

Evaluating the accuracy of counterfactuals
The role of heterogeneous expectations in life cycle
models*

Jochem de Bresser[†]
Tilburg University[‡]
Netspar[§]

January 7, 2020

Abstract

This paper shows that subjective survey information on survival expectations can be used to improve the out-of-sample predictions of a dynamic structural model of labor supply, benefit claiming and saving. We consider three approaches to model survival: life tables, average subjective expectations and individual-specific estimates based on reported survival probabilities. The models are estimated on Dutch data from the 1990s, a period during which workers could retire from age 59 at no actuarial penalty to pension benefits. Such actuarial adjustments were introduced in the early 2000s and we use data from the period 2006-2016 to evaluate the accuracy of the counterfactual predictions. While the three models yield different preference estimates, their within-sample fit is similar. Out-of-sample forecasts do differ markedly. Both models based on fixed expectations anticipate a 5-year increase in the average retirement age in the new regime, compared with an observed increase of 2.6 years. The model with heterogenous expectations, on the other hand, predicts a more realistic increase of 2.5 years. We conclude that expectations matter when it comes to counterfactual predictions, even if different combinations of preferences and expectations appear equivalent within a given institutional setting.

Key words: Subjective expectations, life cycle model

JEL-codes: D84; J14; C34

*This work is part of the research programme Innovational Research Incentives Scheme Veni with project number 451-15-018, which is financed by the Netherlands Organisation for Scientific Research (NWO). I thank Max Groneck and David Sturrock for their careful discussions and Jeff Campbell and seminar participants at Tilburg University, the 2018 Netspar Pension Day and the 2019 Netspar Pension Workshop for their feedback.

[†]Email: j.r.debresser@uvt.nl.

[‡]Tilburg University, P.O. Box 90153, 5000 LE Tilburg the Netherlands.

[§]Netspar, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

1 Introduction

A life cycle model of labor supply and saving provides a lens through which one can understand observed patterns of wealth and work. That lens can be turned to the future to anticipate how those patterns might change if the institutional context is altered, for instance by changes to the pension system. Empirical analyses of such models infer preferences from observed behavior (e.g. French, 2005; De Nardi et al., 2010; French and Jones, 2011). However, agents are assumed to be forward looking, so they base their decisions on expectations as well as inclinations. Saving, for instance, is determined not only by patience, but also by one's perceived longevity. This poses an identification problem since usually neither preferences nor beliefs are directly observed (Manski, 2004). The solution is to pin down expectations using additional information and estimate preferences conditional on those beliefs. Survival expectations are typically approximated by actuarial life tables. In this paper we enrich a structural retirement model with subjective survival expectations based on probabilities reported by survey respondents. We show that doing so improves the accuracy of counterfactual predictions. The ability to capture variation in expectations across individuals in a tractable way is the crucial advantage of our approach.

We formulate a dynamic model of the retirement incentives in place in the Netherlands during the 1990s. Agents choose their labor supply, benefit claiming and saving and the model captures all major sources of income (work, pensions, social insurance and capital income). Future wages, health and longevity are uncertain. Longevity is modelled according to three methods. The first is the baseline approach of combining actuarial tables with survey data on the likelihood of death conditional on current health. This is the usual way beliefs are handled in empirical life cycle models and neither reflects the average level of subjective expectations nor any variation. The second approach sets the *level* of expectations in accordance with subjective probabilities and equates the probability of death with the average reported probability. The third technique uses reported probabilities to their full potential

and introduces individual-level heterogeneity in subjective survival. Both procedures that use subjective beliefs are based on a measurement model that takes into account survey-induced rounding of reported probabilities. Importantly, the incorporation of heterogeneity in subjective survival only introduces a single additional state variable to the life cycle model. It relates behavior to data, rather than increasing flexibility through latent preference types that complicate identification. Furthermore, the method can be applied to other countries and settings, since subjective survival has been elicited in most major household panels (e.g. the HRS in the U.S., SHARE in Europe and the LISS panel in the Netherlands). We estimate preference parameters by matching simulations from the model to observed labor supply, benefit claiming and wealth profiles in the DNB Household Survey (DHS) and document how expectations affect preference estimates, model fit and the simulated impact of a large pension reform.

This paper contributes to the literatures on life cycle models and subjective survival. While lifespan uncertainty and detailed descriptions of pension systems have been central to life cycle retirement models since the early 1990s, much progress has been made in the modelling of uncertainty, budgets and decisions.¹ More recent models add uncertainty in wages and unemployment (Gourinchas and Parker, 2002; French, 2005; Low et al., 2010; De Nardi et al., 2018) and in future health and medical expenditures (De Nardi et al., 2010).²³ Moreover, models take into account public and private health insurance (Rust and Phelan, 1997; Van der Klaauw and Wolpin, 2008; De Nardi et al., 2010; French and Jones, 2011) and means-tested government transfers (Hubbard et al., 1994, 1995) to represent household bud-

¹See De Nardi et al. (2016) for a review that is focused on saving and Blundell et al. (2016) for an overview of the literature on retirement.

²Potentially catastrophic medical costs are important in the U.S. However, the Dutch were covered by universal health insurance during our sample period. Therefore, the model discussed below does not feature uncertain medical expenditures.

³Empirical work to date assumes that agents know the distribution from which future states are drawn: it allows for risk, but not for ambiguity. In a theoretical contribution Groneck et al. (2016) show that ambiguity in survival expectations can explain under-saving at younger ages and slow decumulation of assets at advanced age, two stylized facts that characterize wealth profiles in the U.S.

gets. Heyma (2004) uses Dutch data from the 1990s and shows that it is important to model pensions and social insurance schemes jointly, since they provide alternative pathways into retirement. The literature on subjective survival has established that reported survival probabilities correlate with risk factors in plausible ways and that they predict actual mortality even when controlling for self-reported health (Hamermesh, 1985; Hurd et al., 1998; Hurd and McGarry, 1995, 2002; Gan et al., 2005).⁴ Hence, those probabilities contain information that life cycle models suggest should be relevant for behavior.

Though all modern life cycle models feature uncertain mortality and data on subjective longevity are widely available, relatively few papers have combined these literatures to date and most that do focus exclusively on consumption and saving.⁵ Wu et al. (2015) show that subjective survival expectations are not proportional to life tables, since people are both pessimistic about survival to younger ages and optimistic about survival to older ages. These deviations from actuarial forecasts affect saving in a calibrated life cycle model. Heimer et al. (2018) confirms the pattern of pessimism followed by optimism and uses an estimated model to show that this can explain both low saving early in life and slow decumulation of wealth among the elderly. However, they do not consider individual-level differences in expectations. Gan et al. (2015) formulate a simple model of saving and show that individual-specific survival curves based on subjective data improve model predictions of wealth holdings. They show that subjective beliefs lead to different preference estimates compared to life tables, raising risk aversion and lowering the discount factor. The only study to embed subjective survival in a model of both labor supply and saving is Van der Klaauw and Wolpin (2008), which does not allow for individual-level heterogeneity in expectations.

⁴See Manski (2004) and Hurd (2009) for reviews of the literature on subjective expectations.

⁵Subjective beliefs have been used in structural models of other decisions, mostly where they concern outcomes of decisions rather than exogenous processes such as mortality (Van der Klaauw, 2012). Examples include the perceived effectiveness of contraception (Delavande, 2008) and the career perspectives of college majors (Zafar, 2011; Stinebrickner and Stinebrickner, 2013).

The present study contributes to previous work in several ways. It analyzes individual-level heterogeneity in mortality beliefs in the context of a structural model of retirement and saving, which remains a gap in the literature (De Nardi et al., 2016). Our model is richer than that of Gan et al. (2015) and other related studies in terms of decisions, uncertainty and institutions. This allows analysis of the implications of expectations for the key outcomes of labor supply and pension claiming as well as saving. Moreover, while Gan et al. (2015) compare life tables with heterogeneous subjective beliefs, we shift the level of expectations and add heterogeneity in separate steps, disentangling their effects. Most importantly, this study is the first to validate different versions of a retirement model through counterfactual predictions in a substantially different policy environment. There are at least two reasons why such out-of-sample predictions are the best yardstick to compare models. Firstly, the number of state variables in dynamic programming models is limited by the curse of dimensionality. However, omitting relevant variables that drive retirement may affect policy evaluations (Blundell et al., 2016). Secondly, models that are observationally equivalent for a given dataset may generate different behavior under counterfactual policies (Blundell et al., 2016). This problem is exacerbated when latent preference types are introduced to increase flexibility. While the potential of out-of-sample evaluations to overcome these challenges has been acknowledged, only a single study uses exogenous policy variation to validate a life cycle model. Low et al. (2018) use quasi-experimental variation in the imposition of time limits on receipt of welfare in the U.S. and check whether their model of welfare claiming, employment and divorce reproduces reduced form estimates. While our empirical approach is similar, we model a different set of decisions. Furthermore, we use the policy change to compare three different model specifications, rather than vindicate a single model.

We validate the models by comparing forecasts of the impact of a pension reform with observed behavior. Whereas financial incentives favored early retirement during the 1990s, actuarial adjustments were subsequently introduced that lowered pensions if they were claimed

prior to age 65 and rewarded delayed claiming. As a result the average age of retirement in our sample increases from 61.2 before the reform to 63.8 after the introduction of actuarial adjustments. Quasi-experimental evidence confirms that this delay in retirement is caused by the policy reform (Euwals et al., 2010) and many studies use it as a first stage to measure the effect of retirement on various outcomes (e.g. Montizaan et al., 2010; Grip et al., 2011; Montizaan and Vendrik, 2014; Montizaan et al., 2015). We verify whether the models can anticipate these changes in retirement when estimated solely on data from the 1990s (the old policy regime).

The results show that estimated preferences do depend on survival expectations and that the interaction between the two affects counterfactual predictions. In particular, heterogeneous expectations lead to more risk aversion and a lower and more plausible estimate for the rate of time preference as in Gan et al. (2015). Moreover, they lower the estimated weight on consumption relative to leisure compared to the two models without variation in expectations. However, while model fit within the estimation sample is slightly better for the model with variation in beliefs, we find that all three models show the same strengths and shortcomings. Labor supply and pension claiming are matched reasonably well, though the model over-predicts labor supply in good health prior to age 60. Claiming of disability insurance is reproduced accurately up to the eligibility age for occupational pensions. After age 60 model-implied claiming drops to zero whereas the data show continued use of disability benefits. All wealth quartiles are matched closely up to age bin 70-74, after which confidence intervals for the median and 75th percentile become wide. These features are common to all three specifications of beliefs, so behavior in the estimation sample is largely equivalent across the three models. This is not true for the out-of-sample predictions under the counterfactual of actuarial adjustments to pensions. While the data indicate that the average retirement age increased by 2.6 years to 63.8 after early retirement pensions became less generous, both models with fixed expectations predict a larger increase of 5 years to a retirement age above

65. The model with variation in survival expectations anticipates a more reasonable increase of 2.5 years to an average retirement age of 63.5. Our analysis shows that this improvement in accuracy is driven by the estimated parameters for risk aversion, the consumption weight, the discount factor and the importance of bequests. The estimates obtained under heterogeneous expectations improve counterfactual predictions both overall and conditional on beliefs.

The rest of the paper is organized as follows. Section 2 describes our life cycle model. The estimation routine is explained in section 3, after which section 4 introduces the data. Section 5 contains the results and section 6 concludes. A detailed description of the model used to measure subjective survival expectations is provided in Appendix A.

2 Model

We model the retirement, benefit claiming and saving decisions of Dutch men aged 50 and older at a yearly resolution. The model is based on De Bresser et al. (2017). Individuals derive utility from leisure, consumption and bequests. They face uncertainty in longevity, health and wages. Our model captures all social insurance schemes in place during the sample period and in particular those that provide income in retirement, most importantly the flat-rate public pension and occupational pensions that depend on one's earnings history. Furthermore, unemployment and disability insurance provide alternative pathways into retirement. We estimate preferences on data from the period 1993-2001, during which early retirement was possible from age 59 on extremely generous terms.

In addition to evaluating model fit, we also check the accuracy of predictions in a different policy regime. To this end we use the estimates to simulate the impact of introducing actuarial adjustments in occupational pensions and changing the tax function. This policy regime corresponds to the situation in place in the period 2006-2016 and we compare the

implications of the model with observed behavior during that period. The following subsections explain the model in more detail.

2.1 Decisions

Labor supply and benefit claiming decisions are discrete and consumption is a continuous choice variable. Between the ages (t) of 50 and 69 agents choose one of four levels of labor supply h_t in hours per year: $h_t \in \{0, 1500, 2000, 2500\}$. We assume that nobody works beyond age 70. Disability (DI) and unemployment (UI) benefits can be claimed between ages 50 and 64 after which both are replaced by the public pension. An individual has to be in poor health in order to be able to claim DI. In the Netherlands the onset of public pension benefits is automatic at the age of eligibility, which was 65 for the periods covered by our samples. While people with entitlements can choose to start claiming occupational pensions at any age between 59 and 69, the incentives embedded in the benefit formula push workers to retire early. In contrast to the DI and UI schemes, claiming occupational pensions is an absorbing state. One cannot claim pension, DI or UI benefits and work at the same time. All discrete choices are thus limited to the younger ages and consumption and hence saving decisions are made until the individual dies (longevity is random and capped at age 100). Figure 1 illustrates the decisions available in the model at all ages.

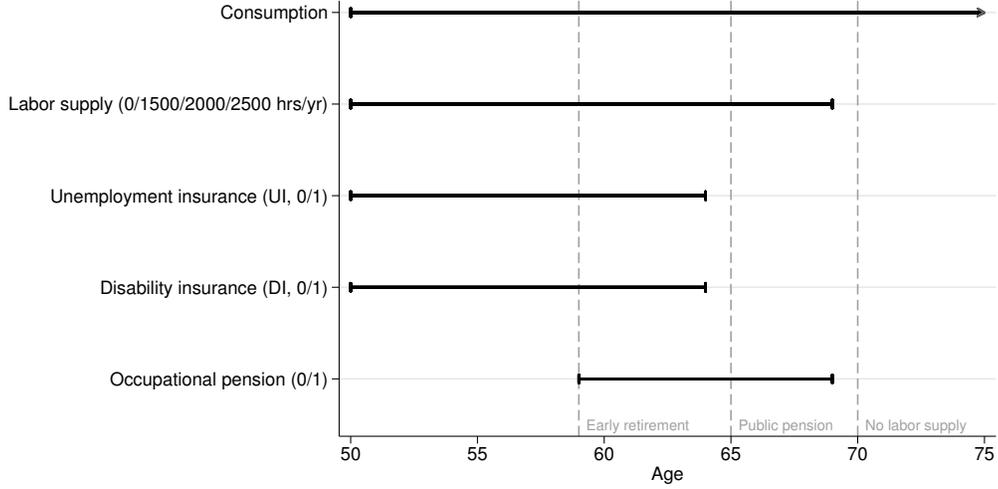


Figure 1: Decisions available at various ages

2.2 Preferences

Individuals derive utility from consumption (c_t), leisure (l_t) and, potentially, from bequests.

The utility function is given by

$$u(c_t, l_t) = \frac{n_t}{1 - \sigma} \left[\left(\left(\frac{c_t}{n_t} \right)^\kappa l_t^{1-\kappa} \right)^{1-\sigma} - 1 \right] \quad (1)$$

where n_t is an equivalence scale that increases with household size according to $n_t = hsize^{0.7}$.

Parameter σ determines the concavity of the utility function and thus both risk aversion and the intertemporal elasticity of substitution and κ sets the relative weights of consumption and leisure. Leisure depends on labor supply h_t through the time constraint

$$l_t = 4,000 - h_t - \gamma \mathbb{I}\{h_t > 0\} - \delta \mathbb{I}\{\text{health} = \text{bad}\} - \text{stigma costs} \quad (2)$$

The total time endowment is set to 4,000 hours per year and individuals incur a fixed cost γ when working a positive number of hours. Moreover, working is more costly when in bad health, which is captured by the time cost δ . Claiming either UI or DI benefits entails stigma

costs, also measured in hours of leisure:

$$\text{stigma costs} = \phi \mathbb{I} \{ \text{claim DI} \} + \xi \mathbb{I} \{ \text{claim UI} \} \quad (3)$$

Given that γ , δ , ϕ and ξ are all measured in hours per year, these parameters are constrained to be between 0 and 4,000.

The bequest motive is important to fit (dis-)saving at older ages. We take the bequest utility function from French (2005) and add one parameter to gain the flexibility required to fit wealth profiles that span all ages from 50 to 84. We model wealth across a broader age range than is typically done: usually the focus is either on the period around retirement, age 50-70 as in French (2005), or exclusively on retirees as in De Nardi et al. (2010). Previous research based on actual bequests shows that on average widow(-er)s with children do not leave behind larger bequests than those without offspring (Hurd, 1989; Kopczuk and Lupton, 2007). In line with that finding, bequests are allowed to be targeted especially to one's spouse, rather than one's children, by making the bequest weight a function of the equivalence scale n_t :

$$b(w_t, n_t) = \exp[\theta_0 + \theta_1 n_t] \times \frac{(w_t + K)^{(1-\sigma)\kappa}}{1 - \sigma} \quad (4)$$

The wealth level is given by w_t and parameters θ_0 and K govern the importance of bequests and the curvature of bequest utility respectively. θ_1 is a new parameter that allows the utility derived from bequests to be higher when the equivalence scale is high, that is, for ages at which many spouses are alive.

2.3 Survival expectations

We pay particular attention to modelling survival expectations. The typical approach when estimating a life cycle model is to combine survey information on health and survival with actuarial life tables, e.g. the Human Mortality Database, to infer the probability of death at a certain age conditional on prior health (e.g. French, 2005). As explained in the introduction, such method may get the level of survival probabilities wrong and does not allow one to capture heterogeneity. While we implement the life table technique as a baseline, we also draw on reported probabilities to approximate expectations.

One complication when analyzing subjective probabilities of any type is rounding: the tendency for respondents to report one number (10%) when the true subjective probability lies in some interval (5-15%) (Manski and Molinari, 2010; De Bresser and Van Soest, 2013; Kleinjans and Van Soest, 2014; Bissonnette and de Bresser, 2018). Rounding is particularly important for survival probabilities in the DHS, since respondents are asked to report on an 11-point scale ranging from 0 (no chance) to 10 (complete certainty). Previous research indicates that such coarse data has adequate test re-test validity when compared with expectations elicited on a finer scale, especially if one models the rounding that is enforced by the survey (De Bresser, 2017). Therefore, we employ a model similar to that estimated by De Bresser (2017) to take rounding into account. The model assumes that reported probabilities are generated from a Gompertz distribution, which has one common baseline hazard for all individuals that is affected proportionately by health and by unobserved heterogeneity. Integrating over the distribution of unobserved heterogeneity, the estimates are used to back out the average survival probabilities net of rounding. Moreover, we calculate the posterior mean individual effect for each respondent conditional on reported probabilities. Those means allow us to introduce heterogeneity in survival expectations into the life cycle model through a single additional state variable. A detailed description of the data and the measurement model for expectations can be found in Appendix A.

2.4 Health and wages

While expected longevity is a key driver of behavior in intertemporal models, agents also face other types of risk. In particular, we allow future health and wages to be uncertain. Health m_t can take two levels: good (2) or bad (1). The probability of being healthy next year is a first order Markov process that depends on age:

$$\Pr(m_{t+1} = 2 | m_t, t) = \frac{\exp[\mu_0 + \mu_1 t + \mu_2 \mathbb{I}\{m_t = 2\}]}{1 + \exp[\mu_0 + \mu_1 t + \mu_2 \mathbb{I}\{m_t = 2\}]} \quad (5)$$

Future wages are also uncertain and evolve according to an AR(1) process as in Gourinchas and Parker (2002). We estimate the parameters of wage transitions from the European Community Household Panel (ECHP).

We use subjective expectations to model longevity, but unfortunately the DHS does not contain similar subjective data regarding either health or wages. Hence, we follow the default approach and model those processes based on the transition rates observed in the data.

2.5 Budget constraint

Decision makers cannot consume more than cash-on-hand x_t :

$$x_t = w_t + \tau(\text{earn}_t, DIinc_t, UIinc_t, \text{pubpens}_t, \text{occpens}_t, w_t, t) + inc_t^{\text{SP}} - OOP_t \quad (6)$$

Here $\tau(\cdot)$ is the tax function that calculates net income from earnings, DI and UI benefits, public and occupational pensions, and wealth. Public pensions provide a fixed subsistence income from age 65 onward. Occupational pensions are set as a replacement rate relative to final earnings and thus depend on one's work history. More information on these income sources is given in Appendix B and Appendix C provides a detailed description of the tax function $\tau(\cdot)$.

Wealth w_t , which is restricted to be positive, enters the budget constraint directly and through the tax function, since it generates capital income at a rate of 4% and is taxed at 0.7%.⁶ In addition to this net income after taxes, the household budget also includes an exogenous stream of net income from the spouse given by inc_t^{sp} . This does not enter the tax function, because income is taxed at the level of the individual. Finally, out-of-pocket medical expenditures OOP_t are at the household level. During both periods studied in this paper medical expenditures were of limited importance in the Netherlands and consisted mostly of monthly premiums for mandatory health insurance. Hence, we do not model medical expenditures as an additional source of uncertainty and consider only the mean expenditure by health and age.

Government transfers ensure subsistence expenditures to households whose cash-on-hand drops below a minimum consumption level (Hubbard et al., 1994, 1995):

$$x_t = \max \{x_t, n_t c_{\min}\} \tag{7}$$

The minimum consumption level c_{\min} is set to 7,000 Euro per year.

2.6 Policy reform

The Netherlands provides an ideal setting to study the accuracy of out-of-sample predictions of retirement models, because occupational pensions were reformed drastically in the early 2000s. We explain the most important aspects here and refer the reader to Appendix B for detailed descriptions of both sets of institutions. In the old system occupational pensions, which could not be claimed while working, were divided into two tiers: early retirement pensions (ages 59-64) and regular pensions (ages 65+). Motivated by the notion that early retirement of older workers would improve the employment rate of the young, the key feature

⁶During the period spanned by our estimation sample, 1993-2001, wealth tax was levied on the stock w_t rather than on capital income.

of the system was a complete absence of incentives to postpone retirement. There was no actuarial adjustment to early benefits, which were set at 85% of final earnings. Moreover, one would continue to accumulate regular pension entitlements while on early retirement: years in early retirement counted as working years as far as subsequent pensions were concerned. At age 65 early retirement pensions were automatically supplanted by regular pensions, which replaced 1.75% of final earnings for each year of work or early retirement.

During the early 2000s actuarial adjustments were introduced for cohorts of workers born in 1950 or later. The two tiers of occupational pensions are now combined into a single system, in which pensions can first be claimed between age 60 and 70. Claiming prior to age 65 lowers benefits by 6% per year and later claiming is rewarded at 7% per year. Furthermore, one can no longer accumulate pension rights while claiming benefits.

A second change concerned taxes: a new law reformed the system of income taxation in 2001. As a consequence the marginal effective tax rate, including tax credit and contributions to social insurance schemes, increased by 5pp for individuals earning between 20,000 and 40,000 Euro per year and dropped by 11-14pp for those with either lower or higher incomes (see Appendices B and C for details).

2.7 Solving the model

Agents in the model are fully rational and take into account all future consequences of their behavior. At age t the individual is described by state \mathbf{s}_t that contains the eight state variables health (m_t), years worked ($yrswrk_t$), previous earnings ($prevearn_t$), wage (j_t), wealth (w_t) and three variables that capture the claiming and entitlement status for occupational pensions and disability and unemployment insurance ($PENstat_t$, $DIstat_t$ and $UIstat_t$):

$$\mathbf{s}_t = (m_t, yrswrk_t, prevearn_t, j_t, w_t, PENstat_t, DIstat_t, UIstat_t) \quad (8)$$

As explained above, health is a binary variable that is either good ($m_t = 2$) or bad ($m_t = 1$). Years worked, previous earnings, wage and wealth are four continuous state variables that we discretize in order to solve the model.⁷ Note that though years worked appears discrete due to the yearly resolution, these years accumulate at a rate of 0.375 and 0.7 respectively when claiming UI or DI and are reduced in case of part time work. Occupational pension status ($PENstat_t$) is discrete and takes three values: 1 for workers who have no entitlement because their employer does not offer a pension; 2 for workers who have an entitlement but have not claimed yet; and 3 for retirees who are currently claiming occupational pensions. Similarly, $DIstat_t$ distinguishes between those who claimed disability insurance in the previous period ($DIstat_t = 2$) and those who did not ($DIstat_t = 1$). $UIstat_t$ keeps track of the number of years one can claim UI benefits:

$$UIstat_t = \begin{cases} 1 & \text{if entitlement is 3 years;} & 2 & \text{if entitlement is 2 years} \\ 3 & \text{if entitlement is 1 year;} & 4 & \text{if entitlement is 0 years} \end{cases} \quad (9)$$

While we do reduce one's UI entitlement with each year of claiming, we do not take into account that entitlements increase by $1/12$ year for each year of work in order to simplify the state space. The model with heterogeneous survival expectations includes an additional continuous state variable that captures expectations (which we discretize into 7 levels).

Decisions are taken to maximize the sum of instantaneous utility derived from consumption and leisure plus the expected present value of future utilities. Denote the set of discrete choices available in state \mathbf{s}_t at age t by $\mathcal{D}(\mathbf{s}_t, t)$. A possible option d in this set specifies labor supply and the decision to claim DI, UI and occupational pensions, but *not* the level

⁷Discretization on square-root grids is based on 5 levels of years worked (10-50 years), 10 levels of previous earnings (0-250,000 Euro/year), 7 levels of wages (5-100 Euro/hour) and 10 levels of wealth (0-1,000,000 Euro).

of consumption. For all $d \in \mathcal{D}$ we calculate the value corresponding to that decision as:

$$\begin{aligned}
V^d(\mathbf{s}_t, t) = \max_{c_t} \{ & u(c_t, l_t) \\
& + \beta [(1 - p_s(m_t, t)) b(w_{t+1}, n_{t+1}) + p_s(m_t, t) \mathbb{E}_t [V(\mathbf{s}_{t+1}, t + 1)]] \} \\
\text{s.t. } & c_t \leq x_t \text{ (budget constraint)}
\end{aligned} \tag{10}$$

where c_t is consumption; β is the discount factor; $p_s(m_t, t)$ is the probability of surviving another year conditional on current age and health; \mathbb{E}_t is the expectation across the distributions of future health and wage conditional on the current state; and x_t is cash-on-hand. In other words: V^d is the highest value the individual can obtain if he takes decision d now and continues to make optimal choices in the future. The agent optimally selects the discrete choice d^* that yields the highest value:

$$V^{d^*}(\mathbf{s}_t, t) = \max_{d \in \mathcal{D}(\mathbf{s}_t, t)} \{V^1(\mathbf{s}_t, t), \dots, V^D(\mathbf{s}_t, t)\} \tag{11}$$

where D is the number of discrete choice alternatives available in the given state \mathbf{s}_t at age t . We find the optimal policies for discrete choices and consumption by means of backward induction.

3 Estimation

We estimate all processes that can be estimated outside the dynamic model separately and focus here on the estimation of preference parameters. Preferences are determined by the 10 parameters collected and explained in Table 1. Our Method of Simulated Moments (MSM) estimation algorithm follows French (2005), De Nardi et al. (2010) and French and Jones (2011). For a given parameter vector, we simulate the life cycles of 5,000 workers using initial conditions taken from the data. Various target moments are calculated and compared

Table 1: Overview of preference parameters

Parameter	Explanation
Utility function^a	
σ	Concavity of utility function: intertemporal rate of substitution and risk aversion.
κ	Relative weight of consumption compared to leisure.
β	Discount factor.
γ	Fixed cost of working positive hours (hours per year).
δ	Leisure cost of being in poor health (hours per year).
ϕ	Stigma cost of claiming disability insurance (hours per year).
ξ	Stigma cost of claiming unemployment insurance (hours per year).
Bequests^b	
θ_0	Intercept of weight of bequests relative to utility from consumption and leisure.
θ_1	Coefficient on equivalence scale in bequest weight.
K	Curvature parameter of bequests (2012 Euro).
^a $u(c_t, l_t) = \frac{n_t}{1-\sigma} \left[\left(\left(\frac{c_t}{n_t} \right)^\kappa l_t^{1-\kappa} \right)^{1-\sigma} - 1 \right];$ $l_t = 4,000 - h_t - \gamma \mathbb{I}\{h_t > 0\} - \delta \mathbb{I}\{\text{health} = \text{bad}\} - \phi \mathbb{I}\{\text{claim DI}\} - \xi \mathbb{I}\{\text{claim UI}\}$	
^b $b(w_t, n_t) = \exp[\theta_0 + \theta_1 n_t] \times \frac{(w_t + K)^{(1-\sigma)\kappa}}{1-\sigma}$	

to data by means of a method of moments objective function. The MSM estimator is the parameter vector that maximizes the fit of simulations to data moments. Finding this vector is complicated by the nature of the objective function, which is not uniformly differentiable and has multiple local minima. In light of such difficulties, we prefer simulated annealing over gradient-based optimization methods (we use the variant explained in Goffe et al., 1994). Since the MSM estimation approach is standard for dynamic models, we refer the reader to the references mentioned above for more details.

The moments targeted in estimation concern labor supply, benefit claiming and wealth. For labor supply we match average yearly hours worked and participation rates by health status for two-year age bins ranging from 50 to 65. These labor supply profiles yield a total of 32 moments ($2 \times 2 \times 8$). For benefit claiming we calculate the fraction of individuals in different two-year age brackets who claim a certain type of benefits. Unemployment and disability benefits can be claimed up to the receipt of the public pension at age 65, so we match eight moments for each. Occupational pensions can be claimed from age 59 onward, for

a total of four moments. Regarding wealth we target quartiles by age brackets. There are ten two-year brackets per quartile for the ages 50-69. At ages 70-84 quartiles are calculated for broader five-year brackets, because we have fewer observations. There are thirteen moments for each quartile (ten for ages 50-69 and three for ages 70-84). We match 91 moments in total ($32 + 8 + 8 + 4 + 39$).

While all parameters are affected by all moments in complex nonlinear models, it is nonetheless useful to provide a heuristic discussion of the identification of parameters in terms of the moments that are most important to pin them down (Eckstein et al., 2019). This is straightforward for the stigma costs of DI and UI, which are primarily determined by the rates at which those benefits are claimed. Similarly, the leisure cost of poor health is driven by the difference in labor supply between poor and good health and the fixed cost of work by the combination of hours worked and the participation rate (higher fixed costs result in fewer hours of work at a given level of participation). It is more difficult to link the remaining parameters to specific groups of moments, since risk aversion, patience, the consumption weight and the bequest motive all have a profound effect on labor supply, benefit claiming and wealth. We verified that the moments identify these parameters by fixing key parameters at different levels and estimating the rest. Comparing function values for these constrained estimation runs confirmed that the model cannot rationalize observed moments if we fix risk aversion, the discount factor or the slope that relates household size to the importance of bequests at levels that are different from those obtained in unconstrained optimization.

Our GMM estimator does not use an optimal weighting matrix, because many covariances between moments would be estimated based on little data. To safeguard robustness we use a diagonal weighting matrix that contains only inverses of the estimated variances of sample moments on the diagonal. The standard errors for the preference parameters reported below take this into account.

Table 2: Overview of sample selection due to different criteria (sample sizes)

	Estimation sample (1993-2001)	Simulation sample (2006-2016)
Auxiliary processes (all men 50+)	7,484 (100%)	10,433 (100%)
Never self-employed	7,114 (95%)	9,569 (92%)
At least 15 yrs work experience	5,363 (72%)	7,141 (68%)
Relevant cohort ^a	7,364 (98%)	4,478 (43%)
Satisfies all criteria (used in moments)	5,056 (68%)	2,482 (24%)

^a Relevant cohort was born before 1950 for the estimation sample and after 1949 for the simulation sample.

4 Data

In order to estimate the model and evaluate its predictions in a different policy regime we need rich data that cover a long period during the 1990s and 2000s. The DNB Household Survey (DHS) is ideally suited to this purpose. The DHS is a yearly survey that is administered to the CentERpanel by CentERdata, which is affiliated with Tilburg University.⁸ The CentERpanel consists of roughly 5,000 individuals in 3,000 households and is representative for the Dutch population. Panel members receive weekly questionnaires over the internet. Prospective members are randomly selected from the address registry of Statistics Netherlands and are provided with internet access and a simple computer if required for participation. The DHS contains a wealth of information on psychological and economic aspects of financial behavior and is comparable in scope to the Health and Retirement Study (HRS) in the U.S.

We model the labor supply and wealth of men who are at least 50 years old. Our sample selection proceeds as follows. The auxiliary processes of (subjective) survival, health, medical expenses and equivalence scales are estimated on all 50+ men in the relevant DHS waves (1993-2001 for the estimation sample and 2006-2016 for the out-of-sample predictions). To obtain the samples from which we calculate data moments to be matched in estimation we drop men who were self-employed at any moment and those with fewer than 15 years of work experience (we drop the self-employed because they are not covered by occupational

⁸See <https://www.centerdata.nl/en/databank/dhs-data-access> for more information.

pensions). Moreover, in order to isolate the groups with and without access to early retirement, we drop from the estimation sample (1993-2001) all men born after 1949. Similarly, we drop all men born before 1950 from the simulation sample (2006-2016). We estimate the net income stream of the spouse on the partners of the men in these reduced samples. Table 2 shows how many observations are lost at each step of sample selection when going from the sample used to estimate auxiliary processes to that from which estimation moments are constructed. For the estimation sample observations are lost mostly due to patchy records for work experience. In the simulation sample the restriction to cohorts born after 1949 results in the largest loss of data.

4.1 Survival expectations

Figure 2 shows the expectations estimated from life tables and the mean and distribution of the probabilities constructed from survey data. The figure plots the probability that an individual dies within one year conditional on reaching a given age. Looking at the top panels, one notices that the life tables approach yields much higher probabilities of dying when in poor health than do subjective expectations. The actuarial method results in a probability of death close to 1 for ages above 90, while subjective expectations in the estimation sample only increase to 0.3. Probabilities in good health, on the other hand, increase to around 0.2 for both approaches. One immediate question is whether the large differences in probabilities between health states that are produced by the life table technique are realistic. Simulation of life courses based on the probabilities in panel a. yields unrealistically short lifespans: the median individual dies at age 71. The subjective probabilities of panel b. give a more realistic, if still a bit low, median lifespan of 79 years. Such implausibly low survival expectations derived from life tables are not specific to our dataset. Appendix D shows that we estimate a similarly high probability of death when in poor health for other commonly used data, namely the HRS (U.S.) and two countries in SHARE (the Netherlands and Spain). Survival

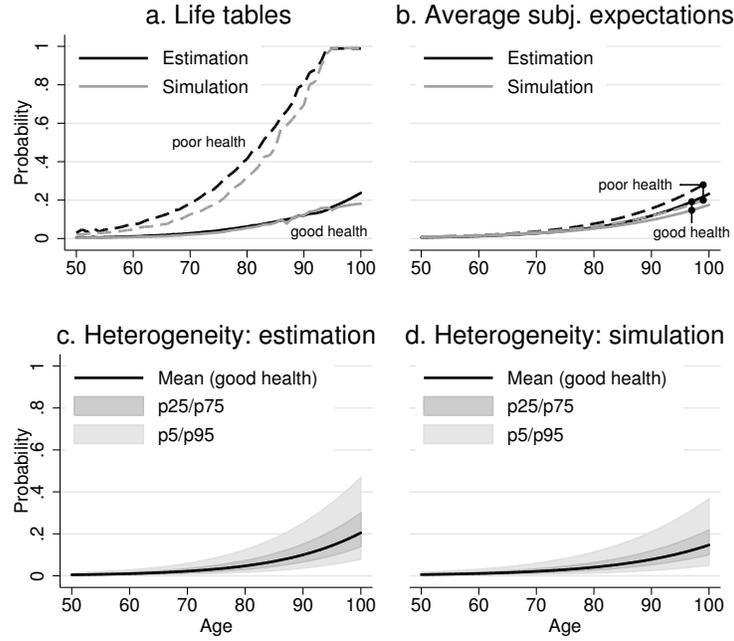


Figure 2: Estimated mortality expectations (probability of dying at a given age conditional on surviving to that age)

probabilities in poor health computed from exit surveys or other information available from surviving spouses thus leads to an overly pessimistic assessment relative to both subjective expectations and actual mortality.

In addition to its potential to better approximate the *level* of survival expectations, subjective information also allows one to capture *variation*. This variation is illustrated in the bottom panels of Figure 2. Panel c. shows the mean mortality probability conditional on good health, as well as the range for probabilities across different percentiles of the distribution of the posterior means for individual effects (all conditional on good current health). While the mean is in line with that in panel b., the variation around it is substantial. For instance, the difference between the mean and the 75th percentile is larger than that between individuals in good versus bad health. While panel c. draws on data from the estimation period, panel d. illustrates that the same pattern holds for the simulation sample. The goal

of the paper is to investigate how such level differences between actuarial and subjective expectations and variation between individuals matter for model fit and for the accuracy of out-of-sample predictions.

Estimates for the other auxiliary processes are reported in Appendix E.

4.2 Target moments

We follow the methodology described by French (2005) and use Fixed Effects (FE) models to remove cohort and family size effects from the target moments used in estimation. In particular, we estimate the following linear FE models for wealth (shown), hours worked, participation and benefit claiming:

$$w_{it} = \sum_{j=1}^J [\beta_j age_{it}^j] + \beta_{J+1} n_{it} + \alpha_i + \varepsilon_{it} \quad (12)$$

The first term on the right hand side is a linear combination of J age dummies corresponding to relevant age bins, followed by the equivalence scale n_t (for wealth the relevant age bins are two-year bins between ages 50 and 69, followed by five-year bins between ages 70 and 84). The two error components are an individual effect α_i and the idiosyncratic error ε_{it} . We calculate the following adjusted variable \tilde{w}_{it} :

$$\tilde{w}_{it} = \sum_{j=1}^J [b_j age_{it}^j] + b_{J+1} 2^{0.7} + a_i + \bar{a}_{1940} - \bar{a}_i + e_{it} \quad (13)$$

where b -coefficients are FE estimates, a_i is the estimate of the individual effect of individual i and e_{it} is the residual. Wealth is transformed in two ways. Firstly, we fix the family size at two. Secondly, we add the difference between the average individual effect in cohort 1940-1949 (\bar{a}_{1940}) and the average individual effect in the cohort of individual i (\bar{a}_i). This removes cohort effects, using the 1940s as the baseline. Other variables used in target moments are

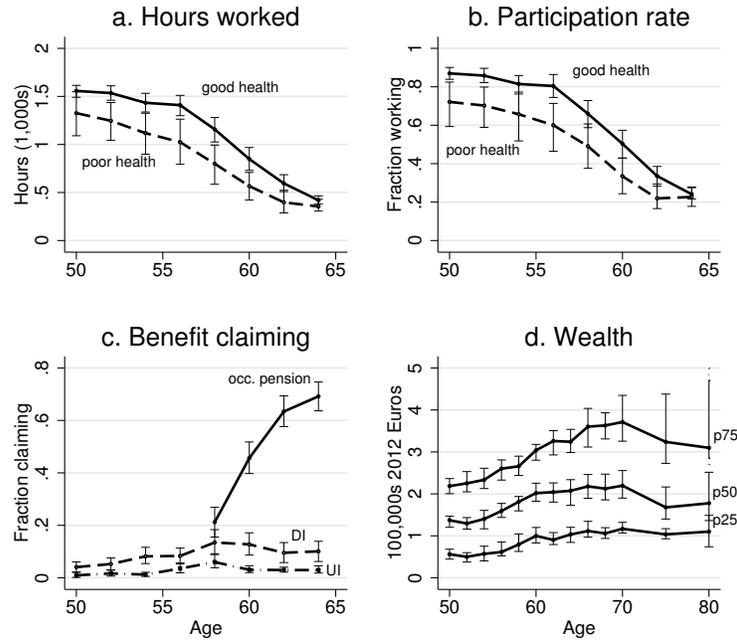


Figure 3: Moments targeted in estimation (averages for panels a.-c., quartiles for panel d.; 95% confidence intervals obtained by clustered bootstrap)

transformed in a similar way. Data for the simulation sample are adjusted with the 1950s as baseline in order minimize cohort differences between the two samples.

The data moments to be matched in the estimation of preference parameters are reported in Figure 3. These are the moments described in Section 3. Panel a. of Figure 3 shows that average labor supply in good health measured in yearly hours worked drops from 1,500 hours around age 50 to below 500 hours around age 64. Labor supply in poor health is lower, which suggests that the leisure cost of poor health is strictly positive. Participation in the labor market is equal to one if someone works positive hours and zero otherwise. It drops from close to 90% to around 20% (panel b.). Benefit claiming, as described in panel c., suggests a preference for disability over unemployment insurance. Claiming rates for the former are around 10%, while the latter is only claimed by less than 5% of the sample. Occupational pensions are claimed much more eagerly, with 20% of the 58/59 year olds claiming early

retirement benefits. This rises to 70% for 64/65-year olds. Finally, panel d. of Figure 3 reports wealth quartiles. All three quartiles increase with age up to age 70, after which they either stabilize (p25) or decline slightly (p50 and p75). However, bootstrapped confidence intervals for p75 at the age bins 75-79 and 80-84 are very wide. For p25 and p50 the data do allow us to rule out strong dissaving at advanced age, suggesting that the bequest motive may be operational.

4.3 Initial conditions

Simulated life courses start from initial conditions provided by the data. We select the first observation for each individual in the sample used to compute target moments, as long as the individual was between 50 and 54 years old at that time. We then replicate each record 50 times and randomly select 5,000 observations for which we simulate labor supply and saving according to the model.

Table 3 contains descriptive statistics of the initial conditions for both samples. At age 50 individuals are generally healthy, around 84% rate their health as “good” in both samples. Median work experience is close to 27 years, which implies that workers entered the labor market around age 23. In contrast to health and work tenure, wages, earnings and wealth do differ between samples. Wages and previous earnings were lower in the simulation sample than in the estimation sample after adjusting for inflation: the median wage dropped by five Euro to 25 Euro per hour. Similarly, previous yearly earnings were about 10,000 Euro lower in the simulation sample (43,000 compared to 53,000). Wealth, on the other hand, was higher in the simulation sample with a median of 193,000 compared to 145,000 Euro. This large increase in the value of wealth is mostly due to the development of house prices between the 1990s and 2016, which increased from an average price of 134,226 in 1995 to 233,307 in 2016 in 2012 Euro (CBS, 2017). Coverage of occupational pensions is 87% in the simulation sample, which is 7pp higher than the estimation sample. Unemployment insurance entitlements are

Table 3: Initial conditions

	a. Estimation sample			b. Simulation sample		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Health (1 = good; 0 = bad)	0.84	1	0.37	0.82	1	0.39
Wage (Euro/hr)	33.3	30.5	13.8	26.9	25.1	11.2
Years worked	25.6	26	6.7	27.3	28	5.0
Previous earnings (1,000s Euro/year)	57.7	52.6	26.8	46.3	42.8	23.1
Wealth (1,000s Euro)	168.1	144.9	138.4	215.0	193.3	171.2
Occ. pension status (1 = has pension)	0.80	1	0.40	0.87	1	0.34
DI status (0 = did not claim last year)	0	0	0	0	0	0
UI status (1 = fully, ..., 4 = not entitled)	1.4	1	0.6	1.5	1	0.6
Heterogeneity in subjective survival	-4.7	-4.7	0.6	-4.0	-3.9	0.6
N		5,000			5,000	

very similar in both periods, and disability insurance claiming status is initialized as not previously claiming in both samples. Finally, variation in expectations is similar in both samples (both standard deviations are 0.6).

5 Results

This section presents results for estimated life cycle models with the three specifications of survival expectations. The discussion proceeds in four steps. First we compare the estimated preference parameters across models. In-sample model fit is then described by comparing simulations with target moments. Thirdly, the models are tested on their ability to successfully anticipate the changes in labor supply and claiming after the attractiveness of early retirement was reduced. The accuracy of out-of-sample predictions is the key yardstick against which to measure these models, since it speaks directly to their usefulness as tools for the prediction of the effects of policy changes. Finally, we analyze how the interaction between preferences and heterogeneity in subjective survival affects counterfactual simulations.

Table 4: Estimates for preference parameters

	(1) Life tables	(2) Average subj. exp.	(3) Heterogeneous exp.
Utility function^a			
σ (concavity)	4.78 (0.000026)	4.70 (0.000031)	5.14 (0.000018)
κ (consumption weight)	0.52 (0.000086)	0.64 (0.000081)	0.43 (0.000048)
β (discount factor)	1.05 (0.000013)	1.02 (0.000039)	0.96 (0.000022)
γ (fixed cost of work; hrs/yr)	83 (7.7)	881 (5.5)	833 (7.8)
δ (time cost of bad health; hrs/yr)	381 (51.8)	303 (6.2)	310 (3.2)
ϕ (stigma cost DI; hrs/yr)	834 (147.9)	1845 (171.0)	2051 (122.8)
ξ (stigma cost UI; hrs/yr)	3093 (2.9)	3418 (1.4)	3581 (3.1)
Bequests^b			
θ_0 (intercept bequest weight)	-8.86 (0.00015)	-3.64 (0.000072)	-10.12 (0.000072)
θ_1 (slope bequest weight)	4.33 (0.00015)	5.11 (0.000062)	2.78 (0.000042)
K (concavity bequests)	941,886 (21.6)	885,709 (9.7)	302,260 (7.1)
Function value	521.75	496.22	450.21

Standard errors in parentheses.

$$^a u(c_t, l_t) = \frac{n_t}{1-\sigma} \left[\left(\left(\frac{c_t}{n_t} \right)^\kappa l_t^{1-\kappa} \right)^{1-\sigma} - 1 \right];$$

$$l_t = 4,000 - h_t - \gamma \mathbb{I}\{h_t > 0\} - \delta \mathbb{I}\{\text{health} = \text{bad}\} - \phi \mathbb{I}\{\text{claim DI}\} - \xi \mathbb{I}\{\text{claim UI}\}$$

$$^b b(w_t, n_t) = \exp[\theta_0 + \theta_1 n_t] \times \frac{(w_t + K)^{(1-\sigma)\kappa}}{1-\sigma}$$

5.1 Parameter estimates

Table 4 presents preference estimates obtained for the three models of survival expectations. The leftmost column contains baseline results for the approach that relies on life tables to approximate beliefs. This is the approach followed by most earlier studies, which serve as a guide to judge the plausibility of these estimates. The estimated concavity parameter of the utility function is 4.78, which is in the ballpark established by previous literature. Consumption is valued slightly more highly than leisure, but the consumption weight of 0.52 indicates that individuals attach importance to both. The estimated discount factor is 1.05.⁹ While this point estimate is high, it is in line with the discount factor of 1.04 reported in French (2005). Working any positive number of hours is associated with a fixed leisure cost of 83 hours per

⁹Because other studies sometimes fix the discount factor at a pre-determined value, we verified that our moments identify it by setting β to 0.97 and estimating the other parameters. Doing so yields a substantially worse objective function value.

year, which is about one third of existing estimates for the U.S. (French (2005)'s estimate is 240 hours per year). The estimated time cost of poor health is higher in the Netherlands at 381 hours per year compared with 202 for the U.S.. As a result, Dutch workers will reduce labor supply more strongly when they become unhealthy relative to their American peers. In contrast to previous studies, which tend to focus exclusively on labor supply, we introduce stigma costs for claiming disability and unemployment insurance. Such costs are required to generate reasonable rates of benefit claiming. The estimates indicate that disability and unemployment insurance come with sizeable stigma costs of 834 and 3,093 hours of leisure respectively. These costs strongly discourage claiming. As for bequests, the estimates suggest that they are directed almost exclusively at other people living in the same household, which in our sample means one's spouse. The bequest weight of a man who lives with a partner is $\exp[-8.86 + 4.33 \times 2^{0.7}] = 0.16$ which declines to $\exp[-8.86 + 4.33 \times 1^{0.7}] = 0.01$ for a widower living alone. French (2005) estimates a constant bequest weight of 0.037, which is in the same range. Moreover, the bequest curvature parameter is far from zero at 941,886, which implies that the disutility of leaving zero bequests is finite (French (2005) fixes this parameter at 500,000 USD). All standard errors are small relative to point estimates, indicating that the model is estimated with considerable precision.

The middle column of Table 4 shows that the estimates of all parameters change by many standard errors when we model survival expectations based on subjective data, even when this only adjusts the *level* of beliefs without allowing for any heterogeneity. The only exception is the time cost of poor health, which decreases by 78 hours per year or 1.5 standard error. We find that the weight of consumption in the utility function is particularly sensitive to the level of mortality expectations: it increases from 0.52 for life tables to 0.64 for average subjective expectations. Furthermore, the leisure costs of work and DI claiming also rise substantially to 881 and 1,845 hours per year respectively. The increased importance of consumption reduces saving and increases labor supply. The higher leisure cost of work offsets this labor

supply effect. In particular, such high cost of working at all pushes agents into part time employment. The net effect of these changes in expectations and parameters is a slightly improved model fit: the criterion value drops from 522 for the specification with expectations modelled by life tables to 496 for average subjective expectations without heterogeneity.

The final column of Table 4 introduces variation in mortality expectations, allowing each individual to have an unique view of his own longevity. Heterogeneity in expectations changes the estimates for most parameters substantially relative to estimation uncertainty (the only exception is the time cost of poor health). The most economically important shifts relative to the model based on average subjective expectations are observed for the level of risk aversion, the consumption weight, the discount factor and for the parameters associated with bequests. Agents modeled by the third set of estimates are more risk averse, care less about consumption and are less patient than suggested by the other results. While these preferences are different in economically meaningful ways, all estimates remain plausible. Risk aversion increases to 5.14 which remains consistent with existing evidence. The consumption weight drops below fifty percent, but at 0.43 agents do desire a balanced mix of leisure and income. Though models (1) and (2) estimate discount factors that are in line with earlier papers, the third model yields a more plausible point estimate of 0.96. The finding that subjective survival expectations increase risk aversion and decrease the discount factor relative to actuarial tables corroborates the results in Gan et al. (2015) in a richer life cycle model.

The bequest motive in model (3) is different from the other two models in two ways. Firstly, the weight attached to bequests is much reduced, declining from 0.004 for a man living with his spouse to 0.0006 for a widower. Secondly, the K parameter that governs the curvature of the bequest utility function drops from around 900,000 for models (1) and (2) to 302,260 for the model with variation in beliefs. The implication is that while bequests are less important overall, differences in the strength of the bequest motive between individuals with higher and lower wealth levels are smaller than for the first two models. These changes in

preference parameters in combination with heterogenous expectations further improve model fit: the function value drops by ten percent from 496 to 450. The next section shows exactly how the fit differs between the three models.

5.2 Model fit

While function values summarize fit in one number, more can be learned by looking at the separate moments targeted in estimation. Figure 4 visualizes model fit by plotting data moments against their simulated counterparts based on all three specifications of expectations. Each column corresponds to a method used to approximate expectations and different rows show different groups of moments related to labor supply, benefit claiming or wealth. Looking first at the baseline model based on life tables, panels a. and d. indicate that it succeeds to match observed labor supply reasonably well, especially for individuals in poor health. Labor supply in good health is overestimated at ages 50-59: predicted yearly hours worked is up to 300 hrs/yr too high at age bins 56-57 and 58-59. Some of this difference may be due to the assumption that there is no early retirement prior to age 59 while in reality some industries and employers did allow workers to retire earlier. The model stays closer to the data for those currently in poor health: it never deviates by more than 100 hours. Participation in the labor force too is predicted more accurately for the unhealthy. Participation of those in good health is overestimated by around 10pp or 12% prior to age 60. Participation in bad health, on the other hand, is matched accurately for all ages. When it comes to benefit claiming, the model fits claiming rates of occupational pensions and unemployment insurance closely, with the exception that it overestimates pension claiming at age 64-65 by 12pp or 15%. While we match DI claiming well up to age bin 58-59, the early retirement age for occupational pensions, the model cannot rationalize the persistent claiming of disability insurance from age 58 onward. Though this discrepancy could be fixed by making DI stigma costs a function of age, we choose not to make such ad hoc changes to preferences. All in all,

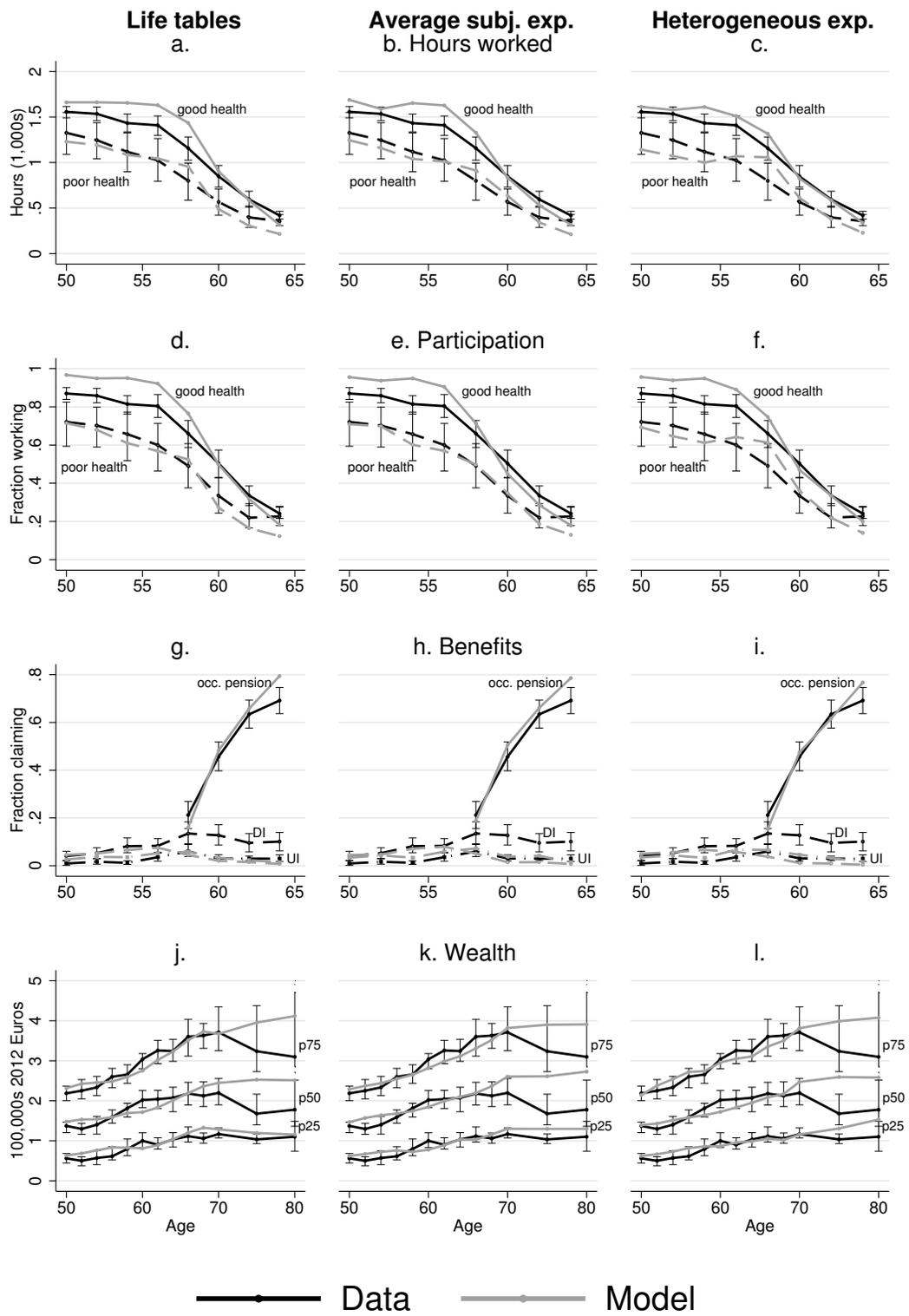


Figure 4: Model fit – target moments and simulations

the fit of the baseline model is adequate for most moments that capture labor supply and benefit claiming. The fit for wealth, on the other hand, varies across the distribution. The first quartile of the simulated wealth distribution tracks the data closely at all ages. However, the median and third quartile fare worse, especially after age 70. As simulated wealth levels off (p50) or continues to increase (p75) at older ages, the data moments decline. While the model substantially over-predicts wealth at advanced age, neither the median nor the 75th percentile are estimated precisely at those ages and model simulations are within the 95% confidence intervals at age 80-84.

The objective function values show that changing the level of survival probabilities to the average found in subjective data leads to an improvement in model fit. The source of that improvement does not stand out in Figure 4 – it is driven by small improvements in moments that are estimated relatively precisely. The primary place where the model based on average subjective beliefs does better is the labor supply of those in good health. For instance, the deviation from the data in hours worked at age 52-53 is 50 hrs/yr, compared to 100 hrs/yr for model (1). At age 58-59 too both the participation rate and the average hours worked are closer to the data than was the case for the model using life tables. Such improvements are picked up by the objective function because labor supply in good health is estimated relatively precisely, as indicated by the narrow confidence intervals. Panels g. and h. show that both models without variation in expectations fit the data roughly equally well when it comes to benefit claiming. Differences between models in terms of wealth are small: changing the level of mortality expectations does not ameliorate the overestimation of median wealth at old age.

Heterogeneity in subjective expectations further improves overall model fit. This is the net result of a better fit for some moments that receive much weight and worse fit for others that are estimated less precisely. Panel c. in the third column of Figure 4 shows that simulated hours worked in good health prior to age 60 are visibly closer to the data than for the other

two models. While simulations stay within the 95% confidence intervals, the fit for labor supply in poor health deteriorates slightly. Panel f. shows that the participation rate in good health is mostly similar to the other models, though it matches the data slightly better at age 64-65. Again, participation of those in poor health is not matched as well as was the case for the other models. Panel i. shows that there are no large changes in fit for pension claiming, or for disability and unemployment insurance. Furthermore, simulated wealth is close to that of the other two models.

Overall we find that all three models fit the data adequately and similarly. The variation between the objective function values shown in Table 4 reflects modest differences in fit for moments that are estimated relatively precisely. These moments are mostly related to labor supply in good health, which the baseline model overestimates substantially for ages 50-59. The better fit for the model based on heterogeneous expectations is due to improvements in hours worked in good health, which that model overestimates by around 100 hrs/yr compared with 200-300 hrs/yr for the baseline model. All three models yield similar profiles for occupational pension claiming and rates of disability and unemployment insurance. While these rates are close to the data for pension and unemployment benefits, the models cannot reproduce sustained disability insurance claiming after early retirement becomes available. Wealth quartiles are matched well up to age 70-74, but the median and third quartile are overestimated at ages 75-84. However, wealth moments are estimated imprecisely at those age brackets and simulations are in the confidence intervals at age 80-84. While there are statistically significant and economically meaningful differences between estimated preferences, the fit of all three models shows the same strengths and weaknesses.

5.3 Out-of-sample predictions

We use the models to predict changes in labor supply and pension claiming when moving to a different policy regime in order to test whether expectations affect counterfactual predictions.

The most important change between the estimation period (1993-2001) and the simulation period (2006-2016) was the introduction of actuarial adjustments to occupational pensions. As explained above, early claiming of occupational pensions was made less financially attractive to encourage workers to delay retirement. Hence, the key outcome targeted by the policy reform was the age at which people first claim their pension, their retirement age. The top panel of Table 5 reports the average retirement ages in both policy regimes, as well as the difference between them. The data indicate that the average age at which pension benefits were first claimed increased from 61.2 to 63.8 when early retirement was made costlier. This is a simple before-after comparison, in which the influence of cohort effects is mitigated by setting the pre-reform cohort to 1940-1949 and the post-reform cohort to 1950-1959 using the method described in section 4. Nonetheless, the increase of 2.6 years in the average retirement age is consistent in sign and magnitude with prior quasi-experimental estimates of the effect of the abolishment of generous early retirement (e.g. Euwals et al., 2010). The recession following the 2008 financial crisis does not affect our estimate for the simulation period substantially once we account for cohort differences between individuals born in the 1950s and 1960s.¹⁰ Moreover, the average retirement age of 61.2 during the 1990s is identical to that reported by the OECD in Duval (2003).

The bottom panel of Table 5 shows that neither the model based on life tables nor that which fixes expectations at the level of average subjective beliefs successfully anticipates this observed change in behavior. Both models predict an average retirement age in the estimation sample of 60.7, which is outside of the 95% confidence interval for the population mean and is six months younger than the sample average of 61.2. Furthermore, both models substantially over-predict the average retirement age for the period without early retirement scheme. According to the models without variation in survival expectations, pensions are

¹⁰Adding an indicator for the period after 2009, during which unemployment increased in the Netherlands, reduces the point estimate of the average retirement age during the post-reform period to 63.4 (95% confidence interval: 63.15-63.69). Hence the difference between pre and post-reform periods becomes 2.2 years (confidence interval: 1.89-2.57).

Table 5: Average retirement age with ('93-'01) and without ('06-'16) generous early retirement schemes

a. Data			
	Estimation sample ('93-'01)	Simulation sample ('06-'16)	Difference (yrs)
Average pension age ^a	61.2 (60.97 – 61.41)	63.8 (63.49 – 64.03)	2.6 (2.23 – 2.92)
N	756	467	1223
b. Simulations			
	Estimation sample ('93-'01)	Simulation sample ('06-'16)	Difference (yrs)
Life tables	60.7	65.6	4.9
Average subj. exp.	60.7	65.7	5.0
Heterogeneous exp.	61.0	63.5	2.5

^a Age of retirement combines information from three sources: 1) actual retirement age reported by retirees; 2) observed retirement age if respondent retires while in the sample; 3) expected retirement age if neither reported nor observed retirement age is available.
95% confidence intervals based on robust standard errors in parentheses.

first claimed at age 65.6-65.7 on average after the introduction of actuarial adjustments. This difference of around 5 years relative to the estimation period is almost twice the 2.6 years observed in the data. Allowing for variation in subjective mortality improves the in-sample predicted retirement age slightly: the average retirement age of 61 is in the 95% confidence interval for the population mean and is only 2.4 months below the sample average. Furthermore, that model produces much more reasonable simulations out-of-sample. It predicts that the average retirement age would increase to 63.5, which is in line with the data. The resulting difference between estimation and simulation samples is 2.5 years, which is close to the observed difference of 2.6 years. Based on these simulations we conclude that while the improvement in within-sample fit is modest, variation in expectations generates substantially better out-of-sample forecasts.

Figure 5 provides more detailed information regarding observed and simulated labor supply during the 1990s and 2000s. The figure contains results regarding the average hours worked by two-year age bins (pooling individuals in good and poor health). Columns cor-

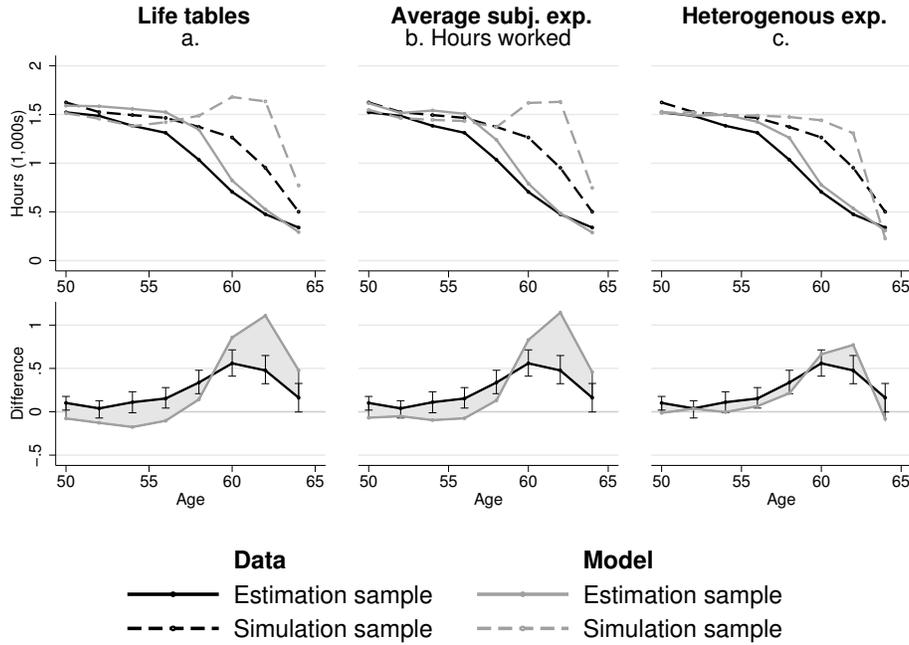


Figure 5: Simulated and observed labor supply in the estimation ('93-'01) and simulation ('06-'16) sample

respond to different models of survival expectations and the bottom graph in each panel highlights the difference between pre- and post-reform periods. The solid lines in the top graphs illustrate the discussion of model fit within the estimation sample: all models over-estimate labor supply prior to age 60 and this discrepancy is larger for the models based on life tables and average subjective expectations. The dashed lines correspond to the simulation sample and indicate that the models with fixed expectations, panels a. and b., substantially over-estimate labor supply out of sample for age bins 60-61 and 62-63. While the data show that at those ages labor supply is 500 hrs/year higher for the simulation sample than before, these models produce differences of 800-1200 hrs/year. At ages 64-65 too the simulated difference between estimation and simulation periods is about twice as large as that observed in the data. This substantial over-estimation of the change in labor supply at ages 60-65 confirms the results for the average retirement age documented in Table 5. Whereas the results

for older ages are most striking, at younger ages too the models with fixed expectations miss the mark. Both models predict that labor supply at age 50-57 is slightly lower during the 2000s than in the 1990s, which is not borne out by the data.

Panel c. of Figure 5 shows that out-of-sample simulations for labor supply improve markedly once variation in life expectancy is taken into account. While the model over-predicts labor supply at ages 60-63, the difference with the data is only about half as large as that observed in panels a. and b. The bottom graph illustrates that the differences between both sets of simulations are much closer to the differences observed in the data than is the case for the other two models. Moreover, simulations lie within 95% confidence intervals for all ages except 62-63 and 64-65. Whereas the results in Figure 5 combine adjustments to the extensive and intensive margins, similar patterns are obtained for participation in the labor market (see Appendix F).

5.4 How do heterogeneous expectations improve predictions?

In the previous section we show that the model with variation in survival expectations generates more realistic counterfactual predictions than do models with fixed beliefs. This section shows how the interaction between preferences and expectations leads to better forecasts. We first analyze behavior conditional on expectations in the estimation sample and then shift attention to the simulation sample.

Preferences, expectations and behavior in the estimation sample

Figure 6 documents behavior conditional on subjective survival. In particular, we compute moments by age bins separately for those who expect to live long (top 45% of subjective survival) or short (bottom 45%).¹¹ Black lines correspond to data moments and grey lines

¹¹The top 45% of subjective survival corresponds to the bottom 45% of posterior means for individual effects for the hazard of death and visa versa. We drop the middle 10% of the sample to make the classification more robust to measurement error in the reported probabilities used to model beliefs.

Estimation sample ('93-'01)

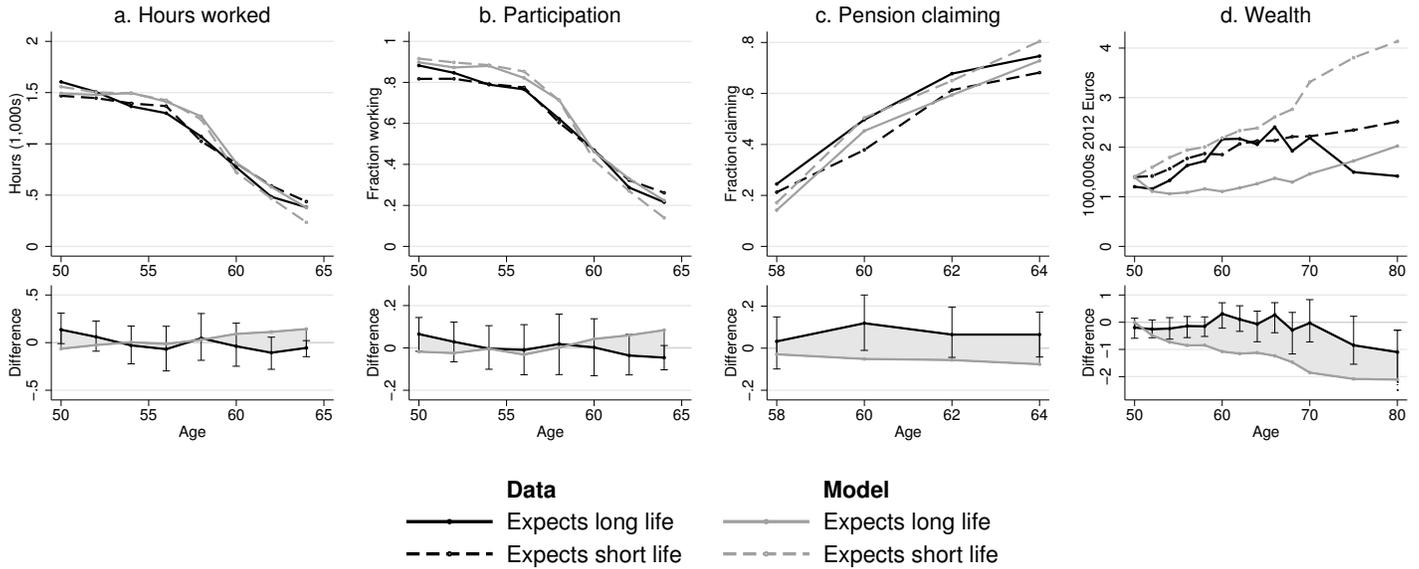


Figure 6: Behavior conditional on survival expectations for the estimation sample (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)

represent model simulations. The data do not reveal large or statistically significant differences along survival expectations for hours worked, participation rates or pension claiming. However, small sample sizes result in wide confidence intervals for the differences between those who expect to live long and those who do not, so economically meaningful variation cannot be ruled out. Though these conditional moments were not targeted in estimation, the model with individual-specific expectations fits the data in that it does not generate large differences in labor supply or pension claiming. For wealth, on the other hand, the model indicates that those who expect to live shorter hold more wealth late in life than their peers who expect a longer life. By age 80-84 the difference between the conditional medians amounts to 200,000 Euro. This divergence is driven by the fact that the bequest motive is directed primarily to other individuals living in the same household (one’s spouse). Those who expect a long life expect to outlive their partner, which reduces their desire to bequeath

and hence suppresses their accumulation of wealth. While uncertainty is considerable, the data do suggest that those who expect to live shorter hold on to more wealth after age 70. The point estimate for the difference at age 80-84 is 100,000 euro and the corresponding confidence interval does not include zero. However, up to age 70 there is little systematic covariation between wealth and expectations in the data, while in the simulations the difference quickly rises to 100,000 euro. Thus, the model fits the qualitative pattern observed for wealth at the oldest ages but not at younger ones.

As can be seen in Appendix G, the model produces larger differences along the distribution of survival expectations for all types of behavior if we evaluate it at the preference estimates obtained for expectations fixed at the subjective average. For labor supply and pension claiming this variation clearly deviates from the data. Furthermore, at the estimates for fixed expectations the model also generates an even larger gap for wealth: the median wealth holding in the top 45% of the distribution of subjective survival reaches zero by age 80-84. Comparing model simulations at these two sets of estimates illustrates how preferences and expectations interact. In particular, the fact that labor supply and pension claiming do not vary with survival in the data is itself valuable information regarding preferences.

Preferences, expectations and behavior in the simulation sample

The pattern of unrealistically strong variation of behavior with expectations when simulations are done at the preference estimates for average subjective survival is even starker in the simulation sample (2006-2016). Figure 7 displays differences in behavior between those who expect to live long and those who anticipate an earlier demise. The variation in simulated labor supply and pension claiming across the distribution of expectations is more pronounced for this sample, in which occupational pensions are subject to actuarial adjustments. Estimating preferences under heterogeneous expectations thus results in behavior that is closer to the data for the counterfactual policy regime, both overall and conditional on beliefs.

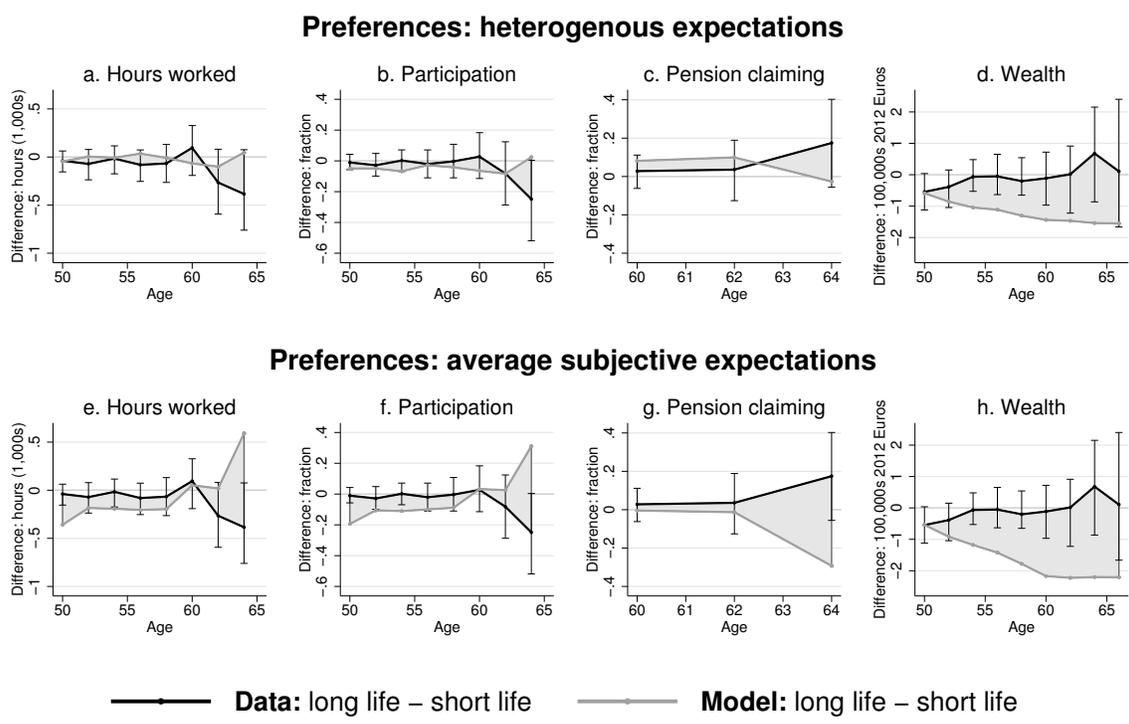


Figure 7: Differences in behavior across the distribution of survival expectations for the simulation sample (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)

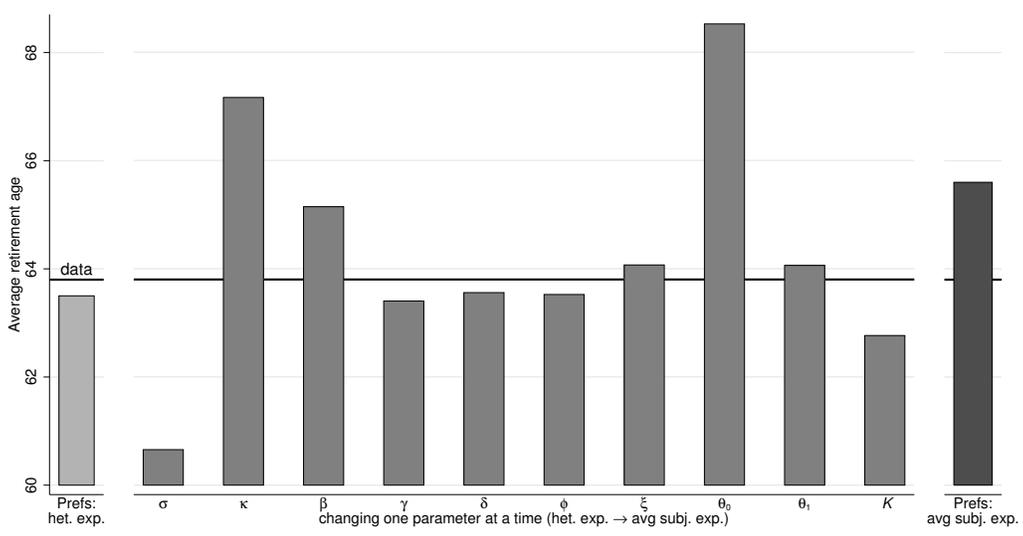


Figure 8: The impact of changing preference parameters on counterfactual simulations for the model with heterogeneous expectations

Figure 8 illustrates which preference parameters are the most important drivers of differences in counterfactual predictions across models. The figure presents the average retirement age in simulations and the data for the period 2006-2016. All simulations are based on the model with heterogeneous expectations, but preferences vary. The leftmost bar corresponds to preference estimates for the same model and reproduces the average simulated retirement age of 63.5 (compared with 63.8 in the data, see Table 5). The rightmost bar is also based on the model with heterogeneous beliefs, but preferences are set to the estimates for the model that equates beliefs to the subjective *average*. Doing so results in an average simulated retirement age of 65.6, which is equal to that based on the model with fixed expectations evaluated at the same parameters. Variation in expectations does not change average retirement for a given set of preferences. The middle panel of Figure 8 illustrates the relative importance of the different parameters. It starts from the estimates obtained under belief heterogeneity and changes one parameter at a time to the estimate from the model without variation in expectations. Risk aversion (σ), the consumption weight (κ), the discount factor (β) and the constant in the bequest weight (θ_0) all have large effects on the average retirement age. The estimates obtained for these parameters cause the model with heterogeneity to outperform the ones without it.

6 Conclusion

This paper shows that incorporation of subjective survival expectations into a life cycle model of labor supply, pension claiming and saving leads to improved out-of-sample forecasts. We model the retirement incentives faced by Dutch workers during the 1990s, a period during which occupational pensions could be claimed as early as age 59 without actuarial adjustments to benefit levels. Preference parameters are estimated under three specifications for life expectancy: life tables adjusted for current health, average subjective expectations and

heterogenous subjective expectations that use reported survival probabilities to construct subjective longevity for each individual in the sample. We use the estimated preferences to simulate behavior in a policy regime that does adjust pension benefits for the age at which they are first claimed – the situation in place during the 2000s. Using data from the DNB Household Survey that spans the period 1993-2016 we evaluate how survival expectations affect estimated preferences and model fit both within the estimation sample and in a different policy environment.

The findings show that expectations matter for the estimation of preferences: different models of expectations yield preferences estimates that are both statistically and economically different. In particular, heterogeneity in expectations results in higher risk aversion, a lower weight on consumption relative to leisure and a lower and more plausible estimate for the rate of time preference compared to the two models without variation in expectations. This extends the results in Gan et al. (2015) and underlines the fact that structural models identify preferences *conditional on expectations*, a point made by Manski (2004).

Given that beliefs matter even if one is primarily interested in preferences, the question arises whether for these life cycle models different combinations of expectations and preferences are equivalent in terms of implied behavior. The values of the objective function at the estimates show that average subjective expectations outperform life tables in terms of model fit and that heterogeneity leads to further improvement. However, a closer analysis of the behavior implied by the models shows that all three models share the same strengths and weaknesses as far as fit is concerned. They produce a reasonably close fit of observed labor supply and occupational pension claiming, though they over-predict labor supply in good health for ages 50-59. Moreover, they reproduce observed rates of disability and unemployment insurance claiming prior to the eligibility age for occupational pensions. They do not, however, match the sustained claiming of disability benefits once pensions become available. Moreover, while the models follow observed quartiles of wealth closely up to age bin 70-74,

the median and 75th percentile deviate from the data at older ages where those quartiles are not estimated precisely. Overall the within-sample fit of all three models is similar.

We do find large differences in the accuracy of out-of-sample forecasts under alternative pension rules. The models based on life tables and average subjective expectations predict an average retirement age above 65 after the introduction of actuarial adjustments, an increase of 5 years relative to the estimation period. This is far higher than that observed in the DHS or reported in quasi-experimental studies: the before-after difference in our sample is 2.6 years to an average post-reform retirement age of 63.8 years. The model with variation in survival expectations does much better, predicting a 2.5 year increase to an average of 63.5 (both are within the corresponding 95% confidence interval). The superiority of the model with heterogeneous beliefs in out-of-sample predictions is confirmed in more detailed analyses of labor supply by age (both for hours worked and for the participation rate). These gains are not explained by covariation between labor supply or pension claiming and subjective longevity: neither in the data nor in simulations do we find systematic differences in behavior between those with different life expectancies. Instead the different preference estimates interact with expectations to produce more accurate forecasts. Risk aversion, the consumption weight, the discount factor and the importance of bequests all play substantial roles. Changing these estimates affects labor supply, benefit claiming and wealth accumulation, both overall and conditional on subjective longevity.

We interpret the findings above as evidence in support of the use of subjective survival in structural economic models. One message is that even though different combinations of expectations and preferences may be more or less equivalent within a given framework of incentives, they may still have very different implications in an alternative policy regime. Incorporating individual-specific expectations entails a modest cost in terms of complexity and require addition of a single state variable to the model.

References

- Bissonnette, L. and de Bresser, J. (2018). Eliciting subjective survival curves: Lessons from partial identification. *Journal of Business & Economic Statistics*, 36(3):505–515.
- Blundell, R., French, E., and Tetlow, G. (2016). Retirement incentives and labor supply. In *Handbook of the economics of population aging*, volume 1, pages 457–566. Elsevier.
- Bovenberg, A. L. and Meijdam, L. (2001). The dutch pension system. In *Pension Reform in Six Countries*, pages 39–67. Springer.
- CBS (2017). Statistics Netherlands (CBS) Statline. <http://statline.cbs.nl>. Accessed: 2017-01-29.
- De Bresser, J. (2017). Test-retest reliability of subjective survival expectations. Mimeo Tilburg University.
- De Bresser, J., Fonseca, R., and Michaud, P.-C. (2017). Retirement behavior in the U.S. and Europe. Netspar Academic Series, DP 11/2017-018.
- De Bresser, J. and Van Soest, A. (2013). Survey response in probabilistic questions and its impact on inference. *Journal of Economic Behavior and Organization*, 96(December 2013):65–84.
- De Nardi, M., Fella, G., and Pardo, G. P. (2018). Nonlinear household earnings dynamics, self-insurance, and welfare. Technical report, National Bureau of Economic Research. forthcoming in *Journal of the European Economic Association*.
- De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenditures. *Journal of Political Economy*, 118(1):39–75.
- De Nardi, M., French, E., and Jones, J. B. (2016). Savings after retirement: A survey. *Annual Review of Economics*, 8:177–204.

- Delavande, A. (2008). Pill, patch, or shot? subjective expectations and birth control choice. *International Economic Review*, 49(3):999–1042.
- Duval, R. (2003). The retirement effects of old-age pension and early retirement schemes in OECD countries. EOCED Economics Department Working Papers, No. 370.
- Eckstein, Z., Keane, M., and Lifshitz, O. (2019). Career and family decisions: Cohorts born 1935–1975. *Econometrica*, 87(1):217–253.
- Euwals, R., Van Vuuren, D., and Wolthoff, R. (2010). Early retirement behaviour in the netherlands: Evidence from a policy reform. *De Economist*, 158(3):209–236.
- French, E. (2005). The effects of health, wealth and wages on labour supply and retirement behaviour. *Review of Economic Studies*, 72(2):395–427.
- French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.
- Gan, L., Gong, G., Hurd, M., and McFadden, D. (2015). Subjective mortality risk and bequests. *Journal of Econometrics*, 188(2):514–525.
- Gan, L., Hurd, M. D., and McFadden, D. L. (2005). Individual subjective survival curves. In *Analyses in the Economics of Aging*, pages 377–412. University of Chicago Press.
- Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of econometrics*, 60(1-2):65–99.
- Gourinchas, P.-O. and Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, 70(1):47–89.
- Grip, A. d., Lindeboom, M., and Montizaan, R. (2011). Shattered dreams: the effects of changing the pension system late in the game. *The Economic Journal*, 122(559):1–25.

- Groneck, M., Ludwig, A., and Zimper, A. (2016). A life-cycle model with ambiguous survival beliefs. *Journal of Economic Theory*, 162:137–180.
- Gustman, A. L., Steinmeier, T. L., et al. (1986). A structural retirement model. *Econometrica*, 54(3):555–84.
- Hamermesh, D. S. (1985). Expectations, life expectancy, and economic behavior. *The Quarterly Journal of Economics*, 100(2):389–408.
- Heimer, R., Myrseth, K. O. R., and Schoenle, R. (2018). Yolo: Mortality beliefs and household finance puzzles. Forthcoming in *Journal of Finance*.
- Heyma, A. (2004). A structural dynamic analysis of retirement behaviour in the netherlands. *Journal of Applied Econometrics*, 19(6):739–759.
- Hubbard, R. G., Skinner, J., and Zeldes, S. P. (1994). The importance of precautionary motives in explaining individual and aggregate saving. In *Carnegie-Rochester Conference Series on Public Policy*, volume 40, pages 59–125. Elsevier.
- Hubbard, R. G., Skinner, J., and Zeldes, S. P. (1995). Precautionary saving and social insurance. *Journal of political Economy*, 103(2):360–399.
- Hurd, M. (2009). Subjective probabilities in household surveys. *Annual Economic Review*, 1:543–562.
- Hurd, M. D. (1989). Mortality risk and bequests. *Econometrica: Journal of the econometric society*, pages 779–813.
- Hurd, M. D., McFadden, D. L., and Gan, L. (1998). Subjective survival curves and life cycle behavior. In *Inquiries in the Economics of Aging*, pages 259–309. University of Chicago Press.

- Hurd, M. D. and McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human resources*, pages S268–S292.
- Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, 112(482):966–985.
- Kleinjans, C. and Van Soest, A. (2014). Non-response and focal point answers to subjective probability questions. *Journal of Applied Econometrics*, 29(4):567–585.
- Kopczuk, W. and Lupton, J. P. (2007). To leave or not to leave: The distribution of bequest motives. *The Review of Economic Studies*, 74(1):207–235.
- Low, H., Meghir, C., and Pistaferri, L. (2010). Wage risk and employment risk over the life cycle. *American Economic Review*, 100(4):1432–67.
- Low, H., Meghir, C., Pistaferri, L., and Voena, A. (2018). Marriage, labor supply and the dynamics of the social safety net. Technical report, National Bureau of Economic Research. Working Paper 24356.
- Manski, C. F. (2004). Measuring expectations. *Econometrica*, 75(5):1329–1376.
- Manski, C. F. and Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics*, 28(2):219–231.
- Montizaan, R., Cörvers, F., and De Grip, A. (2010). The effects of pension rights and retirement age on training participation: Evidence from a natural experiment. *Labour Economics*, 17(1):240–247.
- Montizaan, R., de Grip, A., Cörvers, F., and Dohmen, T. (2015). The impact of negatively reciprocal inclinations on worker behavior: Evidence from a retrenchment of pension rights. *Management Science*, 62(3):668–681.

- Montizaan, R. M. and Vendrik, M. C. (2014). Misery loves company: Exogenous shocks in retirement expectations and social comparison effects on subjective well-being. *Journal of Economic Behavior & Organization*, 97:1–26.
- Rust, J. and Phelan, C. (1997). How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica: Journal of the Econometric Society*, pages 781–831.
- Stinebrickner, R. and Stinebrickner, T. R. (2013). A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472.
- Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181.
- Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge.
- Van der Klaauw, W. (2012). On the use of expectations data in estimating structural dynamic choice models. *Journal of Labor Economics*, 30(3):521–554.
- Van der Klaauw, W. and Wolpin, K. I. (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1-2):21–42.
- Wu, S., Stevens, R., and Thorp, S. (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase. *Journal of Economic Dynamics and Control*, 55(June 2015):39–56.
- Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348.

A Measurement model for subjective expectations

The measurement model for subjective expectations uses multiple reported probabilities for each respondent/wave observation to construct subjective survival curves. It takes into account the rounding enforced by the DHS survey, which limits response to an eleven-point scale ranging from 0 (no chance) to 10 (absolute certainty). The model is similar to those presented in De Bresser and Van Soest (2013); Kleinjans and Van Soest (2014) and especially De Bresser (2017), on which this discussion is based.

Data

Subjective survival expectations are elicited by means of the following questions:

“Please indicate your answer on a scale of 0 thru 10, where 0 means ‘no chance at all’ and 10 means ‘absolutely certain’.

How likely is it that you will attain (at least) the age of [65]?”

etc.

Respondents answer questions that refer to multiple target ages, 65 in the example, in each survey. These target ages depend on the current age of the respondent: older respondents face older target ages. We interpret the answers as probabilities, with 1 corresponding to 10% chance etc. (De Bresser, 2017, shows that this is a valid interpretation).

Table A1 summarizes the data. The average reported probability declines monotonically with target age from 79% for age 65 to 18-26% for age 100. Moreover, there is substantial heterogeneity at all target ages, with standard deviations around 20-30pp and Inter-Quartile Ranges (IQRs) of 30pp. We model these reported probabilities in order to adjust for rounding and to reduce heterogeneity in expectations to a single parameter.

Table A1: Descriptive statistics of reported survival probabilities (0-100%)

	N	Current age	Mean	Std. Dev.	p25	p50	p75
a. Estimation sample (1993-2001)							
Target age 75	1,629	50-65	69	20	50	70	80
Target age 80	2,008	50-70	56	23	40	60	70
Target age 85	646	65-75	48	22	30	50	60
Target age 90	381	70-80	35	24	20	30	50
Target age 95	151	75-85	27	23	0	20	50
Target age 100	43	80-88	26	30	0	10	50
b. Simulation sample (2006-2016)							
Target age 65	879	50-65	79	16	70	80	90
Target age 75	3,706	50-65	70	19	60	70	80
Target age 80	5,169	50-69	59	23	50	60	80
Target age 85	2,505	65-74	55	23	40	60	70
Target age 90	1,792	70-79	41	25	20	40	60
Target age 95	1,185	75-84	30	24	10	30	50
Target age 100	553	80-89	18	21	0	10	30

Model

In the model expectations follow a Gompertz distribution with the baseline hazard shifted proportionally by demographic variables and unobserved heterogeneity. This parameterization of expectations implies that the true subjective probability of surviving to target age ta_k conditional on having survived to current age a_{it} is given by:

$$\begin{aligned}
 S_{itk}|a_{it} &= \Pr(t \geq ta_k | t \geq a_{it}) = \frac{\Pr(t \geq ta_k \ \& \ t \geq a_{it})}{\Pr(t \geq a_{it})} \\
 &= \frac{1 \times \Pr(t \geq ta_k)}{\Pr(t \geq a_{it})} = \frac{\exp\left(-\frac{\gamma_{it}}{\alpha} (\exp(\alpha (ta_k/100)) - 1)\right)}{\exp\left(-\frac{\gamma_{it}}{\alpha} (\exp(\alpha (a_{it}/100)) - 1)\right)}
 \end{aligned} \tag{14}$$

where $\gamma_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta}_1 + \zeta_i + \eta_{it})$ depends on the demographics of respondent i in survey-year t and α determines the shape of the baseline hazard. The most important variable that affects survival expectations is current health. We distinguish two types of unobserved heterogeneity: individual effects ζ_i and question sequence effects η_{it} . The former are common to all probabilities reported by individual i , while the latter are shared only by probabilities

reported in survey wave t . Distributional assumptions for these error components follow. We divide both the target age and the current age by 100 to facilitate estimation of baseline hazard α . Equation 14 shows that the proportional hazard parameterization of expectations reduces differences between individuals to the single parameter γ_{it} .

We do not observe S_{itk} directly. Instead, the reported probabilities are perturbed by recall error ε_{itk} :

$$P_{itk}^* = S_{itk} + \varepsilon_{itk} \quad (15)$$

where $\varepsilon_{itk} \sim \mathcal{N}(0, \sigma_{it}^2)$, independent of all covariates and across thresholds, surveys, years and individuals. Note that positive correlation between recall errors in a survey wave implies that all errors are likely to be either high or low and would be indistinguishable from heterogeneity in expectations. We allow for heteroskedasticity and model the variance of recall errors as $\ln(\sigma_{it}) = \mathbf{x}'_{it} \boldsymbol{\beta}_2$.

Latent probabilities P_{itk}^* are censored between zero and 100 and rounded prior to being reported. We allow for rounding to multiples of 100, 50 and 10. Our rounding model is ordinal:

$$R_{itk} = r \iff \mu_{r-1} \leq y_{it}^* = \zeta_i^r + \eta_{it}^r + \varepsilon_{itk}^r < \mu_r \quad (16)$$

where $r \in \{10, 50, 100\}$. The rounding equation includes individual and question sequence effects, allowing rounding to be correlated across repeated observations for a given individual and to be more strongly correlated within survey waves than between them. Moreover, both types of unobserved heterogeneity may be correlated with their respective counterparts in the equation that shifts survival curves. We assume that the idiosyncratic rounding shocks ε_{itk}^r follow a standard normal distribution and are independent from covariates and all other errors, so the conditional probabilities of each category of rounding $\Pr(R_{itk} = r | \zeta_i^r, \eta_{it}^r)$ take

the shape of an ordered probit.

A reported probability in combination with a particular level of rounding implies an interval for the perturbed probability $P_{itk}^* \in [LB_{itk}^r, UB_{itk}^r]$. For instance, a reported probability of 50% that is rounded to a multiple of 10 yields the interval $P_{itk}^* \in [45, 55]$. The probability of that event is easy to calculate, since $P_{itk}^* \sim \mathcal{N}(S_{itk}, \sigma_{it}^2)$. Rounding is a latent construct, because a given reported probability may result from different degrees of rounding. We therefore average across the different degrees of rounding to obtain the likelihood contribution. In particular, define for each reported probability P_{itk} the set Ω_{itk} that consists of all types of rounding that are consistent with that probability. We obtain the conditional density as:

$$f(P_{itk}|\mathbf{x}_{it}, \boldsymbol{\zeta}_i, \boldsymbol{\eta}_{it}) = \sum_{r \in \Omega_{itk}} \Pr(R_{itk} = r | \zeta_i^r, \eta_{it}^r) \times \Pr(LB_{itk}^r \leq P_{itk}^* < UB_{itk}^r | \mathbf{x}_{it}, \zeta_i, \eta_{it}) \quad (17)$$

where $\Pr(LB_{itk}^r \leq P_{itk}^* < UB_{itk}^r | \mathbf{x}_{it}, \zeta_i, \eta_{it})$ is given by

$$\Pr(LB_{itk}^r \leq P_{itk}^* < UB_{itk}^r | \cdot) = \begin{cases} \Pr(LB_{itk}^r \leq P_{itk}^* | \mathbf{x}_{it}, \zeta_i, \eta_{it}); & \text{if } P_{itk} \geq P_{it,k-1} - 0.5r \\ LB_{itk}^r = P_{it,k-1} - 0.5r & \\ \Pr(LB_{itk}^r \leq P_{itk}^* < UB_{itk}^r | \mathbf{x}_{it}, \zeta_i, \eta_{it}); & \text{if } 0.5r \leq P_{itk} \\ LB_{itk}^r = P_{itk} - 0.5r & \text{and } P_{itk} < P_{it,k-1} - 0.5r \\ UB_{itk}^r = P_{itk} + 0.5r & \\ \Pr(P_{itk}^* < UB_{itk}^r | \mathbf{x}_{it}, \zeta_i, \eta_{it}); & \text{if } P_{itk} < 0.5r \\ UB_{itk}^r = 0.5r & \end{cases} \quad (18)$$

where $r \in \{10, 50, 100\}$ as before. The first case occurs when P_{itk} is censored from above by the preceding probability (the first probability is censored by 100). The second case is not subject to censoring and the third case is censoring from below at zero. All probabilities in

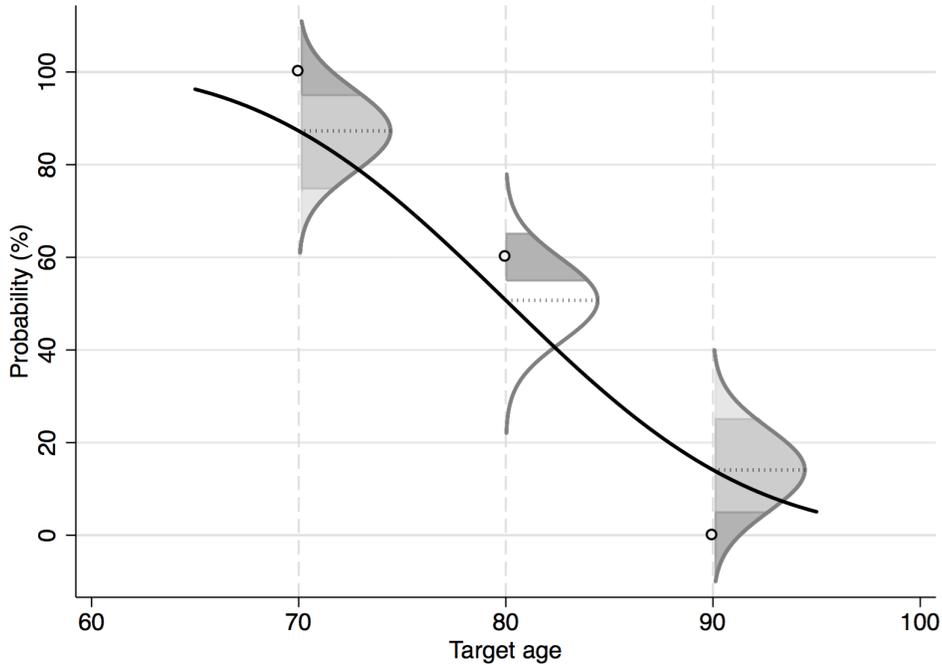


Figure A1: Illustration of the logic of the rounding model

equation 18 above are calculated from univariate normal distributions and are therefore easy to obtain. Whether a given probability is censored or not depends on the degree of rounding and on the preceding reported probability.

Individual effects of expectations and rounding follow a bivariate normal distribution, as do sequence effects. The two types of heterogeneity are independent of one another and of covariates (both types of error components are modelled as random effects). We estimate the measurement model by Maximum Simulated Likelihood and simulate the likelihood using 100 Halton draws (see Train, 2003, for details).

Figure A1 visualizes the logic of the model. The empty circles represent three hypothetical reported probabilities, equal to 100, 60 and 0% respectively. The black curve is the survival function of the Gompertz distribution for the respondent's combination of observed and unobserved characteristics. For each target age k , this Gompertz survival curve gives S_{itk} in equations 14 and 15. Vertical normal distributions capture the recall error ε_{itk} from equa-

Table A2: Estimates for the model of subjective expectations

	a. Estimation sample (1993-2001)				b. Simulation sample (2006-2016)			
	Expectations – hazard rates		Heteroskedasticity recall error		Expectations – hazard rates		Heteroskedasticity recall error	
Poor health	1.363***	(0.0443)	0.00979	(0.0509)	1.244***	(0.0187)	0.155***	(0.0201)
Educ. middle ^a	1.005	(0.0423)	-0.161***	(0.0536)	0.927***	(0.0249)	-0.133***	(0.0215)
Educ high ^a	1.058	(0.0423)	-0.162***	(0.0456)	1.104***	(0.0320)	-0.0993***	(0.0192)
Constant	0.00875***	(6.56e-4)	2.355***	(0.0391)	0.0185***	(4.01e-4)	2.485***	(0.0161)
Var [ζ_i] (ind. effects)	0.406***	(0.0181)			0.503***	(0.0687)		
Var [η_{it}] (seq. effects)	0.0626***	(0.00476)			0.0203***	(0.00159)		
Var [ζ_i]/Var [ζ_i] + Var [η_{it}]	0.866***	(0.0103)			0.961***	(0.00606)		
Baseline hazard	7.788***	(0.100)			6.693***	(0.0280)		
	Rounding				Rounding			
μ_1	1.461***	(0.0895)			2.478***	(0.110)		
μ_2	3.326***	(0.179)			3.996***	(0.141)		
Var [ζ_i^r] (ind. effects)	0.821***	(0.184)			2.729***	(0.320)		
Var [η_{it}^r] (seq. effects)	0.114	(0.0917)			0.00863	(0.00962)		
Var [ζ_i^r]/Var [ζ_i^r] + Var [η_{it}^r]	0.878***	(0.0883)			0.997***	(0.00348)		
Corr [ζ_i ; ζ_i^r]	0.0909	(0.0740)			0.149***	(0.0438)		
Corr [η_{it} ; η_{it}^r]	0.264	(0.231)			-0.397	(0.488)		
Number of individuals	1,371				1,557			
Number of probabilities	4,858				15,789			
Log-likelihood	-8,393.77				-25,947.56			

^a Baseline: low education level (lower secondary education). Educ. middle: intermediate vocational. Educ. high: (applied) university. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

tion 15. In addition to this perturbation, reported probabilities are censored and rounded. The first reported probability is 100, which is a multiple of 10, 50 and 100. Hence, it can result from any of those levels of rounding. If it were rounded to 10s, the latent probability P_{itk}^* must be larger than 95 for the reported probability to be 100. This corresponds to the darkest shaded area under the normal distribution. However, if it were rounded to 50s or 100s, the lower bound for the latent probability would be 75 and 50 respectively. Those degrees of rounding result in the lighter grey areas under the normal. A reported probability of 60, on the other hand, can only result from rounding to multiples of 10. Hence, there is only one shade of grey for that probability, ranging from the lower bound of 55 to the upper bound of 65.

Estimates

Model estimates are reported in Table A2. People in bad health have a higher hazard of death than those in good health: their hazard is 36% higher in the estimation sample and 24% higher in the simulation sample. The estimates of the heteroskedasticity equation show that compared to their lower educated peers, respondents who finished at least intermediate vocational training report probabilities that are closer to Gompertz distributions. Unobserved heterogeneity is primarily due to individual effects, which account for 87-96% of its total variance. Unsurprisingly, the baseline hazard of death increases significantly with age. The estimated rounding thresholds imply that 93% of probabilities in the estimation sample and 99% in the simulation sample are rounded to multiples of 10.

Heterogeneity in expectations: posterior means

Having estimated the model, we calculate posterior means for individual and sequence effects for each respondent given personal characteristics and reported probabilities. The posterior mean for the individual effect of individual i and for the sequence effect pertaining to probability sequence it are obtained by the following algorithm (more information can be found in Train, 2003):

1. Draw 500 pairs of vectors $\boldsymbol{\zeta}_i^s = (\zeta_i^s, \zeta_i^{r,s})$ and $\boldsymbol{\eta}_{it}^s = (\eta_{it}^s, \eta_{it}^{r,s})$ from their bivariate normal distributions ($s = 1, \dots, 500$).
2. For each pair of vectors s calculate the conditional likelihood contribution $\mathcal{L}_i^s(\mathbf{P}_i | \boldsymbol{\zeta}_i^s, \boldsymbol{\eta}_i^s; \mathbf{x}_i, \mathbf{b})$ of individual i . This is the conditional probability of observing all probabilities reported by that individual, given the draw of unobserved heterogeneity, covariates and the parameter estimates. Similarly, compute $\mathcal{L}_{it}^s(\mathbf{P}_{it} | \boldsymbol{\zeta}_i^s, \boldsymbol{\eta}_{it}^s; \mathbf{x}_{it}, \mathbf{b})$: the likelihood contribution of probability sequence it . Note that the likelihood for individual i is the

product of the likelihoods for all his sequences: $\mathcal{L}_i^s = \prod_{t=1}^T \mathcal{L}_{it}^s$. This is true because we condition on unobserved as well as observed heterogeneity.

3. Calculate individual-level weights for each of the 500 draws as $w_i^s = \frac{\mathcal{L}_i^s}{\sum_{s=1}^{500} \mathcal{L}_i^s}$. For the sequence effects we calculate weights for each question sequence ($w_{it}^s = \frac{\mathcal{L}_{it}^s}{\sum_{s=1}^{500} \mathcal{L}_{it}^s}$).
4. Simulate the posterior mean individual effect as $\tilde{\boldsymbol{\zeta}}_i = (\tilde{\zeta}_i, \tilde{\zeta}_i^r) = \sum_{s=1}^{500} w_i^s \boldsymbol{\zeta}_i^s$. The mean of the posterior distribution of the sequence effect is $\tilde{\boldsymbol{\eta}}_{it} = (\tilde{\eta}_{it}, \tilde{\eta}_{it}^r) = \sum_{s=1}^{500} w_{it}^s \boldsymbol{\eta}_{it}^s$.

We simulate $\tilde{\boldsymbol{\zeta}}_i$ for every individual in the data and $\tilde{\boldsymbol{\eta}}_{it}$ for all sequences of probabilities and use these individual and sequence-level estimates to compute an approximation $\tilde{\gamma}_{it}$ of the γ_{it} parameter for each individual/year observation:

$$\tilde{\gamma}_{it} = \exp\left(\mathbf{x}'_{it} \mathbf{b}_1 + \tilde{\zeta}_i + \tilde{\eta}_{it}\right) \quad (19)$$

where $\tilde{\zeta}_i$ is the posterior mean for ζ_i and $\tilde{\eta}_{it}$ is the mean for η_{it} . We leave covariates \mathbf{x}_{it} at the levels observed in the data, except for health which we set to ‘good’ for everybody when calculating the expectations index $\tilde{\gamma}_{it}$. In this way our individual-specific measure of subjective survival controls for current health. The estimates reported in Table A2 show that individual effects are much more important than sequence effects, which means that expectations of a given individual are stable over time. Therefore, we fix expectations of the individual to

$$\tilde{\gamma}_i = \exp\left(\mathbf{x}'_{it} \mathbf{b}_1 + \tilde{\zeta}_i + \tilde{\eta}_i\right) \quad (20)$$

where we substitute the average sequence effect for individual i $\tilde{\eta}_i$. We condition on good health and the other covariate, education, is time-constant since respondents enter our sample at age 50. As a result \mathbf{x}_{it} is constant over time for each individual.

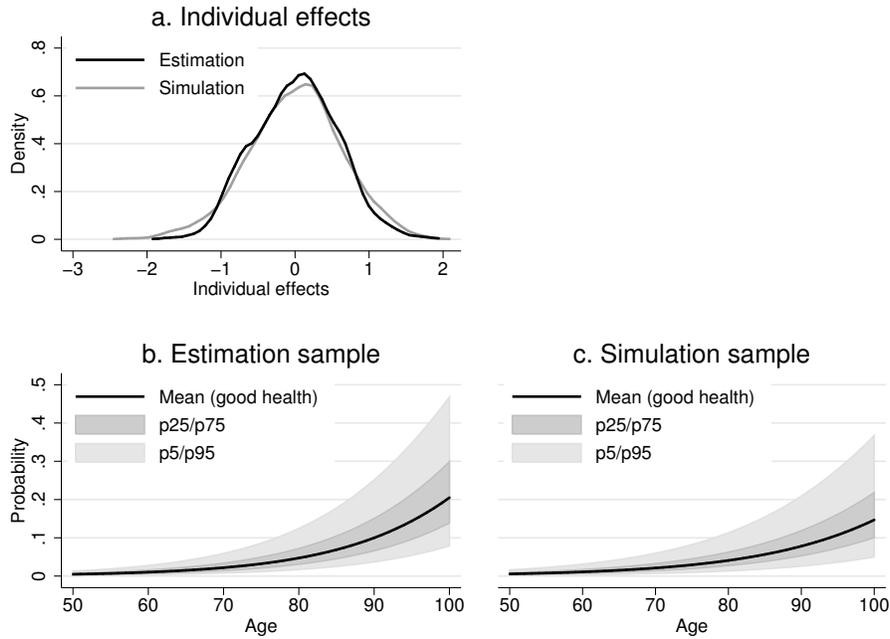


Figure A2: Heterogeneous expectations: distributions of individual effects and corresponding variation in survival probabilities

Figure A2 shows the distributions of the posterior means for the individual effects in both samples and the variation in survival expectations that these distributions generate. Panel a. contains the densities of individual effects, which are almost symmetric around zero. The distributions in both samples look similar, with slightly more dispersion in the simulation sample. Panels b. and c. plot the variation in mortality probabilities that is induced by the distributions from panel a. As discussed in the main text, the individual effects generate substantial differences in expectations between individuals. These differences are large relative to the impact of current health.

B Institutions

Earnings

Earnings are the product of wages j_t and hours worked h_t and wages are adjusted for part-time employment defined as working fewer than 2000 hours per year:

$$earn_t = \exp[-\pi \mathbb{I}\{h_t < 2000\}] j_t \times h_t$$

The penalty π for part time work is set to 20% in accordance with previous papers such as Gustman et al. (1986).

Disability and unemployment insurance

One cannot claim social insurance while working or after age 64. Only one single type of benefits can be claimed at any point in time. If an individual chooses to claim, both disability and unemployment insurance provide benefits that replace 70% of last earned gross earnings. Payouts from both schemes are capped at 50,400 Euro per year. While payments are calculated according to the same formula, UI and DI can be claimed under different circumstances. DI can only be claimed when in bad health ($m_t = 1$) and claiming can continue indefinitely while in poor health (up to age 65). UI provides benefits regardless of health status, but eligibility does depend on prior work experience. Individuals are entitled to one year of UI for each 12 years of work (either part time or full time). Moreover, UI cannot be claimed for more than three years, after which new entitlements can be accumulated at the rate of one month per year of work.

Pensions

Public and occupational pensions are the two main pillars that provide retirement income in the Netherlands. The public pension is a flat rate pay-as-you-go funded scheme that

provides a subsistence income to all individuals who lived in the country between ages 15 and 65. Payments start automatically at age 65 regardless of labor supply. Since the public pension provides all older individuals with a basic income, UI and DI cannot be claimed from age 65 onwards. The level of benefits is 9,000 Euro per year.

In addition to the public pension, around 90% of Dutch employees are covered by a Defined Benefit (DB) occupational pension through their employer (Bovenberg and Meijdam, 2001). Participation in such schemes is mandatory and workers cannot opt out or choose between multiple pension funds (each industry or company is typically covered by a single fund). Furthermore, contributors cannot set their contribution rate or influence the way in which their contributions are invested. Unlike the public pension, occupational pensions can only be claimed if the pensioner does not work. During the period of our estimation sample, 1993-2001, occupational pension payouts had two distinct phases. Early retirement benefits, known in Dutch as *Vervroegde Uittreding* or VUT, could be claimed from age 59 until the onset of the public pension at 65. These early benefits had the explicit aim of discouraging work among older individuals, under the assumption that this would create jobs for the young. Early retirement schemes were financed on a pay-as-you-go basis and the level of benefits was a fixed replacement rate of 85% of last earned gross salary. Not only was this replacement rate not adjusted for the age at which claiming started, the years spent in early retirement even counted towards the accumulation of entitlements in the “regular” occupational pension scheme.

At age 65 the early benefits would be replaced by regular pension payouts. Unlike early retirement pensions, the regular schemes were fully funded. Regular pension benefits were, and still are, determined as a replacement rate relative to previous earnings that is a function of the number of contribution years:

$$occpens_t = 0.0175 \times yrswrk_t \times (prevearn_t - 19,000) \quad (21)$$

One year of work or early retirement adds 1.75 percentage point (pp) to the replacement rate. Hence, after 40 years of contributions a worker receives 70% of last earnings in excess of 19,000 Euro. This threshold, the so-called *franchise*, was taken from the website of APG, a large pension fund. Occupational pensions were capped at the level of final earnings.

Taxes

As can be seen in Figure B1, the tax system in place during the 1990s was highly progressive. The figure shows the marginal effective tax rate as a function of income before tax for an individual with zero wealth whose only source of income is a salary (“worker”) or public and occupational pensions (“pensioner”). This effective rate includes tax credit, contributions to social insurance schemes and income tax. The effective rate of a worker with a gross salary below 20,000 Euro per year was 37%, three quarters of which consisted of contributions to social insurance schemes rather than income tax. The marginal rate increases in three steps to 63% for someone who earns more than 80,000 Euro per year. Pensioners face lower tax rates than workers and slightly different income brackets, because they no longer contribute to social insurance and public pensions. Appendix C gives the exact functional form for the tax function.

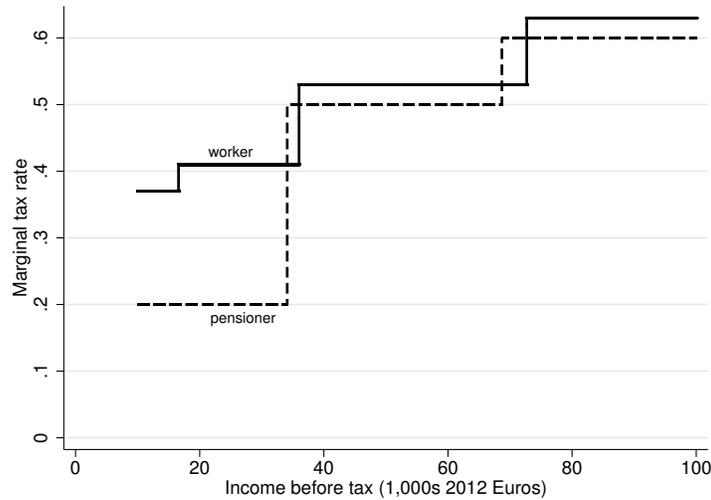


Figure B1: Marginal tax rates for workers and pensioners (estimation sample: '93-'01)

Policy reform: taxes and pensions

The most attractive feature of structural life cycle models is the ability to simulate behavior in counterfactual policy regimes. Therefore, we do not only evaluate how well the model fits the targeted moments in our estimation dataset, but we also check how survival expectations affect the models' ability to predict behavior in a different policy environment. The Netherlands provides an excellent opportunity to study this, since both taxes and occupational pensions were reformed dramatically during the early 2000s.

The Dutch tax system was revised completely in 2001. Figure B2 plots effective marginal tax rates of workers and pensioners as a function of pre-tax income for both the old regime (1993-2001) and the new one (2006-2016). As before, these rates include tax credit as well as contributions to social insurance schemes and income tax. Marginal rates dropped from 37% to 23% for workers with incomes below 20,000 Euro and from 63% to 52% for those with an annual income above 80,000 Euro. Rates increased for lower-middle income individuals with incomes between 20,000 and 40,000 Euro. The effects of the reform on pensioners mirror those on workers. The exact tax functions for both sample periods are given in Appendix C.

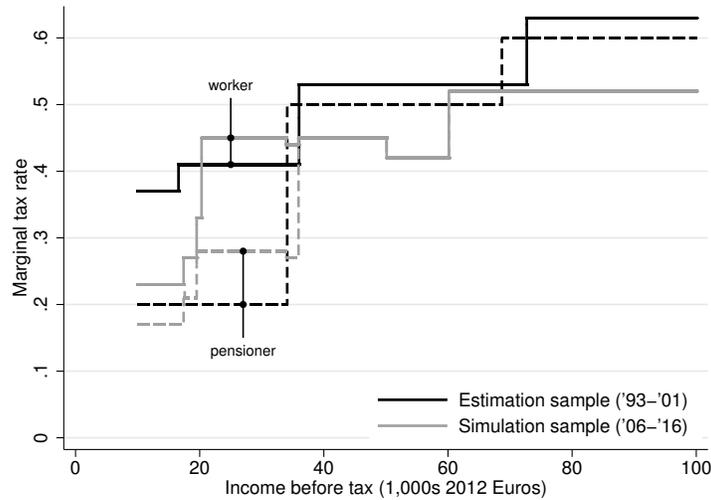


Figure B2: Marginal tax rates for the estimation and simulation samples

While marginal tax rates changed markedly in 2001, the reforms of occupational pensions in the late 1990s and early 2000s were even more consequential. Policymakers realized that while generous early retirement did succeed in pushing people into retirement, there was little evidence that it helped the young to find jobs. Moreover, pay-as-you-go early retirement pensions were not financially sustainable due to population ageing. As a result representatives of employers and employees agreed to replace the two-tier scheme of early and regular pensions by a unified system of occupational pensions that are adjusted actuarially for the age at which they are first claimed. The consequences for workers were twofold. Firstly, they would no longer continue to accumulate pension entitlements while claiming their occupational pension. Secondly, the level of benefits would now be lower if they claim benefits early. These actuarial adjustments amount to 6% for each year of claiming prior to age 65. Claiming after age 65

is rewarded at 7% per year:

$$occpens_t = \begin{cases} \exp[-0.06(65 - t_0)] \\ \quad \times 0.0175 \times yrswrk_t \times (prevearn_t - 19,000) & \text{if } 60 \leq t_0 < 65 \\ \exp[0.07(t_0 - 65)] \\ \quad \times 0.0175 \times yrswrk_t \times (prevearn_t - 19,000) & \text{if } 65 \leq t_0 \leq 70 \end{cases} \quad (22)$$

where t_0 is the age at which occupational pensions are first claimed. Not only did this reduce the attractiveness of early retirement for all individuals, the fact that the same actuarial adjustment is applied for all retirees means its effect may interact with the subjective survival expectations that individuals hold.

Occupational pension plans are organized at the level of the industry or, for large corporations, the firm. Hence, the revisions were not implemented simultaneously across the board. Prior research uses this piecemeal abolishment of the two-part system to estimate the causal effect of the policy change on retirement (e.g. Euwals et al., 2010). The new rules were only applied to younger cohorts of workers born in 1950 or later. We thus limit the estimation sample, with early retirement option, to men born prior to 1950 observed in the 1993-2001 survey waves. The simulation sample consists of men born in 1950 or later observed in 2006-2016. The five year gap between the two samples in combination with the separation by birth year ensures that the samples distinguish between those who do and do not have access to the early retirement scheme in place in the 1990s. While the abolishment of early retirement had a profound impact on the financial tradeoffs around retirement, early claiming now entailed a lower pension, the policy reforms did not offer workers any more choice during the accumulation phase. In the new system too they are automatically enrolled in the one fund that covers their industry and have no discretion regarding the level of contributions or the investment strategy.

C Tax functions

As in the rest of the paper, all monetary amounts in this appendix are denoted in 2012 Euros.

Estimation sample: 1993-2001

We take the tax function for the 1990s from the OECD report *Taxing Wages 1999*. Nominal amounts are transformed to 2012 Euro using the CPI published by Statistics Netherlands.

Net income is a function of earnings; DI and UI benefits; public and occupational pensions; wealth; and age:

$$netinc_t = \tau(earn_t, DIinc_t, UIinc_t, pubpens_t, occpens_t, w_t, t) \quad (23)$$

Individuals have a basic allowance of 5,266 Euro per year and a deductible for work-related expenses that is only available when working positive hours (strictly positive earnings):

$$workcredit_t = \begin{cases} 0 & \text{if } earn_t = 0 \\ 154 & \text{if } 0 < earn_t < 1,283 \\ 0.12 \times earn_t & \text{if } 1,283 \leq earn_t < 15,825 \\ 0.12 \times 15,825 & \text{if } earn_t \geq 15,825 \end{cases} \quad (24)$$

Contributions to unemployment insurance are paid from earnings according to the following function:

$$UIcontr_t = \begin{cases} 0 & \text{if } earn_t < 16,588 \\ 0.0645 \times (earn_t - 16,588) & \text{if } 16,588 \leq earn_t < 744,357 \\ 0.0645 \times 744,357 & \text{if } earn_t \geq 744,357 \end{cases} \quad (25)$$

Contributions to public healthcare insurance are also a function of earnings:

$$pubmed_t = \begin{cases} 0.012 \times earn_t + 132 \times n_t^{1/0.7} & \text{if } earn_t < 32,539 \\ 0.012 \times 32,539 + 132 \times n_t^{1/0.7} & \text{if } 32,539 \leq earn_t < 37,890 \\ 0.012 \times 32,539 & \text{if } earn_t \geq 37,890 \end{cases} \quad (26)$$

Taxable income $taxable_t$ consists of all income components except returns on wealth:

$$taxable_t = earn_t + DIinc_t + UIinc_t + pubpens_t + occpens_t \quad (27)$$

The general allowance, work credit and contributions to UI and public medical insurance are deducted from taxable income:

$$taxable_t = \max\{0, taxable_t - 5,266 - workcredit_t - UIcontr_t - pubmed_t\} \quad (28)$$

Taxes are then levied over taxable income in two steps. The first step concerns contributions to social insurance schemes $scontr_t$. These contributions depend on age and are reduced when the payout of public pensions starts at age 65:

$$scontr_t = \begin{cases} 0.275 \times taxable_t & \text{if } taxable_t < 28,835 \text{ \& } t < 65 \\ 0.275 \times 28,835 & \text{if } taxable_t \geq 28,835 \text{ \& } t < 65 \\ 0.11 \times taxable_t & \text{if } taxable_t < 28,835 \text{ \& } t \geq 65 \\ 0.11 \times 28,835 & \text{if } taxable_t \geq 28,835 \text{ \& } t \geq 65 \end{cases} \quad (29)$$

The second step is the income tax $inctax_t$ levied across three income brackets:

$$inctax_t = \begin{cases} 0.0885 \times taxable_t & \text{if } taxable_t < 28,835 \\ 0.0885 \times 28,835 \\ \quad + 0.5 \times (taxable_t - 28,835) & \text{if } 28,835 \leq taxable_t < 63,419 \\ 0.0885 \times 28,835 \\ \quad + 0.5 \times (63,419 - 28,835) \\ \quad + 0.6 \times (taxable_t - 63,419) & \text{if } taxable_t \geq 63,419 \end{cases} \quad (30)$$

Wealth w_t is taxed at a rate of 0.7%:

$$wtax_t = 0.007 \times w_t \quad (31)$$

The final step is to calculate gross income $grossinc_t$ that includes income from capital (which is not taxed separately):

$$grossinc_t = earn_t + DIinc_t + UIinc_t + pubpens_t + occpens_t + capinc_t \quad (32)$$

The rate of return on wealth is fixed at 4%: $capinc_t = 0.04 \times w_t$. Net income is gross income minus all taxes and contributions:

$$netinc_t = grossinc_t - UIcontr_t - pubmed_t - scontr_t - inctax_t - wtax_t \quad (33)$$

Simulation sample: 2006-2016

The Dutch tax system was revised drastically in 2001, so we code a different tax function for the 2006-2016 sample on which we evaluate the predictions of our model (note that the 2001

wave asks questions about 2000, so it was subject to the old tax regime described above). This tax function is based on the OECD publication *Taxing Wages 2004*, with all nominal amounts expressed in 2012 Euro using the CPI published by Statistics Netherlands.

As was the case during the 1990s, net income is a function of earnings; DI and UI benefits; public and occupational pensions; wealth; and age:

$$netinc_t = \tau(earn_t, DIinc_t, UIinc_t, pubpens_t, occpens_t, w_t, t) \quad (34)$$

The general allowance was replaced by general tax credit which is deducted from income tax rather than pre-tax earnings (the work credit was revised analogously). Moreover, social insurance benefits and pensions are taken into account when calculating contributions to UI and public health insurance. Hence, those contributions are calculated from gross income $grossinc_t$ (which excludes capital income):

$$grossinc_t = earn_t + DIinc_t + UIinc_t + pubpens_t + occpens_t \quad (35)$$

Unemployment insurance contributions $UIcontr_t$:

$$UIcontr_t = \begin{cases} 0 & \text{if } grossinc_t < 17,409 \\ 0.058 \times (grossinc_t - 17,409) & \text{if } 17,409 \leq grossinc_t < 50,115 \\ 0.058 \times (50,115 - 17,409) & \text{if } grossinc_t \geq 50,115 \end{cases} \quad (36)$$

Contributions to public health insurance $pubmed_t$:

$$pubmed_t = \begin{cases} 0.0125 \times grossinc_t + 446 & \text{if } grossinc_t < 33,917 \\ 0.0125 \times 33,917 + 446 & \text{if } 33,917 \leq grossinc_t < 37,490 \\ 0 & \text{if } grossinc_t \geq 37,490 \end{cases} \quad (37)$$

Both types of contributions are deducted from gross income to arrive at taxable income $taxable_t$:

$$taxable_t = \max \{0, grossinc_t - UIcontr_t - pubmed_t\} \quad (38)$$

Social insurance contributions $scontr_t$ are levied over taxable income and are lower at ages in which public pensions are paid out:

$$scontr_t = \begin{cases} 0.324 \times taxable_t & \text{if } taxable < 33,974 \text{ \& } t < 65 \\ 0.324 \times 33,974 & \text{if } taxable \geq 33,974 \text{ \& } t < 65 \\ 0.145 \times taxable_t & \text{if } taxable < 33,974 \text{ \& } t \geq 65 \\ 0.145 \times 33,974 & \text{if } taxable \geq 33,974 \text{ \& } t \geq 65 \end{cases} \quad (39)$$

Income tax $inctax_t$ is levied across four income brackets:

$$inctax_t = \begin{cases} 0.01 \times taxable_t & \text{if } taxable < 18,705 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (taxable_t - 18,705) & \text{if } 18,705 \leq taxable_t < 33,974 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (33,974 - 18,705) \\ \quad + 0.42 \times (taxable_t - 33,974) & \text{if } 33,974 \leq taxable_t < 58,250 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (33,974 - 18,705) \\ \quad + 0.42 \times (58,250 - 33,974) \\ \quad + 0.52 \times (taxable_t - 58,250) & \text{if } taxable_t \geq 58,250 \end{cases} \quad (40)$$

Both the general credit of 2,099 Euro and a work credit that depends on earnings are deducted from income tax. The latter, denoted as $workcredit_t$, is given by:

$$workcredit_t = \begin{cases} 0.0175 \times earn_t & \text{if } earn_t < 9,298 \\ 0.0175 \times 9,298 & \text{if } 9,298 \leq earn_t < 9,316 \\ 0.0175 \times 9,298 \\ \quad + 0.112 \times (earn_t - 9,316) & \text{if } 9,316 \leq earn_t < 19,583 \\ 0.0175 \times 9,298 \\ \quad + 0.112 \times (19,583 - 9,316) & \text{if } earn_t \geq 19,583 \end{cases} \quad (41)$$

The total amount of income tax due taking into account both types of tax credit is:

$$inctax_t = \max \{0, inctax_t - 2,099 - workcredit_t\} \quad (42)$$

Capital income $capinc_t$, determined by a fixed 2% rate of return, is not taxed directly, but instead wealth is taxed at a rate of 1.2% above the threshold of 44,280 Euro:

$$wtax_t = \begin{cases} 0 & \text{if } w_t < 44,280 \\ 0.012 \times (w_t - 44,280) & \text{if } w_t \geq 44,280 \end{cases} \quad (43)$$

Finally, net income $netinc_t$ is given by the sum of gross income and capital income, minus all types of taxes and contributions:

$$netinc_t = grossinc_t + capinc_t - UIcontr_t - pubmed_t - scontr_t - inctax_t - wtax_t \quad (44)$$

D Estimated mortality processes using life tables for other household surveys

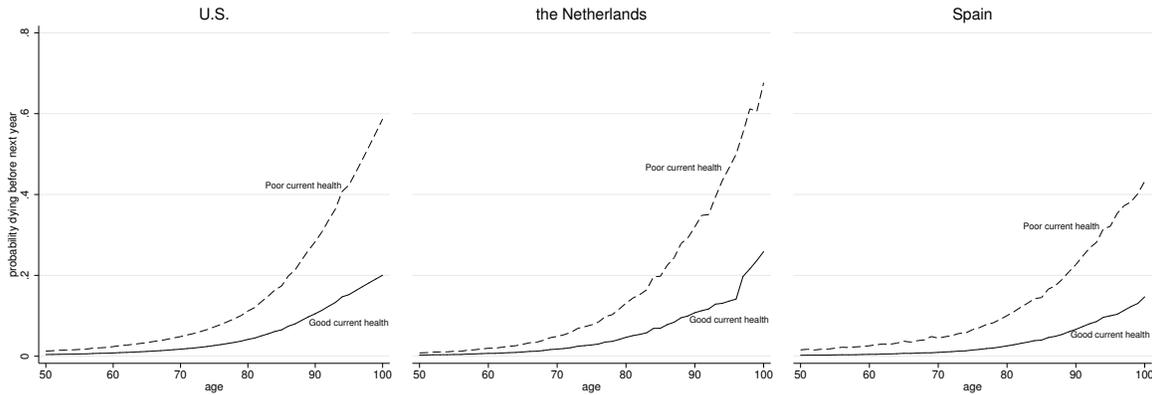


Figure D1: Estimated mortality processes using the HRS (U.S.) and SHARE (the Netherlands and Spain)

Figure D1 shows estimated mortality processes conditional on current health for three samples from prominent household panels (the HRS for the United States and SHARE for the Netherlands and Spain). For each of these samples we combine information on the probability of being in either health state conditional on dying before the next survey wave and the relevant baseline death probabilities from the Human Mortality Database. In other words: we calculate mortality probabilities using the life tables approach. We find that the probabilities conditional on good health are similar to those presented in Figure 2 of the main text for all three countries. Moreover, we corroborate that the probabilities in bad health are much higher than those estimated from subjective data. Though they do not increase to close to 1, as they do for the DHS in panel a. of Figure 2, the probabilities increase with age to 0.6 for the HRS and for the Dutch sample in SHARE and to 0.4 for the Spanish sample. All of these are much higher than the expectations constructed from subjective DHS data, which top out at 0.3. Hence, the high rate of mortality conditional on poor health that we document in the main text is not unique to our dataset.

E Auxiliary processes

In addition to survival we estimate processes for health, equivalence scales, medical expenses and exogenous income of the spouse. The estimates are shown in Figure E1. Black lines refer to the estimation sample and grey lines to the simulation sample. Panel a. contains the probability of being in good health next year conditional on current health and age. Current health is hugely important: the probability of being in good health next year is 65pp higher for individuals who are currently healthy compared to those who are not. Health deteriorates with age, reducing the probability of being in good health next year for the healthy from 95% at age 50 to 80% at age 100. Differences between the estimation and simulation samples are small.

Equivalence scales, panel b. of Figure E1, show that household size declines with age. This reflects both children moving out and the onset of widowerhood as spouses pass away. Panel c. illustrates that medical expenses are low at 2,000-4,000 Euro per year during both sample periods. These expenses mostly consist of premiums for mandatory health insurance and, during the simulation period, small deductibles of no more than 300 Euro. The fact that most medical expenses are generated by insurance rather than treatment explains why the difference between poor and good health is small.

Finally, panel d. contains the net income process of the partner conditional on husband's age and health. Exogenous income was relatively steady around 10,000 Euro per year during the estimation period. During the 2000s spousal income starts out slightly higher at 11,000 Euro per year and declines with age to 6,000 for husbands around age 95. While in the estimation sample the income of the spouse is estimated from men between the ages of 50 and 85, cohort restrictions in the abolition of early retirement mean that the income process for the simulation sample is extrapolated from observations of men aged 50-66. Hence, the steady decline observed for the simulation sample may be an artefact of extrapolation of the downward pattern between ages 50 and 60 also observed for the estimation sample. However,

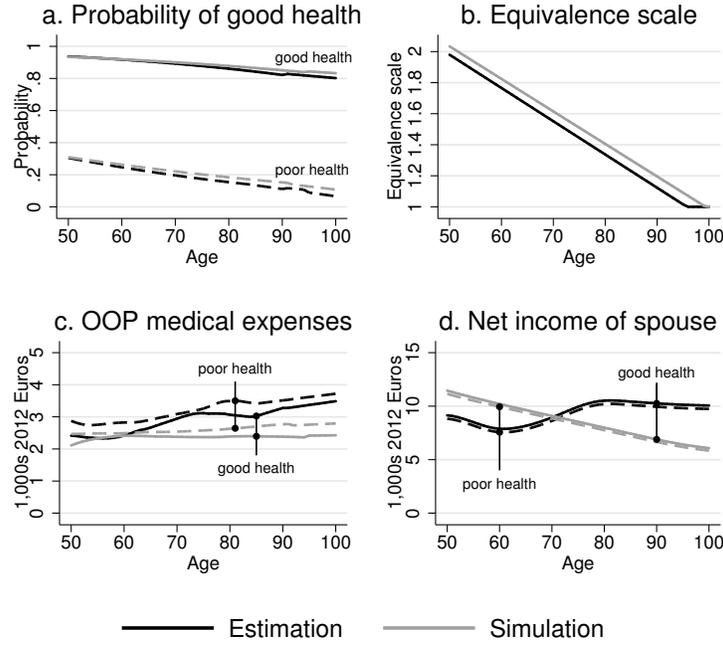


Figure E1: Auxiliary processes (estimated from DHS dataset)

robustness analysis indicates that none of the results for the simulation sample discussed in the main text change when we set the income of the spouse equal to 1.25 times that for the estimation sample (so that the age profiles for the two samples are analogous and the level difference is in line with that observed for ages 50-60).

In addition to the processes shown in Figure E1, we also estimate an AR(1) process for wages following the methods of Gourinchas and Parker (2002). We estimate the autoregressive coefficient at 0.95 and the standard deviation of the innovations at 0.01 and use the same parameters for the estimation and simulation samples. We compute the transition matrix for discretized wages based on the approach described in Tauchen (1986).

F Out-of-sample predictions for participation in the labor market

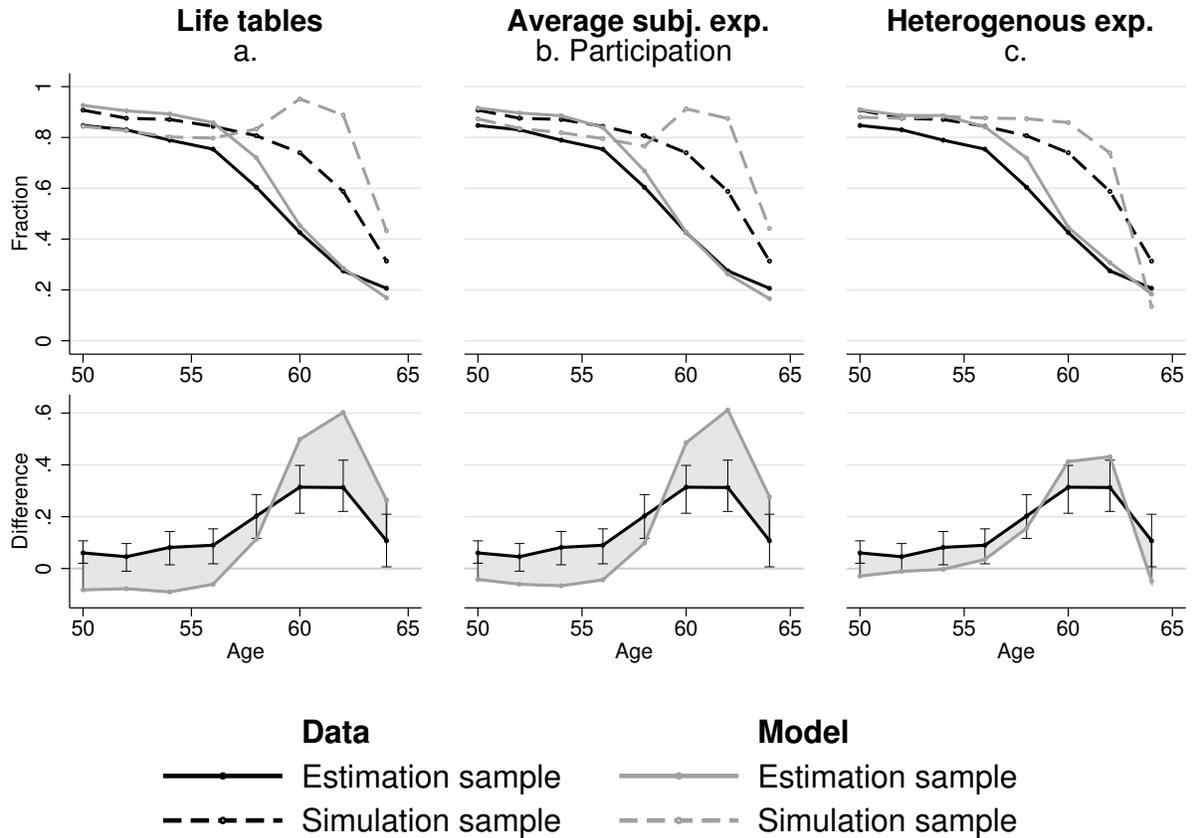
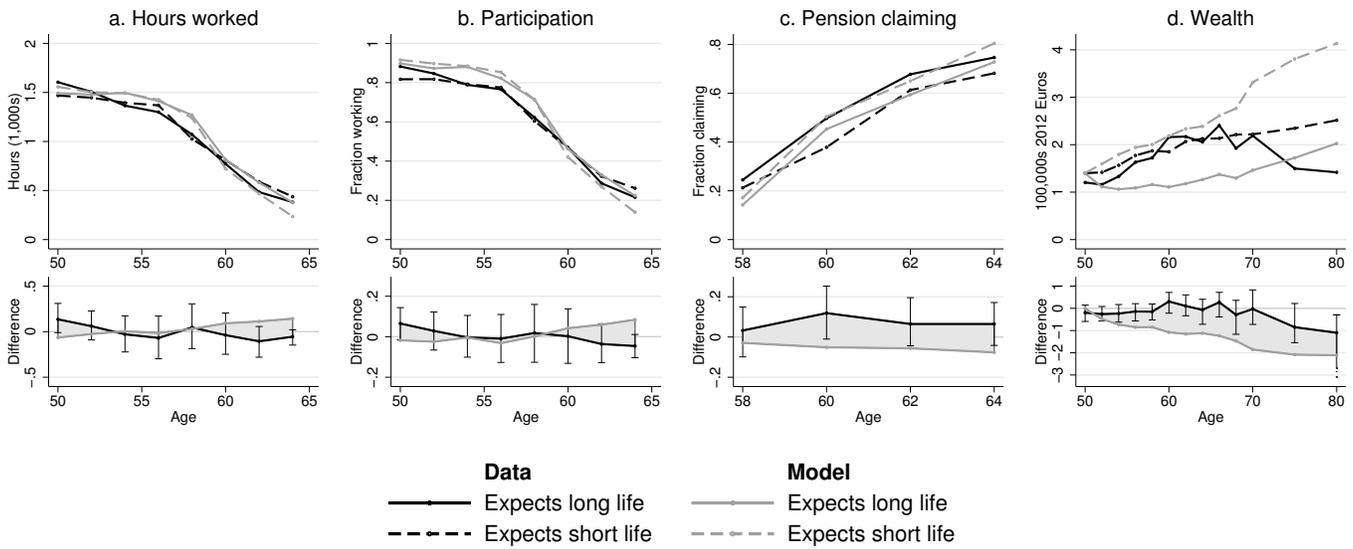


Figure F1: Simulated and observed labor supply in the estimation ('93-'01) and simulation ('06-'16) sample

G Model predictions conditional on survival for different preferences (estimation sample: '93-'01)

Preferences: estimates for heterogeneous expectations



Preferences: estimates for average subj. expectations

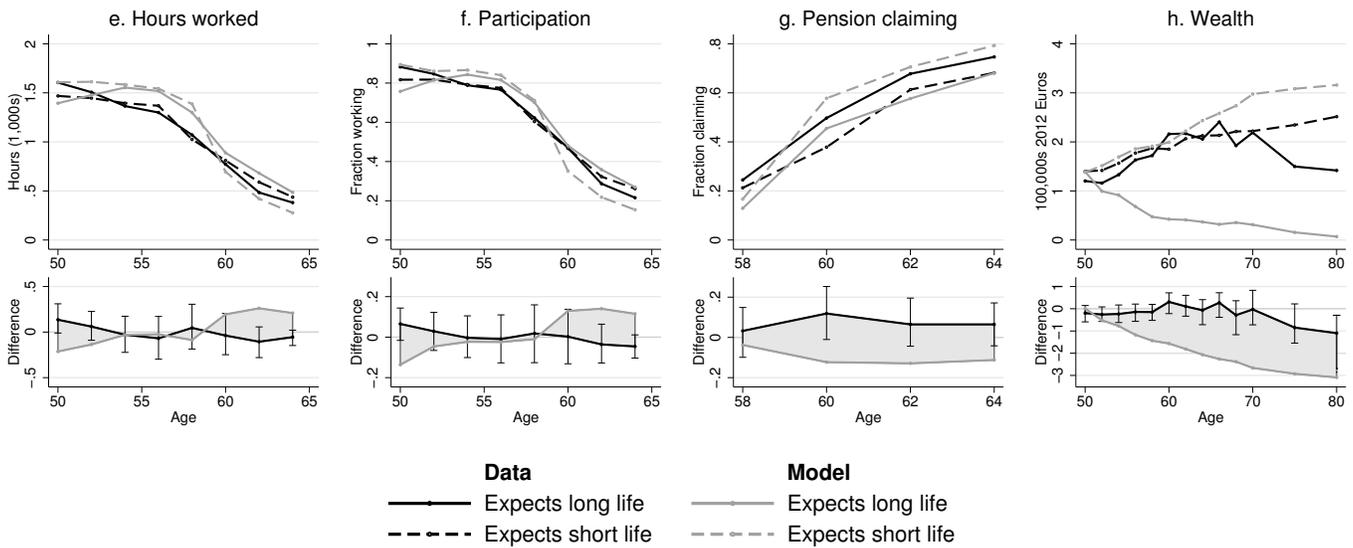


Figure G1: Behavior conditional on survival expectations for the *estimation sample* for different sets of preference estimates (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)