilburg University) is gratefully acknowledged. We thank Esen Erdogan Ciftci and Tom Wansbeek for helpful comments.

1 Introduction

The statutory retirement age in the Netherlands is proposed to increase in 2020 by one full year, to age 66, followed in 2025 by another one-year increase to age 67 (Stichting van de Arbeid, 2010). These and other pension reforms, such as the abolishment of the favorable fiscal treatment of early retirement contributions (CPB, 2005), were decided upon in response to both the aging of the Dutch population, which is increasing the tax burden for the working population, as well as government budget deficit problems caused by the recent credit crunch and the ensuing economic downturn. Furthermore, the same crisis has dampened pension fund resources, leading to proposals to increase occupational pension premia or decrease benefits. Of interest to policy makers are individuals' expectations of future pension entitlements, since they play an instrumental role in response to such pension reforms. For instance, the standard life-cycle model with uncertain (pension) income predicts that consumption during one's working life is positively related to what individuals *expect* to receive as pension income, and negatively related to income *uncertainty*. This paper examines the use of probabilistic survey questions to measure these expectations and uncertainty and to examine how these are affected by the inability of some individuals to answer probabilistic questions correctly.

In most empirical work (see e.g., Feldstein (1974); Attanasio and Rohwedder (2003); Kapteyn, Alessie, and Lusardi (2005)), individual-specific expectations and uncertainty are not available, leading authors to assume static, rational expectations. To avoid making such strict assumptions regarding the expectation formation process, several studies suggest using individuals' subjective expectations of future income (Dominitz and Manski, 2006; Guiso, Jappelli, and Padula, 2009). Of particular interest, and the ones used in this study, are probabilistic questions of the type suggested by Dominitz and Manski (1997b) and Manski (2004) which allow the researcher to elicit the subjective cumulative distribution function (CDF) of an individual's pension income. Manski (2004) provides an overview of the use of probabilistic questions, which have become increasingly popular in recent years and have been used to assess the likelihood of general events (such as inflation and social security benefits) as well as person-specific events (such as mortality and one's economic situation). The author suggests two reasons why eliciting expectations in a probabilistic way is better than eliciting expectations from vaguely defined answer categories (e.g., an event is "very likely" or "not too likely"). First, the numerical scale allows comparisons among individuals. Second, the consistency of a respondent's answers can be checked using the laws of probability. Our paper contributes to the empirical literature by investigating whether incorrect (or inconsistent) answers to probabilistic survey questions, and their subsequent removal from the sample, lead to endogenous sample selection. For example, less-educated and less financially literate persons may be less likely to answer such questions in a meaningful way—that is, their answers may not satisfy certain laws of probability—and may also have a different retirement income replacement rate.[1] Simply excluding these observations when analyzing the determinants of the subjective replacement rate or subjective uncertainty, as is commonly done in other

[1] The less-educated often have a higher income replacement rate, since the state pension in the Netherlands is a flat rate benefit, independent of past earnings, and occupational pension benefits increase with years of work experience.

papers (e.g., Dominitz and Manski (2006); De Bresser and van Soest (2010)), can therefore result in endogenous sample selection and bias the parameter estimates.

We use responses to the Dutch Pension Barometer survey, described in detail in Section 2, which follows the Dominitz and Manski (1997a,b, 2006) approach. By eliciting points on the (subjective) distribution function of future pension income, these questions allow the researcher to compute estimates of the expected replacement rate, that is, the ratio of expected pension income to current income, as well as its standard deviation, which can be interpreted as uncertainty regarding the replacement rate. De Bresser and van Soest (2010) use the same data to investigate what Dutch employees expect to receive as pension benefits and how uncertain they are about these benefits, as well as their relation to other individual characteristics. The authors report an average "correct" response rate of 60% and show that the response rate increases with education and income. To qualify as correct answers in both De Bresser and van Soest (2010) and this paper, the elicited probabilities should fulfill the requirements of a distribution function. In particular, the probabilities answered should be between 0% and 100% and should be monotonically non-decreasing over the support of the distribution function.

The quality of subjective expectations, elicited using probabilistic survey questions, is examined in other papers. Hurd and McGarry (1995, 2002) investigate the validity of subjective survival probabilities and find that individuals are well able to predict their own mortality, underlining Manski's (2004) conclusion that probabilistic survey questions are informative. Dominitz (1998) shows that next-year income expectations are able to predict subsequent realizations reasonably well. In addition, data on expectations and subsequent realizations are used in Dominitz (2001); Das and van Soest (1997); Das and Donkers (1999); Stephens (2004). The overviews of Hurd (2009) and Pesaran and Weale (2006) also emphasize the predictive power of subjective probabilities.

However, Dominitz and Manski (1996, 1997b) signal some evidence that not all respondents answer correctly.[2] Dominitz and Manski (1996) use survey software that automatically signals mistakes in the probabilities entered by the respondent, after which the respondent must correct the mistake, but still allows the researcher to keep track of them.[3] The authors find that 7% of respondents violate the monotonicity of answers, and 40% provide answers incompatible with the (previously elicited) median of the subjective distribution of future income. Dominitz and Manski (1997b) report 30% non-response, and 5% violating monotonicity. Kleinjans and van Soest (2010) show that two common fears associated with probabilistic questions, namely, non-response and focal points (e.g., answering 0%, 50%, or 100%), do not affect the determinants of retirement expectations, but that individuals round off probabilities instead. Manski and Molinari (2010) investigate the extent of rounding in more detail and find heterogeneity in answering patterns, with a small fraction (11% of the respondents) always rounding up to multiples of 50. These authors propose the use of statistical methods to analyze interval data, using an algorithm

[2] One notable difference with this paper is that the Dominitz and Manski questions first explicitly ask for the minimum and maximum household incomes the respondent expects and thereafter elicit expectations for amounts within this interval. Since the questions used in this paper concern pension benefits expressed as a percentage of the current wage, there is no need to ask for minimum and maximum benefits.

[3] The survey software used for the Pension Barometer we use does not signal mistakes to the respondent. See also Delavande and Rohwedder (2008), who study (non-)response patterns to probabilistic questions exploiting the flexibility of Internet surveys.

to determine the length of the interval for each respondent. We do not address the issues of rounding or focal points. More closely related to our study are Dominitz and Manski (2006), who compare the sample statistics of non-respondents to those in their final sample, using data from the Survey of Economic Expectations (SEE). Non-respondents[4] in the SEE are more likely to be female, less likely to be non-Hispanic whites, less likely to be labor force participants, less likely to be married, and less likely to be high school or college graduates. Furthermore, the probability of non-response is non-monotonic in age. The degree of non-response, Dominitz and Manski (2006) claim, compares favorably to that from the Health and Retirement Study, and hence the authors conclude that selection effects are not an issue. By contrast, our empirical findings provide strong evidence of endogenous sample selection effects when omitting incorrect answers from an analysis of expected retirement income replacement rates.

We find that about one-third of the respondents are unable to answer correctly.[5] Like De Bresser and van Soest (2010), we also find that the incidence of violations correlates with observable background variables, such as education, income, and gender. De Bresser and van Soest (2010) use only the sample with correct responses, thus implicitly assuming exogenous sample selection, and apply a least squares estimator. A new finding is that excluding those individuals for whom it is not possible to compute the expectation or standard deviation of future pension income results in endogenous sample selection. The resulting biases are quantified by predicting both the expected replacement rate and the standard deviation of the replacement rate using a linear model without correcting for selection effects and using a Heckman model that does correct for possible endogenous selection effects. This quantification shows that ignoring endogenous sample selection yields a downward bias in the predicted expected replacement rate and an upward bias in the predicted uncertainty (standard deviation) of the replacement rate. These biases are largest for less-educated individuals.

The paper is organized as follows. Section 2 discusses the data and the exact wording of the survey questions. Section 3 examines the incidence of violations of the laws of probability and how it relates to individual characteristics. Section 4 discusses the computation of the expected value and standard deviation of the replacement rate and relates these to individual characteristics. It also quantifies the consequences for the parameter estimates of ignoring endogenous sample selection. Finally, Section 5 presents our conclusions.

[4]Non-respondents' answers are either missing values, incomplete, non-valid, or unusable for estimating the subjective distribution.

[5]Although we use the same data as De Bresser and van Soest (2010), we find a different fraction of correct responses due to merging the Pension Barometer survey responses with the DNB Household Survey (DHS) and restricting the initial sample to non-missing observations on all variables used (see Section 2).

2 Data

Since the summer of 2006, CentERdata has been collecting data on the pension benefit expectations of Dutch households with the Pension Barometer survey. CentERdata, affiliated with Tilburg University, specializes in data collection via (Internet) surveys and administers the Pension Barometer survey to members of their CentERpanel, a representative sample of the Dutch population aged 16 and above. Those without access to the Internet are provided with a set-top box for their television. The CentERpanel households are interviewed regularly about various subjects. In particular, the DNB Household Survey is sent out to the same CentERpanel members. Hence, background information on household panel members is available from previous interviews, since unique identifiers for household members are available.

The Dutch pension system consists of three pillars.[6] The first pillar is the flat-rate public pension, provided to all inhabitants aged 65 and above. In 2010, this amounted to €1057 for singles and €735 for married individuals. The second pillar, the occupational pensions, are mandatory for most employees, and both employers and employees contribute to a pension fund. Finally, the third pillar concerns private pension products, such as annuities bought from banks or insurers. This paper concerns pension benefit expectations from the first and second pillars together.

There are two versions of the Pension Barometer: a monthly survey and a yearly survey. The monthly survey has a rotating panel structure, such that each member receives the survey once every three months. The main focus is on expectations regarding the public pension (see Bissonnette, Nelissen, and van Soest (2009); Bissonnette and van Soest (2010)).[7] The yearly survey is presented in two parts to the panel members, the first part in March and the second in June. The June version of the annual survey is our main source of information, for which the data for 2007, 2008, and 2009 are available. The probabilistic survey questions concerning the retirement income replacement rate are only asked to employees (see below). For this reason, the final sample we use for our analysis is restricted to employees.

We use the responses to the following sequence of questions to obtain a probability distribution of the replacement rate at two different ages for each individual.

Question 1 *At what age do you think you can retire at the earliest, following your employer's pension scheme?*

The answer to this question, say, age Y, is used in the subsequent question:

[6] See Bovenberg and Gradus (2008) for an overview of the Dutch pension system and its reforms.

[7] Van der Wiel (2008) combines expectations on first-pillar pension reforms from the monthly Pension Barometer survey with information on third-pillar pension products to find that participation in third-pillar pension products is greater for individuals who expect a dismantling of the public pension system, in terms of lower benefits or later retirement.

Question 2 *If you would retire at age Y, please think about your total net pension income, including social security, compared to your current total net wage or salary. What do you think is the probability that the purchasing power of your total net pension income in the year following your retirement will be*

a) more than 100% of your current net wage? ... %
b) less than 100% of your current net wage? ... %
c) less than 90% of your current net wage? ... %
d) less than 80% of your current net wage? ... %
e) less than 70% of your current net wage? ... %
f) less than 60% of your current net wage? ... %
g) less than 50% of your current net wage? ... %

Question 3 *Can your employer fire you for reaching a certain (pension eligibility) age?*

The answer can be either yes or no. If the respondent answers yes, a follow-up question is posed:

Question 4 *At what age is this?*

The answer to this question, say, age Z, is used in the next question. If, however, the respondent answered no to question 3, age Z is fixed to be either 65 if the answer to question 1, age Y, was lower than 55, or Y+5 years if Y was larger than 54. In any case, in question 5 respondents are shown a particular age that should be higher than the age for which they answered question 2. Hence, this age Z can be interpreted as their latest retirement age.

Question 5 *If you would retire at age Z, please think about your total net pension income, including social security, compared to your current total net wage or salary. What do you think is the probability that the purchasing power of your total net pension income in the year following your retirement will be*

a) more than 100% of your current net wage? ... %
b) less than 100% of your current net wage? ... %
c) less than 90% of your current net wage? ... %
d) less than 80% of your current net wage? ... %
e) less than 70% of your current net wage? ... %
f) less than 60% of your current net wage? ... %
g) less than 50% of your current net wage? ... %

These questions are intended to reveal the distribution of the net retirement income replacement rate of Dutch employees. For the remainder of this paper, we will call this the replacement rate,[8] or RR. Throughout this paper, the notation for the probability that, say, $RR > 100$ is $\mathbb{P}(RR > 100)$, as, for instance, asked in question 2a. Since the entire probability distribution of the replacement rate is elicited, we are able to compute the expected replacement rate, as well as the standard deviation (uncertainty) of the replacement rate for each individual (see section 4.2 for details). There is some evidence of item non-response for these questions, since some respondents answer questions 1 and/or 3 but not the probability questions. These are classified as "Don't know" in the remainder of this paper and concern 27 observations for Q2 and 35 observations for Q5. Table 1 shows sample statistics for each question in the three survey years. The number of observations varies between years, as well as between the two questions.

Table 1: Response patterns[a]

Year	**2007**				**2008**				**2009**			
Variable	Earliest[b]		Latest		Earliest		Latest		Earliest		Latest	
Number of employees	575				446				459			
Don't know	9		15		8		9		10		11	
Retirement age	64	(63.3)	65	(66.6)	65	(63.5)	65	(66.8)	65	(63.8)	65	(66.9)
$\mathbb{P}(RR > 100)$	0	(11.8)	0	(21.5)	0	(9.6)	0	(19.0)	0	(11.1)	0	(21.2)
$\mathbb{P}(RR < 100)$	92.5	(69.1)	90	(64.4)	98	(68.6)	90	(66.2)	90	(67.9)	85	(65.1)
$\mathbb{P}(RR < 90)$	75	(61.0)	70	(55.8)	80	(61.2)	70	(55.6)	75	(59.5)	60	(54.2)
$\mathbb{P}(RR < 80)$	60	(53.9)	50	(44.6)	60	(53.8)	50	(46.5)	50	(51.6)	50	(45.0)
$\mathbb{P}(RR < 70)$	40	(39.6)	20	(30.7)	40	(40.0)	25	(31.9)	30	(37.3)	25	(31.8)
$\mathbb{P}(RR < 60)$	10	(25.2)	10	(19.6)	12.5	(24.7)	10	(20.5)	10	(22.7)	10	(19.4)
$\mathbb{P}(RR < 50)$	2	(17.7)	1	(14.3)	5	(15.0)	1	(14.1)	5	(15.0)	1	(13.6)
Observations	566		560		438		437		449		448	

[a] This table shows the cross-sectional median (mean) of each variable. The initial sample is restricted to non-missing explanatory variables in Table 2.

[b] Earliest refers to earliest retirement age, in questions 1 and 2, and latest refers to latest retirement age, in questions 3, 4, and 5.

The DHS, formerly known as the VSB-CentER Savings Study, is a yearly survey that started in 1993 and covers about 2000 Dutch households, that also belong to the CentER-panel mentioned above. The survey is administered between February and December of the specific year. Respondents answer questions on a broad range of topics, including household income, assets and liabilities, health, and economic and psychological concepts (see Nyhus (1996) and Alessie, Hochguertel, and van Soest (2002) for an extended description). This study uses the three waves from 2007 to 2009 to obtain individual characteristics. In addition, and as will be explained in detail in Section 4, our analysis uses four variables that identify a Heckman selection model which takes into account possible endogenous sample selection. These four additional variables are the so-called exclusion restrictions. To construct them, we consider the following questions from the Pension Barometer survey and the DHS that provide information on whether or not people are able to answer probabilistic questions and which are not related to retirement.

[8] Traditionally, the replacement rate has been defined as the ratio between pension income and income just prior to retirement. This question asks instead the net replacement rate compared to the current net wage, which better suits the average pay system presently in place in the Netherlands.

Question 6 *How likely is it that you will attain (at least) the age of* 65 / 75 / 80 / 85 / 90 / 95 / 100 *?*[9] *Please indicate your answer on a scale of 0 through 10, where 0 means "no chance at all" and 10 means "absolutely certain."*

The survival probability should be decreasing in age; surviving until age 75 implies survival up to age 65. We construct the variable *Survival probability error*, which is equal to one if the respondent violates monotonicity, and zero otherwise.

Question 7 *What is the probability that the purchasing power of your total household income, in one year from now, will be* higher / lower *than it is now?*

Respondents provide two probabilities, one for the higher expected income and one for the lower. We construct the variable *Expected income adding-up error*, which is equal to one if these probabilities sum up to more than 100%, and zero otherwise.

The next question first elicits the minimum and maximum expected incomes with the following question:

Question 8 *We would like to know a little bit more about what you expect will happen to the net income of your household in the next 12 months. What do you expect to be the* LOWEST/HIGHEST *total net yearly income your household may realize in the next 12 months?*

Then, a series of four follow-up questions are posed based on those answers:

Question 9 *What do you think is the probability that the total net yearly income of your household will be less than* $€[LOWEST + (HIGHEST - LOWEST) * \{0.2/0.4/0.6/0.8\}]$ *in the next 12 months?*

In words, respondents are asked the probability that their net household yearly income will be less than 20% above their lowest expected income, and similarly for 40%, 60%, and 80% above. These four probabilities should be increasing with the threshold level (less than 20% above the lowest expected income implies less than 40% above the lowest expected income), and we construct the variable *Expected income probability error*, which is equal to one if this is violated, and zero otherwise. The final question has a setup similar to that of questions 8 and 9 but concerns expected inflation.

Question 10 *We now would like to learn what you expect will happen to the prices in the next twelve months. What do you think will be the* MINIMUM/MAXIMUM *percentage with which prices could increase over the next 12 months? If you think prices will decrease, you can fill in a negative percentage by using a minus sign in front of the number.*

In the surveys of 2008 and 2009, but not in 2007, this is followed up by the following question:

[9] Respondents answer at most three but mostly two questions, depending on their actual age. Respondents younger than 55 provide survival probabilities up to ages 65 and 75, while people aged $80 - 85$ provide survival probabilities for ages 95 and 100, with a gradual transition for the ages in between. This ensures that respondents do not have to provide survival probabilities for ages lower than their actual age or ages in the near future (five years).

Question 11 *What do you think is the most likely (consumer) price increase over the next 12 months?*

Respondents choose an integer from 1 to 10. This answer is used to define four thresholds (say, G_1 to G_4) for the series of follow-up questions:

Question 12 *Of course it is difficult to predict how much (consumer) prices will increase. The increase can be lower or higher than expected. Therefore we would like to ask you how certain you are about your prediction. How likely do you think it is that the increase in prices in the next 12 months will be* less than G_1 / less than G_2 / more than G_3 / more than G_4*?*

For 2007, the thresholds are again 20%, 40%, 60%, and 80% above the minimum expected inflation from question 10, as in question 9, and probabilities of inflation less than these thresholds are elicited. We construct the variable *Inflation probability error*, which is equal to one if monotonicity is violated, taking into account the different elicitation methods between survey years, and zero otherwise.

Table 2 shows the mean and standard deviation of individual characteristics used in this study: Female (=1 if female), age, education dummies, partner (=1 if married or in a relationship), gross monthly income in euros, years worked full-time and/or part-time (Experience), and the four exclusion restrictions. For education, we divide respondents into Elementary, Secondary, College, and University educational attainment.

Table 2: Descriptive statistics explanatory variables[a]

Variable	2007		2008		2009	
Female	0.396	(0.489)	0.374	(0.485)	0.364	(0.482)
Single	0.26	(0.439)	0.244	(0.43)	0.233	(0.423)
Income	2968.821	(7613.048)	3183.258	(8635.731)	2867.634	(1323.109)
Experience	22.512	(12.967)	23.325	(12.17)	24.473	(12.209)
Age	44.443	(10.024)	45.46	(10.05)	46.747	(9.997)
Elementary	0.241	(0.428)	0.238	(0.426)	0.235	(0.425)
Secondary	0.337	(0.473)	0.336	(0.473)	0.307	(0.462)
College	0.278	(0.448)	0.289	(0.454)	0.305	(0.461)
University	0.144	(0.351)	0.137	(0.344)	0.153	(0.36)
Survival probability error	0.007	(0.083)	0.002	(0.047)	0.007	(0.081)
Expected income adding-up error	0.059	(0.236)	0.054	(0.226)	0.037	(0.189)
Expected income probability error	0.172	(0.378)	0.157	(0.364)	0.198	(0.399)
Inflation probability error	0.144	(0.351)	0.296	(0.457)	0.338	(0.473)
Observations	575		446		459	

[a] Table shows the mean (standard deviation) of each explanatory variable for the sample of employees for which all dependent and explanatory variables are known.

The sample consists of mostly males and married or cohabiting persons. Gross monthly income is around €3000. More than 42% of respondents have higher vocational or university education. As for the potential exclusion restrictions, the incidence of errors in answering probabilistic questions is the highest for questions 9 and 12. There is some evidence of an increase in the incidence of errors after the change in elicitation method for question 12, but since we are merely interested in who is able to answer the probabilistic questions, we do not explore this issue further.

3 Violations

Questions 2 and 5 ask respondents to fill in seven probabilities. This section checks the answers for consistency. Question 2a asks $\mathbb{P}(RR > 100)$; question 2b asks $\mathbb{P}(RR < 100)$. Since these events are mutually exclusive, the sum of the answers to these questions should not exceed 100% so as not to violate adding up.[10] In order not to violate monotonicity, the answers to questions 2b to 2g should be non-increasing; if the replacement rate is less than 90% of the current income, it is, of course, also less than 100% of the current income. Table 1 shows that, for the average respondent, the answers to 2a and 2b add up to less than 100% and that, on average, there are no violations of monotonicity. Table 3 shows the number of individuals (by year) for whom the sum of their answers to questions 2a and 2b add up to more than 100% and/or who violate monotonicity in their answers to questions 2b to 2g. The percentage of correct responses was about 65% in 2007 and about 70% in 2008 and 2009, which suggests a slight decrease in the incidence of violations. Furthermore, there exists a clear pattern in the violations, in that for each year and each type of violation (i.e., six tests), we find a significant positive association[11] between correctly answering questions 2 and 5.

Table 3: Violations of the laws of probability[a]

	Panel A: 2007					
Replacement rate	Earliest retirement age (Q2)			Latest retirement age (Q5)		
	Monotonicity			**Monotonicity**		
Adding up	Correct	Incorrect	Total	Correct	Incorrect	Total
Correct	368	143	511	379	103	482
Incorrect	45	19	64	71	22	93
Total	413	162	575	450	125	575
	Panel B: 2008					
Replacement rate	Earliest retirement age (Q2)			Latest retirement age (Q5)		
	Monotonicity			**Monotonicity**		
Adding up	Correct	Incorrect	Total	Correct	Incorrect	Total
Correct	310	94	404	314	75	389
Incorrect	29	13	42	45	12	57
Total	339	107	446	359	87	446
	Panel C: 2009					
Replacement rate	Earliest retirement age (Q2)			Latest retirement age (Q5)		
	Monotonicity			**Monotonicity**		
Adding up	Correct	Incorrect	Total	Correct	Incorrect	Total
Correct	320	107	427	322	74	396
Incorrect	19	13	32	47	16	63
Total	339	120	459	369	90	459

[a] Table shows the number of respondents that violate adding up ($\mathbb{P}(RR > 100) + \mathbb{P}(RR < 100) > 1$), violate monotonicity ($\mathbb{P}(RR < 100) > \mathbb{P}(RR < 90)$or similar for other thresholds), violate both or answer correctly, in each year and for both questions 2 and 5

[10] Less than 100% is allowed, since this indicates a positive probability for the event of pension benefits equaling current income.

[11] The p-values corresponding to the null hypothesis of no association are all equal to 0.000.

The number of Pension Barometer respondents violating monotonicity and/or adding up is considerably larger than for the SEE respondents, reported by Dominitz and Manski (1997b): 5% of respondents are unable to report correct probabilities concerning next-year income. For pension benefit expectations, studied in Dominitz and Manski (2006), the percentage of correct responses is similar to our finding, 66%.

3.1 Explaining the ability to answer correctly

This section examines the extent to which the ability to answer probabilistic questions correctly—that is, without violations of adding up or monotonicity—is related to individual characteristics. We use gender, age, education, (household) income, and marital status as the explanatory individual characteristics. In addition, since the incidence of violations is lower in 2008 and 2009 compared to 2007 (see Table 3), we include a learning variable that equals one if the respondent is answering these questions for the first time, two if for the second time, and three if for the third time. For both questions separately, we define the variable *Able* to indicate the ability to answer correctly, where $Able_i = 0$ if respondent i violates adding up and/or monotonicity or is classified as "Don't know".

Table 4 reports the effects of individual characteristics on the ability to answer correctly (*Able*). We estimate (pooled) probit models with time-specific intercepts. Table 4 reports parameter estimates, standard errors clustered at the individual level, and marginal effects.

First, we observe that the parameter estimates and significance levels are stable across the questions. We observe that older people are less likely to produce inconsistent answers, but the effect is not significant. Education is very important in explaining ability: Persons with a secondary, college, or university background are significantly less likely to make mistakes than the benchmark group with the lowest level of education. Jointly, the education dummies are highly significant (p-value = 0.000). A perhaps counterintuitive finding is that higher-income individuals are more likely to make mistakes, but the effect is not significant; the education dummies make the inclusion of income redundant, since more educated persons have higher incomes, in general.[12] Women are more likely to violate, while marital status is insignificant. We will use the incidences of incorrectly answering the probabilistic questions corresponding to the survival probability, income expectation, and inflation expectation as exclusion restrictions in the Heckman selection model in Section 4 to explain the expected replacement rate and the standard deviation of the replacement rate. The bottom of Table 4 shows that these excluding variables are jointly significant. Those individuals who make errors in these other probabilistic questions are more prone to make mistakes in the probabilistic questions concerning the replacement rate. Time effects show that significantly fewer people made mistakes in 2008 and 2009, yet this cannot be attributed to learning effects, which have an insignificant impact. We have no explanation for this finding. A Chow test for structural breaks shows that the coefficients, other than the intercept, are stable over time; the null hypothesis of time-invariant slope parameters cannot be rejected (see the bottom of Table 4).

[12] We estimated the equations without education dummies and found that income has a significantly positive

Table 4: Explaining violations[a]

Dependent variable: Able[b]	At the earliest retirement age (Q2)			At the latest retirement age (Q5)		
Covariates	Parameter estimate	Standard error	Marginal effect	Parameter estimate	Standard error	Marginal effect
Age	0.051	0.034	-0.0002	0.026	0.036	0.001
Age^2	-0.0006	0.0004		-0.0003	0.0004	
Secondary	0.140	0.102	0.049	0.233**	0.105	0.078
College	0.431***	0.106	0.143	0.470***	0.108	0.151
University	0.684***	0.141	0.204	0.711***	0.142	0.204
log income	-0.023	0.079	-0.008	-0.137	0.089	-0.047
Female	-0.120	0.089	-0.042	-0.247***	0.091	-0.086
Single	0.043	0.093	0.015	0.044	0.095	0.015
Survival probability error	0.217	0.494	0.072	0.577	0.582	0.161
Expected income adding-up error	-0.175	0.172	-0.064	-0.395**	0.157	-0.146
Expected income probability error	-0.327***	0.094	-0.120	-0.268***	0.093	-0.096
Inflation probability error	-0.234***	0.079	-0.084	-0.126	0.081	-0.044
Learning	0.052	0.049	0.018	0.040	0.049	0.014
Dummy 2008	0.185**	0.082	0.064	0.127	0.082	0.043
Dummy 2009	0.208**	0.084	0.071	0.123	0.086	0.041
Constant	-0.751	0.962		0.707	1.020	
Observations		1453			1445	
Pseudo-R^2		0.046			0.039	
log L		-862.0			-845.5	
p-value Wald test model		0.000			0.000	
p-value Wald test Age		0.328			0.646	
p-value Wald test Education		0.000			0.000	
p-value Wald test Exclusion restrictions		0.000			0.001	
p-value Wald test Time dummies		0.020			0.220	
p-value Chow test		0.510			0.950	

[a] Table shows parameter estimates, standard errors clustered at the respondent level, and marginal effects after probit. Here ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

[b] The dependent variable *Able* equals one if the respondent answers correctly, and zero if the probabilities violate adding up and/or monotonicity or the answer is classified as "Don't know".

4 Replacement rates

This section first computes the expected value and variance of the replacement rate. Second, it analyzes the determinants of the expected replacement rate and uncertainty (the standard deviation of the replacement rate) and relates these to individual characteristics using Heckman selection models that take into account possibly endogenous sample selection (Heckman, 1979). Finally, we quantify the selection bias by predicting the expected value and standard deviation of the replacement rate with and without sample selection corrections for different types of individuals.

4.1 Computation

The answers to question 2 provide information on the subjective cumulative distribution function for each respondent. There are seven thresholds for which we know the probability that the replacement rate is lower or higher than this threshold. We assume a nonparametric, piecewise-linear subjective distribution function for each respondent; that is, the distribution function is uniform between two thresholds.[13] The CDF, denoted $F(RR)$, can then be written as

$$F(RR) = \begin{cases} \mathbb{P}(RR < 50)\left(\frac{RR}{50}\right) & \text{if} \quad 0 \leq RR < 50 \\ \mathbb{P}(RR < 50) + \mathbb{P}(50 \leq RR < 60)\left(\frac{RR-50}{10}\right) & \text{if} \quad 50 \leq RR < 60 \\ \mathbb{P}(RR < 60) + \mathbb{P}(60 \leq RR < 70)\left(\frac{RR-60}{10}\right) & \text{if} \quad 60 \leq RR < 70 \\ \mathbb{P}(RR < 70) + \mathbb{P}(70 \leq RR < 80)\left(\frac{RR-70}{10}\right) & \text{if} \quad 70 \leq RR < 80 \\ \mathbb{P}(RR < 80) + \mathbb{P}(80 \leq RR < 90)\left(\frac{RR-80}{10}\right) & \text{if} \quad 80 \leq RR < 90 \\ \mathbb{P}(RR < 90) + \mathbb{P}(90 \leq RR < 100)\left(\frac{RR-90}{10}\right) & \text{if} \quad 90 \leq RR < 100 \\ \mathbb{P}(RR < 100) + \mathbb{P}(RR = 100) & \text{if} \quad RR = 100 \\ \mathbb{P}(RR \leq 100) + \mathbb{P}(100 < RR < 120)\left(\frac{RR-100}{20}\right) & \text{if} \quad 100 < RR < 120 \end{cases} \quad (1)$$

All the probabilities are known from the answers given by respondents. Writing the CDF as above allows us to work with a continuous distribution function. There is also one discrete element, since the answers to questions 2a and 2b can add up to less than 100%, which indicates that there is a positive probability associated with the event that the retirement income is exactly equal to the current income. We fix the upper bound of the distribution function (arbitrarily) at 120%, and the lower bound at 0%. These choices influence the value of the computed replacement rate but do not induce additional cross-sectional variation and, therefore, do not influence the results of the empirical analysis.

---

effect on the ability to answer correctly. These results are available upon request.

[13]Dominitz and Manski (1997b), Manski (2004), and De Bresser and van Soest (2010) fit a log-normal distribution to the probabilities. This parametric approach relaxes the need to consider small violations in adding up or monotonicity. However, this approach has disadvantages as well. First, it is not clear why the replacement rate should be log-normally distributed. Second, the probabilities need to vary with the threshold to make a non-linear least squares estimation feasible. Third, if the violation becomes "too large", the non-linear least squares estimates hardly converge, leading to implausible estimates of the mean or standard deviation of the replacement rates. Yet, what exactly is too large is not at all clear. Therefore, we propose a nonparametric distribution of the replacement rate.

To illustrate, consider the following two hypothetical respondents, depicted in Table 5.

Table 5: Answers of hypothetical respondents

	$\mathbb{P}(RR < 50)$	$\mathbb{P}(RR < 60)$	$\mathbb{P}(RR < 70)$	$\mathbb{P}(RR < 80)$	$\mathbb{P}(RR < 90)$	$\mathbb{P}(RR < 100)$	$\mathbb{P}(RR > 100)$
Respondent 1	0	10	25	50	80	90	10
Respondent 2	0	0	0	0	100	100	0

For each respondent, we derive an observation-specific lower and upper bound of the replacement rate. The upper bound is the lowest threshold (starting from the highest) with zero probability, and the lower bound is the highest threshold to which the respondent attaches zero probability. For the first hypothetical respondent, these are $\min(RR) = 50$ and $\max(RR) = 120$. The entire support of the distribution is found by computing $\text{Range}_i = \max_i(RR) - \min_i(RR)$. A special case is the second respondent, who attaches 100% probability to some threshold and 0% to the next lower threshold (see Table 5). This respondent reports that the replacement rate will be between 80% and 90% of the net current income, which is the type of answer that corresponds to the least uncertainty. In this case, we assume that the minimum and maximum are the same and equal to 90%, so that there is no uncertainty about the level of the replacement rate. In the latter case, the range equals zero.

To compute the expected replacement rate, we use the CDF to find the probability density function $f(RR)$ by differentiation. The expected value equals

$$E(RR) = \int_0^{\infty} RRf(RR)\, dRR \tag{2}$$

A more detailed computation can be found in the Appendix.

The measure for pension income uncertainty is the standard deviation of the replacement rate. Again, we use the CDF to find $E(RR^2) = \int_0^{\infty} RR^2 f(RR)\, dRR$ and, next, compute

$$\text{SD}(RR) = \sqrt{E\left(RR^2\right) - \left(E(RR)\right)^2} \tag{3}$$

The standard deviation is set to zero if the range equals zero. For our first hypothetical respondent, we find that $E(RR) = 80$ and $\text{SD}(RR) = 15.275$; for the second, $E(RR) = 90$ and $\text{SD}(RR) = 0$.

Table 6 shows summary statistics for the expected value and standard deviation of the replacement rate. Since these cannot be computed for respondents violating the laws of probability, the samples are smaller than those used for Table 4. The differences between the years are small, although we observe a slightly lower replacement rate in 2008. We see a large variation in the expected replacement rate, between 25% (which is the minimum due to the lower bound assumption) and 110% (the maximum) of the current income, with an average of 75%. Furthermore, uncertainty about pension entitlements varies greatly in the sample. On average, respondents have a standard deviation of about

15%, but this varies between 0% and 43.9% of the current income. The expected replacement rate is higher if the respondent faces question 5 with a higher retirement age. A back-of-the-envelope calculation shows that, on average, employees expect 1.67% (5% for a retirement delayed, on average, by three years; see Table 1) more pension income for each additional year of employment.[14] In addition, uncertainty is lower at this later retirement age. Figure 1 shows the empirical distribution of the expected value and standard deviation of the replacement rates, both as a histogram and as a kernel-smoothed estimate of the density (with a bandwidth of 9). The expected replacement rates (Figures 1a and 1c) are symmetrically distributed, with outliers in both tails. The median is around 70% of the current income. Figures 1b and 1d show a spike at a standard deviation of zero. These individuals have expressed their certainty about the level of the replacement rate. Compared to the other respondents, they have, on average, a higher expected replacement rate, are six years older, are slightly less educated, and have similar monthly incomes. Women and individuals with a partner express certainty more often.[15]

Table 6: Replacement rate summary statistics[a]

	At the earliest retirement age (Q2)					At the latest retirement age (Q5)				
Variable	Mean	Std. Dev.	Min.	Max.	N	Mean	Std. Dev.	Min.	Max.	N
Panel A: 2007										
$E(RR)$	75.104	16.848	25	110	368	80.636	17.378	25	110	379
$SD(RR)$	14.599	9.021	0	43.899	368	14.579	9.319	0	43.899	379
Panel B: 2008										
$E(RR)$	73.273	17.708	25	110	310	78.851	18.74	25	110	314
$SD(RR)$	15.639	9.757	0	43.899	310	14.866	10.001	0	43.899	314
Panel C: 2009										
$E(RR)$	75	17.043	25	110	320	80.131	17.658	25	110	322
$SD(RR)$	15.616	9.884	0	43.899	320	14.922	10.193	0	43.899	322

[a] Table shows the (cross-sectional) sample statistics for the expected replacement rate ($E(RR)$), computed according to equation 2, and the standard deviation of the replacement rate ($SD(RR)$), using equation 3, in each year for both questions 2 and 5

The correlation coefficients between the two questions for $E(RR)$ and $SD(RR)$ are equal to, respectively, 0.78 and 0.90 (877 observations).[16]

4.2 Explaining expected value and standard deviation

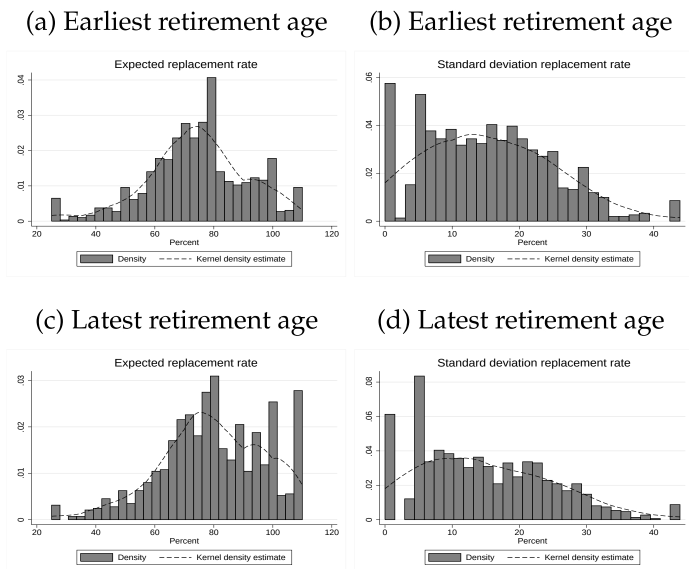
Next, we examine the determinants of the expected value and standard deviation of the replacement rate for both questions by estimating Heckman selection models (Heckman,

[14] For 106 observations, we find a lower expected replacement rate for the latest retirement age, with a 5.7% lower replacement rate, on average, which is counterintuitive. The vast majority do expect a positive effect of later retirement on the replacement rate.

[15] These results are not shown but are available upon request.

[16] Interestingly, for 127 observations, we find the same answer to questions 1 and 4, but different responses to questions 2 and 5. The correlations between the two questions are high and equal 0.85 and 0.81, respectively. We do not treat these observations differently.

Figure 1: Expected value and standard deviation replacement rates



1979). The Heckman selection model consists of two equations. The first equation is a selection equation to determine who is able to answer probabilistic questions correctly (see Section 3.1). The second equation is an outcome equation in which the expected replacement rate or the standard deviation of the replacement rate is explained. The error terms of these two equations are assumed to be bivariate normally distributed with correlation coefficient ρ (see, e.g., Cameron and Trivedi (2005) for more details). Both equations are estimated simultaneously by the maximum likelihood (ML) method.[17] We use time effects, age, gender, marital status, education, income, and work experience as explanatory variables.[18] The estimation results of the selection equation are already reported in Table 4. As discussed earlier, the exclusion restrictions are on the variables *Survival probability error*, *Expected income adding-up error*, *Expected income probability error*, and *Inflation probability error*. We assume, after controlling for incorrectly answering the replacement rate probabilistic questions, that these variables only affect the outcome variables through the selection adjustment and have no independent impact on the outcome variables. We believe this is a reasonable assumption, since these variables are solely related to the ability to answer probabilistic questions, and not to retirement income expectations. Among others studies, Delavande, Perry, and Willis (2006) and O'Donnell, Teppa, and van Doorslaer (2008) show that subjective survival probabilities can explain retirement behavior. Still,

[17] For the Heckman (1979) two-step estimator, the results of the selection equation are exactly equal to those in Table 4. For maximum likelihood, the results are quantitatively and qualitatively similar and are therefore omitted here. We opt for ML to be able to cluster standard errors at the individual level.

[18] We experimented with other specifications, in particular by including the expected retirement age from questions 1 and 4 and several interaction effects, but these estimated effects turned out to be insignificant.

we are confident that violations of the laws of probability in answering these questions are exogenous to retirement expectations after conditioning on incorrect answering the retirement expectation questions. As discussed in Section 3.1 and again reported in Table 7 for the sample selection model, the excluding variables are jointly significant in explaining the ability to answer probabilistic survey questions correctly. The estimation results of the outcome equations for the expected value and standard deviation of the replacement rate are displayed in Table 7 (for both questions). Standard errors are clustered at the individual level. For comparison, Table 7 also shows the results of estimating a linear model for the outcome equations by ordinary least squares (OLS), that is, without the endogenous sample selection correction.

The selection and outcome equations are not independent: The correlation between the error terms, ρ, is significant for all four models. The negative sign of $\hat{\rho}$ in the expected replacement rate equations implies that the unobservable factors in the selection (ability) equation, such as financial literacy, will have a positive effect on the probability of answering correctly while at the same time reducing the expected value of the replacement rate. Furthermore, the unobservable factors in the selection equation are positively correlated with the uncertainty in the replacement rate (a positive $\hat{\rho}$ in the $\mathrm{SD}(RR)$ models). These significant correlation coefficients imply that biased parameter estimates are obtained from a model explaining the expected value or standard deviation of the replacement rate that ignores endogenous sample selection resulting from incorrectly answering the probabilistic questions. Comparison with the OLS estimates shows that ignoring these sample selection effects biases the parameter estimates upward. This is important if these parameters are, for example, used to predict expected replacement rates and pension risk for explaining household consumption or savings behavior (Bottazzi, Jappelli, and Padula, 2006; Van der Wiel, 2008; Guiso, Jappelli, and Padula, 2009).

The assumption of the bivariate normality of the errors of the selection and outcome equation is tested using the approach suggested by Lee (1984) and Pagan and Vella (1989). The normality of the errors of the (probit) selection equation is tested by means of a Lagrange Multiplier test. We regress the constant on the fitted values multiplied by the generalized residual, as well as the squared and cubed fitted values multiplied by the generalized residuals, and compute $n * R^2$; the bottom of Table 7 shows that the null of normality is not rejected (p=0.655 and p=0.273 for the two questions, respectively). The key to testing bivariate normality is that, under the null hypothesis of bivariate normality, the error term of the outcome equation is a linear function of the error term of the selection equation. Lee (1984) and Pagan and Vella (1989) derive testable implications from this observation, and Pagan and Vella (1989) suggest adding fitted values of the selection equation multiplied by the inverse Mills ratio[19] and the second and third powers of the fitted values, again multiplied by the inverse Mills ratio, to the outcome equation. Bivariate normality implies that the coefficients of these added variables equals zero. The p-value of a Wald test for this implication is presented at the bottom of Table 7. The results show that the null hypothesis of normality is not rejected at the 5% significance level, with the most convincing evidence for normality found for the error terms of $E(RR)$ Q2.

[19] If $\phi(z)$ represents the standard normal density and $\Phi(z)$ the cumulative standard normal density, the inverse Mills ratio equals $\lambda(z) = \phi(z)/\Phi(z)$.

Table 7: Estimation results of Heckman and linear models[a]

Covariates	E(RR) Q2		E(RR) Q5		SD(RR) Q2		SD(RR) Q5	
Model	Heckman	Linear	Heckman	Linear	Heckman	Linear	Heckman	Linear
Age	-2.104***	-1.550**	-1.437*	-1.104*	1.031***	0.756**	0.515	0.330
	(0.702)	(0.623)	(0.748)	(0.670)	(0.385)	(0.327)	(0.411)	(0.359)
Age^2	0.022***	0.015**	0.016*	0.013*	-0.014***	-0.012***	-0.008*	-0.007*
	(0.008)	(0.007)	(0.008)	(0.007)	(0.004)	(0.004)	(0.005)	(0.004)
Secondary	-7.178***	-5.574***	-7.702***	-5.518***	3.822***	2.660**	4.217***	2.549**
	(2.211)	(1.942)	(2.193)	(1.867)	(1.247)	(1.124)	(1.208)	(1.065)
College	-11.960***	-7.766***	-12.830***	-8.432***	4.446***	1.671	5.282***	2.165**
	(2.272)	(1.904)	(2.342)	(1.937)	(1.292)	(1.140)	(1.197)	(1.048)
University	-12.740***	-6.771***	-12.440***	-6.387**	6.784***	2.902**	6.601***	2.242*
	(2.913)	(2.590)	(2.955)	(2.604)	(1.573)	(1.429)	(1.518)	(1.335)
log Income	-2.334	-2.377	0.171	-0.727	0.109	0.083	-0.723	-0.260
	(1.659)	(1.481)	(1.839)	(1.619)	(1.013)	(0.984)	(0.910)	(0.763)
Female	-3.037*	-3.976***	-1.125	-3.428**	-0.781	-0.279	-1.205	0.262
	(1.669)	(1.514)	(1.814)	(1.618)	(0.931)	(0.837)	(0.977)	(0.824)
Single	-1.027	-0.476	-2.105	-1.492	0.091	-0.326	0.321	-0.051
	(1.732)	(1.582)	(1.725)	(1.533)	(0.891)	(0.797)	(0.933)	(0.808)
Experience	-0.008	0.013	0.017	0.014	-0.023	-0.031	-0.036	-0.036
	(0.080)	(0.076)	(0.080)	(0.074)	(0.042)	(0.037)	(0.043)	(0.039)
Dummy 2008	-2.964**	-1.761	-2.785**	-1.894	2.004***	1.380**	1.034	0.557
	(1.281)	(1.121)	(1.382)	(1.166)	(0.742)	(0.613)	(0.785)	(0.626)
Dummy 2009	-1.075	0.120	-1.663	-0.914	2.271***	1.657***	1.413*	1.086*
	(1.338)	(1.160)	(1.383)	(1.162)	(0.759)	(0.618)	(0.807)	(0.641)
Constant	161.280***	137.819***	128.640***	115.536***	-10.679	3.283	5.168	14.841
	(20.237)	(17.957)	(21.282)	(19.460)	(12.072)	(10.669)	(11.557)	(10.092)
Observations	1453	998	1445	1015	1453	998	1445	1015
log L	-5078	-4224	-5174	-4340	-4441	-3589	-4504	-3686
R^2		0.058		0.052		0.143		0.130
p-value Equation	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
p-value Age	0.004	0.010	0.149	0.188	0.000	0.000	0.000	0.000
p-value Education	0.000	0.001	0.000	0.000	0.000	0.075	0.000	0.107
p-value Time dummies	0.063	0.182	0.122	0.266	0.005	0.017	0.181	0.238
ρ	-0.823		-0.875		0.902		0.964	
p-value $\rho = 0$	0.000		0.000		0.000		0.000	
p-value Exclusion restr.	0.000		0.022		0.003		0.002	
p-val. Normality selection [b]	0.655		0.273		0.655		0.273	
p-val. Normality outcome [c]	0.279		0.061		0.056		0.058	

[a] Heckman models are fitted in one step by the maximum likelihood method. The results of the selection equation are comparable to those of Table 4. Clustered standard errors in parentheses. Here ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

[b] The normality assumption of the probit error terms is tested by means of the Lagrange multiplier test for detecting skewness and excess kurtosis (see the text for details).

[c] The normality of the outcome equation error term is tested by means of the procedure suggested in Pagan and Vella (1989) (see the text for details).

Table 7 shows that, compared to less-educated individuals, higher-educated individuals expect a significantly lower replacement rate and are more uncertain about it. Two (possibly reinforcing) explanations for this finding are as follows. First, since higher-educated individuals, on average, earn higher wages, while the Dutch state pension system is redistributive in nature, they should expect a lower replacement rate than less-educated individuals. Furthermore, supported by the surprising fact that past work experience is not significant, the career path of higher-educated individuals is usually steeper and surrounded by greater uncertainty, making it harder to predict pension benefits, resulting in more subjective uncertainty. Second, these findings may reflect the fact that higher-educated individuals are better informed about their future pension entitlements by, for instance, keeping closer track of news and developments regarding their pension income. The recent turmoil on the financial markets following the credit crunch of 2008 affected both state and occupational pensions in the Netherlands, lowering pension benefits and increasing the eligibility age. Furthermore, recently implemented changes, including the change from final-pay to average-pay occupational pensions and the abolishment of tax-favorable early retirement contributions, are mostly negative for the level of pension benefits. Hence, those who keep track of (possible) changes in the pension system (i.e., the higher educated) may reasonably expect lower benefits and a more uncertain future.

The expected replacement rate decreases with age until age 48 (45) for the early (late) retirement age, and uncertainty increases with age until age 36 (31). Indeed, the current proposed changes in retirement income (increase in eligibility age, increasing premia) increases uncertainty for the young and decreases their income. Furthermore, uncertainty increases over time, with pension income expected to decrease. Income is not significant, due to both the inclusion of education dummies and the strong correlation between income and education, as well as the fact that the respondents are asked to condition on current income in answering the probabilistic questions.[20] Marital status is not significant, which is not surprising, since the differences between singles and couples are small for state pension benefits and nonexistent for occupational pensions. The uncertainties in pension benefits elaborated upon above (reforms, financial crisis) hold for both couples and singles, and hence there is no reason to expect significant differences between the two.

To gain more insight in the magnitude of the selection bias, we predict both $E(RR)$ and $SD(RR)$ using the Heckman model and the linear model. We consider a benchmark individual who is a married 50-year-old male with 25 years of experience and a university degree. His gross income is €3000 per month, which is in between the median and mean incomes in our sample. Furthermore, we show the effects of changing one of these characteristics, as well as plug in sample averages for all the characteristics from Table 2. The survey year is 2008, and the results are shown in Table 8 below.[21]

[20] The results without education are available upon request. The second reason, however, cannot be proven from the data.

[21] The standard errors of the predictions are presented in parentheses. A formal test of significant differences between the predictions of the Heckman and linear models would boil down to testing the significance of the error correlation ρ, which is provided in Table 7.

Table 8: Heckman model versus linear model predictions[a]

Person / Model	E(RR) Q2 Heckman	E(RR) Q2 Linear	E(RR) Q5 Heckman	E(RR) Q5 Linear	SD(RR) Q2 Heckman	SD(RR) Q2 Linear	SD(RR) Q5 Heckman	SD(RR) Q5 Linear
Benchmark[b]	75.165	71.104	83.582	78.718	13.895	16.450	11.599	14.302
	(2.304)	(2.178)	(2.606)	(2.286)	(1.264)	(1.083)	(1.319)	(1.137)
Age 40	76.968	72.790	83.303	78.051	16.794	19.570	14.108	17.429
	(2.291)	(2.173)	(2.628)	(2.334)	(1.283)	(1.075)	(1.333)	(1.166)
Age 60	77.654	72.460	87.078	81.956	8.112	11.024	7.467	9.827
	(2.670)	(2.557)	(2.901)	(2.517)	(1.418)	(1.225)	(1.415)	(1.197)
Elementary education	87.903	77.875	96.017	85.105	7.111	13.547	4.998	12.060
	(2.263)	(1.772)	(2.492)	(1.750)	(1.415)	(1.081)	(1.186)	(1.003)
Secondary education	80.725	72.301	88.315	79.587	10.933	16.207	9.215	14.609
	(1.894)	(1.568)	(2.151)	(1.621)	(1.164)	(0.849)	(1.056)	(0.888)
College education	75.946	70.110	83.189	76.672	11.557	15.218	10.280	14.224
	(1.677)	(1.434)	(2.009)	(1.575)	(1.003)	(0.798)	(1.002)	(0.839)
Single	74.138	70.628	81.477	77.226	13.986	16.124	11.919	14.251
	(2.646)	(2.590)	(2.816)	(2.500)	(1.391)	(1.213)	(1.427)	(1.249)
Female	72.128	67.128	82.457	75.290	13.114	16.171	10.393	14.564
	(2.412)	(2.242)	(2.781)	(2.370)	(1.391)	(1.146)	(1.415)	(1.216)
Income €1800	76.332	72.293	83.496	79.082	13.840	16.409	11.960	14.432
	(2.634)	(2.452)	(2.984)	(2.628)	(1.506)	(1.346)	(1.522)	(1.304)
Sample average[c]	79.014	71.495	86.618	77.663	11.956	16.757	9.663	15.402
	(1.484)	(1.180)	(1.840)	(1.229)	(1.003)	(0.677)	(0.875)	(0.684)

[a] This table shows the predicted values of the expected replacement rate ($E(RR)$) and the standard deviation ($SD(RR)$) from the estimated Heckman and linear models (see Table 7). Standard errors are in parentheses.

[b] The benchmark individual is a 50-year-old married male with a university degree and 25 years of work experience who earns €3000 gross per month. The survey year is 2008. The other rows show the only deviation from this benchmark.

[c] The sample averages correspond to the 2008 column in Table 2.

We see that the bias is quite substantial for the benchmark individual: about 4 percentage points for the expected replacement rate and about 2.7 percentage points for the standard deviation. The differences in the predictions are similar whether we consider earliest retirement (Q2) or latest retirement (Q5). The bias increases dramatically if we lower the education level: more than 10 percentage points for $E(RR)$ and more than six percentage points for $SD(RR)$. For the sample averages, the bias is larger than the benchmark as well: more than 7.5 percentage points for $E(RR)$ and more than five points for $SD(R)$. In all cases, the linear model underestimates the predicted expected replacement rate and overestimates the degree of uncertainty. To summarize, ignoring endogenous sample selection due to incorrect responses to probabilistic survey questions concerning pension entitlements biases the results toward a more pessimistic expectation and excess uncertainty in the replacement rate.

5 Conclusion

This paper examines possible endogenous sample selection effects due to incorrectly answering probabilistic survey questions concerning pension entitlements in a sample of Dutch employees. The questions are designed to elicit the subjective distribution of the replacement rate of pension income at the individual level, which makes it possible to compute the expected value and standard deviation of the retirement income replacement rate. Selection effects arise, since some individuals are not able to answer or incorrectly answer probabilistic survey questions: They violate either adding up, the monotonicity of the distribution function, or both or refuse to give any answer. We find that about one-third of respondents give answers that violate the laws of probability. Furthermore, the incidence of incorrect answers correlates with education and gender: higher-educated individuals and men make fewer mistakes.

For the sample of respondents who answered the probabilistic questions correctly, we find that the average expected replacement rate equals 75% of the current income, with an increase of 1.67% per extra year in the labor force. Pension risk, measured as the standard deviation of the subjective replacement rate, averages around 15% of the current income. Both the expected value and standard deviation show a great deal of heterogeneity between respondents.

We use a Heckman sample selection model and find strong evidence of endogenous sample selection due to incorrect answering, both for the expected replacement rate and the standard deviation of the replacement rate. The variables excluded from the outcome equations are based on incorrectly answering probabilistic survey questions on survival, income, and inflation expectations, which are assumed not to be directly related to retirement decisions or pension income after controlling for incorrectly answering the replacement rate question. These excluding variables significantly predict the ability to correctly answer the two questions concerning the replacement rate.

To quantify the bias from using a linear model as opposed to a Heckman selection model, we predict the expected replacement rate and standard deviation using both models. The bias is large, especially for less-educated individuals, namely, 12 percentage points of the expected replacement rate and 6.5 percentage points of the standard deviation of the replacement rate. To summarize, ignoring endogenous sample selection due to incorrect responses to probabilistic survey questions biases the results toward a more pessimistic expectation and excess uncertainty in the replacement rate.

For further research, one can use the expected pension income and pension risk based on the (computed) replacement rates and, for instance, estimate a life cycle model of consumption without the need to arbitrarily assume how expectations are formed. An important implication of our findings for such research is that one must account for the endogenous selection effects due to incorrect responses. An interesting extension of our study is to examine how persistent expectations are. That is, we have shown some dispersion in the expected replacement rate in a (pooled) cross-section, but to obtain further insights in expectation formation, it would be of interest to analyze at the individual level how expectations are updated when new information becomes available. Hence, the persistency of these expectations can tell us something about how expectations are formed. Such an extension may require a longer panel than we currently have and is left for future research.

References

ALESSIE, R., S. HOCHGUERTEL, AND A. VAN SOEST (2002): "Household portfolios in the Netherlands," in *Household portfolios*, ed. by L. Guiso, M. Haliassos, and T. Jappelli, pp. 341–388. MIT press, Cambridge, MA.

ATTANASIO, O. P., AND S. ROHWEDDER (2003): "Pension Wealth and Household Saving: Evidence from Pension Reforms in the United Kingdom," *The American Economic Review*, 93(5), 1499–1521.

BISSONNETTE, L., J. NELISSEN, AND A. VAN SOEST (2009): "Retirement Expectations in the Netherlands," *Paper presented at Netspar Conference on Labour Force Participation and Well-being of the 50+ Population, Utrecht, 17-09-2009*.

BISSONNETTE, L., AND A. VAN SOEST (2010): "Retirement Expectations, Preferences, and Decisions," Panel Paper 18, Netspar.

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.

BOVENBERG, A., AND R. GRADUS (2008): "Dutch policies towards ageing," *European View*, 7, 265–275.

CAMERON, A., AND P. TRIVEDI (2005): *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

CPB (2005): "Centraal Economisch Plan," Chapter 6.3: Stelselwijziging vut, prepensioen en levensloop, CPB Netherlands Bureau of Economic Policy Analysis.

DAS, M., AND B. DONKERS (1999): "How Certain Are Dutch Households about Future Income? An Empirical Analysis," *Review of Income and Wealth*, 45(3), 325–338.

DAS, M., AND A. VAN SOEST (1997): "Expected and Realized Income Changes: Evidence from the Dutch Socio-Economic Panel," *Journal of Economic Behavior & Organization*, 32(1), 137 – 154.

DE BRESSER, J., AND A. VAN SOEST (2010): "Retirement expectations of Dutch employees," *Paper presented at the Netspar Theme Conference: Pensions, savings and retirement decisions, Utrecht, May 7, 2010*.

DELAVANDE, A., M. PERRY, AND R. WILLIS (2006): "Probabilistic Thinking and Early Social Security Claiming," Working Papers 129, University of Michigan, Michigan Retirement Research Center.

DELAVANDE, A., AND S. ROHWEDDER (2008): "Eliciting Subjective Probabilities in Internet Surveys," *Public Opinion Quarterly*, 72(5), 866 – 891.

DOMINITZ, J. (1998): "Earnings Expectations, Revisions, and Realizations," *The Review of Economics and Statistics*, 80(3), 374–388.

——— (2001): "Estimation of Income Expectations Models using Expectations and Realization Data," *Journal of Econometrics*, 102(2), 165 – 195.

DOMINITZ, J., AND C. F. MANSKI (1996): "Eliciting Student Expectations of the Returns to Schooling," *The Journal of Human Resources*, 31(1), 1–26.

——— (1997a): "Perceptions of Economic Insecurity: Evidence From the Survey of Economic Expectations," *The Public Opinion Quarterly*, 61(2), 261–287.

——— (1997b): "Using Expectations Data To Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92(439), 855–867.

——— (2006): "Measuring Pension-benefit Expectations Probabilistically," *Labour*, 20(2), 201–236.

FELDSTEIN, M. S. (1974): "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *The Journal of Political Economy*, 82(5), 905–926.

GUISO, L., T. JAPPELLI, AND M. PADULA (2009): "Pension Risk, Retirement Saving and Insurance," *CSEF Working Papers*, 30(223).

HECKMAN, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

HURD, M. D. (2009): "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, 1(1), 543–562.

HURD, M. D., AND K. MCGARRY (1995): "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *The Journal of Human Resources*, 30, S268–S292.

——— (2002): "The Predictive Validity of Subjective Probabilities of Survival," *The Economic Journal*, 112(482), 966–985.

KAPTEYN, A., R. ALESSIE, AND A. LUSARDI (2005): "Explaining the wealth holdings of different cohorts: Productivity growth and Social Security," *European Economic Review*, 49(5), 1361 – 1391.

KLEINJANS, K. J., AND A. VAN SOEST (2010): "Nonresponse and Focal Point Answers to Subjective Probability Questions," *Paper presented at the Netspar Pension Workshop, Amsterdam, January 2010*.

LEE, L.-F. (1984): "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity," *Econometrica*, 52(4), pp. 843–863.

MANSKI, C. F. (2004): "Measuring Expectations," *Econometrica*, 72(5), 1329–1376.

MANSKI, C. F., AND F. MOLINARI (2010): "Rounding Probabilistic Expectations in Surveys," *The Journal of Business and Economic Statistics*, 28(2), 219 – 231.

NYHUS, E. K. (1996): "The VSB-CentER Saving Project: Data Collection Methods, Questionnaires and Sampling Procedures," Progress Report 42, VSB-CentER Savings Project, Tilburg University.

O'DONNELL, O., F. TEPPA, AND E. VAN DOORSLAER (2008): "Can subjective survival expectations explain retirement behaviour?," Discussion Paper 188, DNB Working Papers.

PAGAN, A., AND F. VELLA (1989): "Diagnostic Tests for Models Based on Individual Data: A Survey," *Journal of Applied Econometrics*, 4, pp. S29–S59.

PESARAN, M. H., AND M. WEALE (2006): *Survey Expectations,*vol. 1 of *Handbook of Economic Forecasting*, chap. 14, pp. 715 – 776. Elsevier.

STEPHENS, M. (2004): "Job Loss Expectations, Realizations, and Household Consumption Behavior," *Review of Economics and Statistics*, 86(1), 253–269.

STICHTING VAN DE ARBEID (2010): "Pension Accord Spring 2010," Agreement between Dutch unions and employer associations.

VAN DER WIEL, K. (2008): "Preparing for Policy Changes: Social Security Expectations and Pension Scheme Participation," Discussion Paper 3623, IZA.

Appendix. Detailed computation replacement rates

$$
\begin{aligned}
E(RR) &= \int_0^{\infty} RR f(RR)\, dRR = \int_0^{\infty} RR \frac{\partial F(RR)}{\partial RR}\, dRR \\
&= \int_0^{50} \mathbb{P}(0 \le RR < 50) \frac{RR}{50}\, dRR + \int_{50}^{60} \mathbb{P}(50 \le RR < 60) \frac{RR}{10}\, dRR + \int_{60}^{70} \mathbb{P}(60 \le RR < 70) \frac{RR}{10}\, dRR \\
&+ \int_{70}^{80} \mathbb{P}(70 \le RR < 80) \frac{RR}{10}\, dRR + \int_{80}^{90} \mathbb{P}(80 \le RR < 90) \frac{RR}{10}\, dRR + \int_{90}^{100} \mathbb{P}(90 \le RR < 100) \frac{RR}{10}\, dRR \\
&+ \mathbb{P}(RR = 100) \cdot 100 + \int_{100}^{120} \mathbb{P}(100 < RR < 120) \frac{RR}{20}\, dRR \\
&= 25 \cdot \mathbb{P}(0 \le RR < 50) + 55 \cdot \mathbb{P}(50 \le RR < 60) + 65 \cdot \mathbb{P}(60 \le RR < 70) \\
&+ 75 \cdot \mathbb{P}(70 \le x < 80) + 85 \cdot \mathbb{P}(80 \le RR < 90) + 95 \cdot \mathbb{P}(90 \le RR < 100) \\
&+ 100 \cdot \mathbb{P}(RR = 100) + 110 \cdot P(100 < RR < 120) \\
E(RR^2) &= \int_0^{\infty} RR^2 f(RR)\, dRR = \int_0^{\infty} RR^2 \frac{\partial F(RR)}{\partial RR}\, dRR \\
&= \int_0^{50} \mathbb{P}(0 \le RR < 50) \frac{RR^2}{50}\, dRR + \int_{50}^{60} \mathbb{P}(50 \le RR < 60) \frac{RR^2}{10}\, dRR + \int_{60}^{70} \mathbb{P}(60 \le RR < 70) \frac{RR^2}{10}\, dRR \\
&+ \int_{70}^{80} \mathbb{P}(70 \le RR < 80) \frac{RR^2}{10}\, dRR + \int_{80}^{90} \mathbb{P}(80 \le RR < 90) \frac{RR^2}{10}\, dRR + \int_{90}^{100} \mathbb{P}(90 \le RR < 100) \frac{RR^2}{10}\, dRR \\
&+ \mathbb{P}(RR = 100) \cdot 100^2 + \int_{100}^{120} P(100 < RR < 120) \frac{RR^2}{20}\, dRR \\
&= 833\frac{1}{3} \cdot \mathbb{P}(0 \le RR < 50) + 3033\frac{1}{3} \cdot \mathbb{P}(50 \le RR < 60) + 4233\frac{1}{3} \cdot \mathbb{P}(60 \le RR < 70) \\
&+ 5633\frac{1}{3} \cdot \mathbb{P}(70 \le RR < 80) + 7233\frac{1}{3} \cdot \mathbb{P}(80 \le RR < 90) + 9033\frac{1}{3} \cdot \mathbb{P}(90 \le RR < 100) \\
&+ 10000 \cdot \mathbb{P}(RR = 100) + 12133\frac{1}{3} \cdot \mathbb{P}(100 < RR < 120)
\end{aligned}
$$