

No. 2021-034

**EVALUATING THE ACCURACY OF COUNTERFACTUALS
THE ROLE OF HETEROGENEOUS EXPECTATIONS
IN LIFE CYCLE MODELS**

By

Jochem de Bresser

25 November 2021

ISSN 0924-7815
ISSN 2213-9532

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

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

November 1, 2021

Abstract

This paper shows that subjective survey information on survival expectations can be used to improve the out-of-sample predictions of a dynamic model of retirement 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. The models based on homogeneous expectations anticipate a 4-5-year increase in the average retirement age in the new regime, compared with an observed increase of 2.6 years. The model with heterogeneous expectations, on the other hand, predicts a more realistic increase of 2.7 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.

*This version of the paper was resubmitted to the *Review of Economic Studies* in May 2021, which has since requested a second round of revisions. A new version is expected by spring 2022. 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). The paper benefitted substantially from the insightful comments of four anonymous referees and editor Áureo de Paula. I thank Max Groneck and David Sturrock for their careful discussions and Jeff Campbell and seminar participants at Berlin Applied Micro Seminar, 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.

Key words: Subjective expectations, life cycle model

JEL-codes: D84; J14; C34

1 Introduction

Life cycle models are useful tools to predict the effects of policy changes, such as pension reforms, on labor supply and saving. Such models postulate forward-looking agents, who base their decisions on expectations as well as preferences (e.g. French, 2005; De Nardi et al., 2010; French and Jones, 2011). 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. Saving, for instance, is determined not only by patience, but also by one's perceived longevity. Those 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 out-of-sample predictions after a substantial pension reform. 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 modeled 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 across individuals. The second approach sets the *level* in accordance with subjective expectations and equates the probability of death with the average subjective probability. The third technique uses reported probabilities to their full potential and introduces individual-level heterogeneity in subjective survival.

Both approaches that incorporate subjective beliefs are based on a measurement model

that takes into account the elicitation of reported probabilities on a coarse 11-point scale. Expectations are assumed to be exogenous and stable, and the incorporation of heterogeneity in subjective survival only introduces a single additional state variable in the life cycle model. It has the advantage of relating 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. The analysis is based on the idea that people act on their expectations, regardless of whether those are reasonable for a given information set. Hence, it is not concerned with the rationality of expectations.

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).

This paper fits in 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 modeling of uncertainty, budgets and decisions.¹ Recent models add uncertainty in wages and unemployment (Gourinchas and Parker, 2002; French, 2005; Low et al., 2010; De Nardi et al., 2020) 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 budgets. 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.

The paper makes three contributions. Firstly, it incorporates heterogeneous subjective survival expectations in a rich life cycle model of labor supply, benefit claiming and saving.⁵ The few studies that used subjective survival have mostly assumed homogeneous expectations and/or have been limited to saving (e.g. Heimer et al., 2019; Van der Klaauw and

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.

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, where they mostly concern outcomes of decisions rather than exogenous processes such as mortality (Van der Klaauw, 2012; Pantano and Zheng, 2013). Examples include the perceived effectiveness of contraception (Delavande, 2008); the career perspectives of college majors (Zafar, 2011; Stinebrickner and Stinebrickner, 2013); and risk perceptions regarding HIV/AIDS and fertility (Shapira, 2017).

Wolpin, 2008). Gan et al. (2015) introduces individual-specific subjective survival in a model of saving and shows that doing so affects preference estimates, raising risk aversion and lowering the discount factor. The present paper builds on Gan et al. (2015) in terms of the modeling of expectations, taking into account measurement error and allowing beliefs not to be proportional to life tables in line with recent evidence (O’Dea and Sturrock, 2020; Heimer et al., 2019; Wu et al., 2015). Moreover, it embeds those beliefs in a richer life cycle model that includes labor supply and benefit claiming. Secondly, this paper validates the models by their predictions in a counterfactual policy regime. Such extrapolation is a key application for structural models and provides a strong criterion for model validation, because it limits the scope for data mining (Keane and Wolpin, 2007; Keane, 2010; Keane et al., 2011). While the use of hold-out samples is most powerful if data remain embargoed during the model building phase (Schorfheide and Wolpin, 2012, 2016), this is not feasible for publicly available data such as the DHS. Few papers have used this type of validation, since large policy reforms that generate quasi-random variation in choice environments are rare.⁶ Two examples that use policy reforms to validate structural models in the different context of welfare incentives and female labor supply are Keane and Moffitt (1998) and Low et al. (2018). To our knowledge, Lumsdaine et al. (1992) is the only previous effort to apply this validation approach to models of retirement. Thirdly, the institutional context of the Netherlands provides an interesting departure from the U.S., which has been analyzed in almost all previous work. In particular, universal, mandatory and generous pension annuities insure Dutch workers against longevity risk to a much greater extent than is common in the U.S.

The results show that estimated preferences and counterfactual predictions depend on survival expectations. In particular, heterogeneous expectations lead to more risk aversion

⁶The structural research that has employed validation through extrapolation has mostly relied on social experiments, in which one treatment arm is used for estimation and the others for validation. Examples include Wise (1985) on housing subsidies; Lise et al. (2005) on financial incentives to exit welfare; Todd and Wolpin (2006) on the PROGRESA experiment of cash transfers conditional on school attendance; Duflo et al. (2012) on financial incentives to reduce teacher absenteeism in India; Galiani et al. (2015) on location decisions in Boston; and Arcidiacono et al. (2016) on school vouchers in India.

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, within-sample model fit is similar for all three models. Labor supply and pension claiming are matched reasonably well, though the models over-predict 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 homogeneous expectations predict a larger increase of 4-5 years to a retirement age above 65. The model with variation in survival expectations anticipates a more reasonable increase of 2.7 years to an average retirement age of 63.7. 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. These results show that the specification of the mortality process matters for counterfactuals. Subjective data yield more accurate forecasts than actuarial tables, since they better approximate the beliefs that underlie behavior.

The rest of the paper is organized as follows. Section 2 describes the life cycle model and the different approaches to model survival expectations. The estimation routine is explained in section 3, after which section 4 introduces the data. Section 5 contains the results and section 6 concludes.

2 Models

This section describes the retirement model (section 2.1) and the different methods used to approximate survival expectations (section 2.2).

2.1 Retirement 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 public and occupational pensions. 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 within-sample fit, the model is validated through the accuracy of its 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 forecasts generated by the model with observed behavior during that period. The following sub-sections explain the model in more detail.

2.1.1 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: $h_t \in \{0, 1500, 2000, 2500\}$. These levels are set to cover the

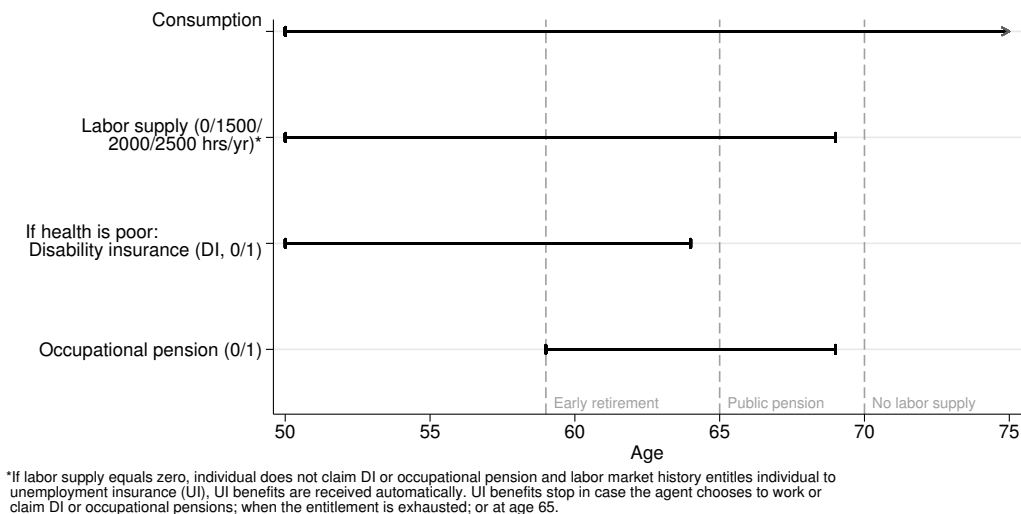


Figure 1: Decisions available at various ages

empirical distribution of hours worked. We assume that nobody works beyond age 70 in accordance with observed behavior.⁷

The model includes simplified disability (DI) and unemployment (UI) insurance schemes that provide alternative routes into retirement. UI and DI benefits can be claimed between ages 50 and 64 after which both are replaced by the public pension. Receipt of UI is automatic if the individual does not work or receive DI or occupational pension benefits and is eligible based on his work history. The choice to claim DI is restricted to those in poor health: an individual has to be in poor health to have the option to claim DI. The model does not include an application process for DI or incomplete acceptance. DI is not an absorbing state: agents choose whether to claim in each period in which they are in poor health. Hence, the evolution of health drives the dynamics of DI claiming. UI also has a dynamic aspect, since claiming depletes a limited entitlement as described below. This treatment of UI and DI as easily accessible sources of income for pre-retirees is in line with the institutions in place at the time. By the 1990s the use of these schemes as pathways into retirement was

⁷In the data fewer than 1% of individuals report a retirement age outside the interval 50-70. We drop those observations from the validation exercise.

common practice and widely acknowledged and condoned by employers, unions and the Dutch government (Heyma, 2004; Trommel and De Vroom, 1994; De Jong and Aarts, 1992).

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 analysis. Hence, claiming the *public* pension is not contingent on any choice made by the agent. 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. That is, from age 65 onward those who receive occupational pensions collect both occupational and public pensions. One cannot claim occupational pension, DI or UI benefits and continue working. 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.

Together public and occupational pensions accounted for 94% of retirement income in 2013. The model does not include a decision to switch employers, because occupational pension funds are organized at the level of broad sectors or industries. Switching employers within such sector does not affect accumulation of entitlements: workers would continue to contribute to the same fund. Similarly, when moving to another sector current entitlements can either be transferred to the new fund or remain dormant in the fund in which they were accumulated. Pension contracts typically do not include a minimum vesting period and do not create incentives for workers to stay put. Therefore, the model allows agents to choose labor supply, but abstracts away from the choice of employer.

2.1.2 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 (Scholz et al., 2006). The equivalence scale is the sample average at a given age and does not vary between decision makers (household size and marital status are not state variables in the model).⁸ 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

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

The total time endowment is set to 4,000 hours per year⁹ and individuals incur a fixed cost ψ when working a positive number of hours. Such fixed cost of work reduces the time budget beyond the hours spent working and raises the marginal value of leisure. Consequently, higher fixed costs induce workers to work fewer hours. Analogously, the time budget is reduced by δ when in bad health. Such time cost increases the marginal value of leisure and rationalizes lower labor supply in poor health. The parameterization of the leisure costs of work and poor health is based on French (2005).

⁸The equivalence scale n_t at age t is the sample average across all I_t households in which the male has age t : $n_t = 1/I_t \sum_{i=1}^{I_t} hhsizet_{it}^{0.7}$. The size of one such household may be larger than 2 if it contains (adult) children. This is the case for 38% of households in age bracket 50-60 and 7% for ages 61 and older.

⁹The time endowment is based on $(24 - 7) \times 5 \times 47 = 3995$ hours per year that that decision maker considers when deciding on work and leisure. That reserves a minimum of 7 hours per day for sleep and does not count weekends and 5 weeks of vacation. These constraints are assumed to be outside the individual's control.

Claiming either UI or DI benefits entails stigma costs, also measured in hours of leisure:

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

Stigma costs have been proposed as an explanation for low take-up rates of benefits among the eligible population (Moffitt, 1983). Welfare stigma is assumed constant throughout the pre- and post-reform periods, since there has been no major reform to UI or DI and the fractions of the sample that receive these types of benefits are stable. Given that ψ , 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 the ages 50 to 84. We match 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 people living in the same household. The bequest weight varies as 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.¹⁰ 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

¹⁰The overall curvature of bequest utility is governed by the risk aversion σ . K shifts around the ‘effective’ level of bequests, which determines local curvature since the function is more curved at lower levels of bequests.

at which many spouses are alive.

2.1.3 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. In contrast to survival, the DHS does not elicit subjective probabilities that measure expectations regarding these variables. Therefore, we follow the usual approach and assume that agents expect health and wages to develop according to estimated processes.

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:

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

These probabilities are estimated by a logit model which regresses the lead of health on current health and age.

Future wages evolve according to an AR(1) process. While the initial conditions allow for cross-sectional dependence between tenure and wage, the dynamics of wages do not depend on work experience. This is in line with the institutional context of the Netherlands, in which the relation between tenure and salary is fixed through collective labor agreements between employers and employees. In the pre-retirement age range employees are usually at or near the top of the salary ladder for their occupation, which means that additional experience does not affect wages (Ter Weel, 2003; Deelen and Euwals, 2014). In line with this, the data do not indicate that additional experience is associated with changes in wages for the age group 50-64 (estimates from Fixed Effects models available on request).

2.1.4 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 A and Appendix B 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%.¹¹ 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} .¹² 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. Medical expenditures were of limited importance in the Netherlands during both periods studied in this paper 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.

¹¹During the pre-reform period wealth tax was levied on the stock w_t rather than on capital income.

¹²For reasons of computational complexity, the model limits the role of the spouse to the exogenous income stream inc_t^{sp} and the equivalence scale n_t . An extended model that includes the spouse as an additional decision maker and explicitly accounts for her survival is left for future work.

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

$$x_t = \max \{x_t, n_t c_{\min}\} \quad (7)$$

The minimum consumption level c_{\min} is set to 7,000 euro per year, in line with the income level at which one qualifies for social assistance.

2.1.5 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 2005. A law in effect from January 1st 2006 places a 52% tax penalty on early retirement pensions, effectively prohibiting such schemes. We explain the most important aspects of the reform here and refer the reader to Appendix A 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 prospects of the young, the key feature 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.

In the new policy regime 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 no longer accumulates pension rights while claiming benefits.¹³

A second change concerned taxes: a new law reformed the system of income taxation on the first of January 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 A and B for details).

2.1.6 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 ($yrswrk_t$, including years in early retirement), 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.¹⁴ The value function is approximated by linear interpolation between grid points. It is possible to accrue a partial work-year if one receives transfers or works part-time (one year of UI receipt adds 0.375 working year to one's pension

¹³Occupational pensions include a provision for the partner in case of widowhood. Modeling such partner pensions would require one to add the partner's age and subjective longevity to the model as additional state variables. The model abstracts away from partner pensions to reduce computational complexity.

¹⁴Discretization 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).

entitlement and a year on DI counts for 0.7). 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 of consumption. For all $d \in \mathcal{D}$ we calculate the value corresponding to that decision as:

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

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. Golden section search is used to find optimal consumption for each discrete choice d . The agent optimally selects the discrete choice d^* that yields the highest value:

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

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.

2.1.7 Survival expectations and behavior

Equation 10 shows that survival expectations determine the relative importance of bequests and the utility of continued survival (which reflects future consumption and leisure). This is distinct from patience, which drives the relative weight of the present versus both types of future rewards. A highly patient person will accept less utility in the present, in order to secure more in the future. He will want to work and save now to afford consumption, leisure or bequests tomorrow. Survival probabilities determine how he weights tomorrow's states of the world against each other. While a more patient individual will unambiguously save more, someone who expects to live longer may either save more or less depending on the relative strengths of the bequest, pre-cautionary and life-cycle motives. If saving is primarily motivated by the latter two, by events occurring within one's life-span, an individual who expects a long life will save more. However, the opposite is true if savings are motivated by the desire to leave bequests. Hence, the relationship between wealth accumulation and

subjective survival depends on preferences and on the institutional context. In light of generous social security and universal pensions and health insurance in the Netherlands, we expect the bequest motive to be relatively important compared to the U.S.

2.2 Survival expectations

We pay particular attention to modeling survival expectations. The typical approach when estimating a life cycle model, dating back at least to French (2005), is to combine survey information on health and survival with actuarial life tables. We implement this life table technique as a baseline and contrast the results with an approach that uses reported probabilities to approximate expectations. The following two subsections explain both methods.

2.2.1 Life tables

The probability of dying in the current period conditional on health M_t and current age t is given by:

$$p_d(M_t, t) = \Pr(t \leq T < t + 1 | M_t = m, T \geq t) = \frac{\Pr(t \leq T < t + 1 \cap M_t = m | T \geq t)}{\Pr(M_t = m | T \geq t)} \quad (12)$$

$$= \frac{\Pr(M_t = m | t \leq T < t + 1, T \geq t) \times \Pr(t \leq T < t + 1 | T \geq t)}{\Pr(M_t = m | T \geq t)} \quad (13)$$

Life tables report forecasts of $\Pr(t \leq T < t + 1 | T \geq t)$, the conditional probability of dying at a given age. We use the Human Mortality Database. Survey data are used to estimate the probabilities of being in a certain health state, $\Pr(M_t = m | T \geq t)$, and of being in that state conditional on dying before the end of the year, $\Pr(M_t = m | t \leq T < t + 1, T \geq t)$. Health is a binary variable in the model, it is either good ($M_t = 2$) or bad ($M_t = 1$), and logit models of health on age and, for the probabilities that condition on mortality, a death-indicator are used to estimate the probabilities.

Construction of the death-dummy requires one to observe the year within which panel members pass away. Unfortunately, the DHS survey does not field exit interviews with surviving relatives and does not provide direct information showing when a panel member dies. Instead, we infer death from spouses' transitions in marital status from married to widowed. The fact that the data do not include a better measurement of death renders direct estimation of $p_d(M_t, t)$ unreliable.¹⁵

2.2.2 Subjective expectations

Subjective survival expectations are elicited by means of the following survey 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 two or three target ages, 65 in the example, in each survey. These target ages depend on the current age of the respondent. We interpret the answers as probabilities, with 1 corresponding to 10% chance etc. (De Bresser, 2019, shows that this is a valid interpretation).

Table 1 summarizes the data. The average reported probabilities decline monotonically with target age from 79% for age 65 to 18-26% for age 100. Average survival probabilities for most target ages are higher in the post-reform period and there is substantial heterogeneity at all target ages, with standard deviations around 20-30 percentage points (pp). The measurement model described below aggregates these subjective probabilities into individual-specific survival curves and achieves three goals. Firstly, it enables extrapolation of the probabilities

¹⁵While mortality rates cannot be estimated straight from the DHS because death is under-reported, comparison with other surveys that provide better measures of in-sample mortality shows that $\Pr(M_t = m | t \leq T < t + 1, T \geq t)$ is measured reliably. In particular, the life tables approach yields very similar survival probabilities when applied to the Dutch subsample of the Survey of Health, Ageing and Retirement in Europe (SHARE), which includes exit interviews.

Table 1: Descriptive statistics of reported survival probabilities (0-100%)

	Current age	Pre-reform sample ('93-'01)			Post-reform sample ('06-'16)		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
Target age 65	50-65	–	–	–	879	79	16
Target age 75	50-65	1,629	69	20	3,706	70	19
Target age 80	50-70	2,008	56	23	5,169	59	23
Target age 85	65-75	646	48	22	2,505	55	23
Target age 90	70-80	381	35	24	1,792	41	25
Target age 95	75-85	151	27	23	1,185	30	24
Target age 100	80-89	43	26	30	553	18	21

given in equation 12 to all ages between 50 and 100. Secondly, the model reduces heterogeneity in expectations to a single parameter. Thirdly, it mitigates measurement error, both because it combines information from multiple probabilities and because it explicitly accounts for rounding.¹⁶ The measurement model and the explanation below are based on De Bresser (2019). More details regarding the likelihood are provided in Appendix C.

In the model expectations follow Gompertz distributions with a common baseline hazard tilted 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 t is given by:

$$S_{isk}(ta_k, t) = \Pr(T \geq ta_k | T \geq t) = \frac{\exp\left(-\frac{\gamma_{is}}{\alpha} (\exp(\alpha (ta_k/100)) - 1)\right)}{\exp\left(-\frac{\gamma_{is}}{\alpha} (\exp(\alpha (t/100)) - 1)\right)} \quad (14)$$

α determines the shape of the baseline hazard, i.e. the relation between mortality and age, and is common to all individuals and survey waves.¹⁷ Variation in expectations is generated by γ_{is} , where the indices i and s reflect heterogeneity across individuals and survey-waves respectively. This variation is driven by both observed and unobserved components: $\gamma_{is} =$

¹⁶Rounding is the tendency for respondents to report one number (e.g. 10%) when the true subjective probability lies in some interval (e.g. 5-15%). It has been shown to affect reported probabilities of many types (Manski and Molinari, 2010; De Bresser and Van Soest, 2013; Kleinjans and Van Soest, 2014; Bissonnette and de Bresser, 2018). In addition, the DHS induces respondents to round their answers by offering a coarse 11-point response scale.

¹⁷We divide both the target age and the current age by 100 to facilitate estimation of baseline hazard α .

$\exp(\mathbf{x}'_{is}\boldsymbol{\beta}_1 + \zeta_i + \eta_{is})$. The vector \mathbf{x}_{is} contains observed covariates of subjective mortality, most importantly health. Unobserved heterogeneity is captured by two error components: individual effects ζ_i and survey wave effects η_{is} . The former are common to all probabilities reported by individual i , while the latter are shared only by probabilities reported in survey wave s and generate additional persistence between the probabilities reported in a single survey. Individual and survey-wave effects are modeled as random effects with normal mixing distributions.

We do not observe S_{isk} directly, since reported probabilities are not equal to Gompertz-probabilities. Instead, the reported probabilities are perturbed by recall error ε_{isk} :

$$P_{isk}^* = S_{isk} + \varepsilon_{isk} \tag{15}$$

where $\varepsilon_{isk} \sim \mathcal{N}(0, \sigma_{is}^2)$, independent of all covariates and across target ages, survey waves 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. Based on previous research documenting that lower educated respondents report probabilities that are further from parametric distributions, we allow for heteroskedasticity and model the variance of recall errors as $\ln(\sigma_{is}) = \mathbf{x}'_{is}\boldsymbol{\beta}_2$ (De Bresser and Van Soest, 2013; De Bresser, 2019).

The recall errors allow P_{isk}^* to take any value and hence equation 15 could be used to construct a likelihood. However, such continuous model would not account for the coarse answer scale, or for excess bunching at zero, fifty and one hundred percent. The measurement model captures bunching by censoring probabilities between zero and one hundred percent and by rounding. In light of the 11-point answer scale, we distinguish between rounding to multiples of 10 (possible answers: 0, 1, 2, ..., 9, 10), 50 (0, 5, 10) or 100 (0, 10). Note that while some answers are only consistent with a single level of rounding, e.g. ‘1’ must be

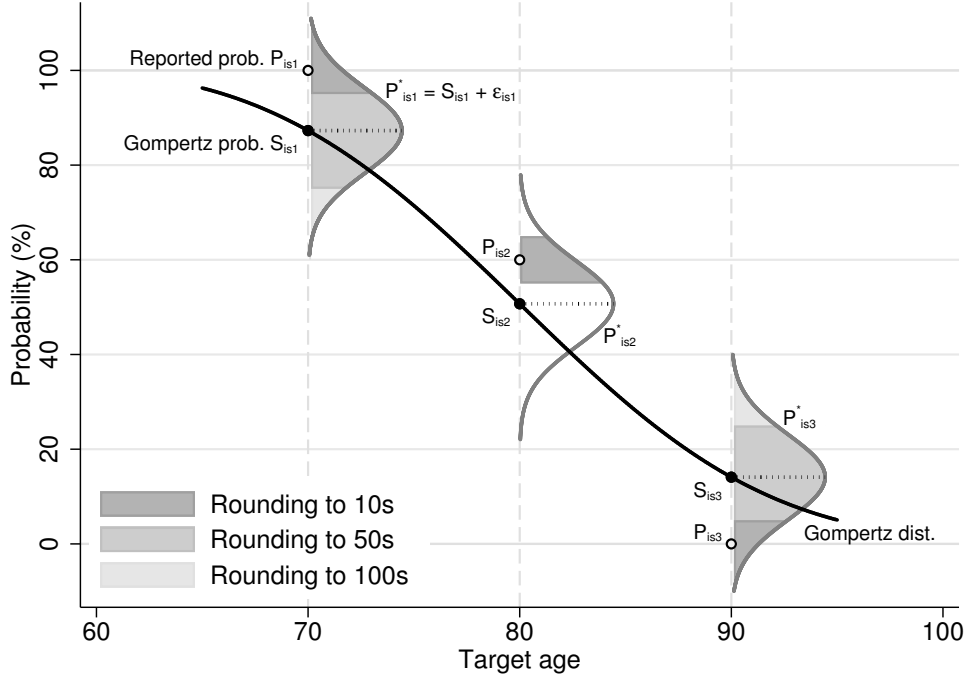


Figure 2: Illustration of the logic of the rounding model

rounded to a multiple of 10%, others may result from multiple levels (a ‘5’ may be rounded to 10s or 50s and a ‘0’ or ‘10’ can result from all three levels). Hence, rounding is a latent construct and the likelihood averages across the relevant levels of rounding that may yield the observed response. Each level of rounding implies a different interval for P_{isk}^* : a ‘5’ rounded to multiples of 10% implies that $P_{isk}^* \in [45, 55)$ and $P_{isk}^* \in [25, 75)$ if rounded to 50s.

Figure 2 visualizes the logic of the model. The empty circles represent three hypothetical reported probabilities P_{isk} , equal to 100, 60 and 0% . 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_{isk} in equations 14 and 15. Vertical normal distributions capture the recall error ε_{isk} . 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, a reported probability of 100% implies that the latent

probability P_{itk}^* must be larger than 95. 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 shaded area for that probability, ranging from the lower bound of 55 to the upper bound of 65.

The model is estimated using Maximum Simulated Likelihood. Observed covariates and individual and survey wave effects generate a (prior) distribution for γ_{is} which captures variation in expectations. Average subjective expectations are calculated as the average one-year survival probabilities across the distributions of observed and unobserved heterogeneity (based on 10,000 draws of individual and survey-wave effects). Hence, the subjective average is computed from the prior distribution of unobserved heterogeneity and does not condition on the probabilities or covariates reported by any given respondent.

Individual-specific expectations are approximated by the posterior mean, conditional on the probabilities and observed covariates for a respondent. This modeling approach implies that expectations are exogenous, known at age 50 and vary over time only as a function of health. The focus on variation between rather than within individuals is based on previous work indicating that the former is both more reliable and quantitatively more important than the latter (De Bresser, 2019). This is especially relevant for the present study, since the measurement model conditions on health and respondents are observed for a relatively short period (86% of respondents in the pre-reform period are observed in 1 or 2 waves). Appendix C describes in detail how observed data, covariates and reported probabilities, are used to infer γ_{is} for each individual in the sample.

Table 2: 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.
ψ	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 $u(c_t, l_t) = \frac{n_t}{1-\sigma} \left[\left(\left(\frac{c_t}{n_t} \right)^\kappa l_t^{1-\kappa} \right)^{1-\sigma} - 1 \right];$ $l_t = 4,000 - h_t - \psi \mathbb{I}\{h_t > 0\} - \delta \mathbb{I}\{\text{health} = \text{bad}\} - \phi \mathbb{I}\{\text{claim DI}\} - \xi \mathbb{I}\{\text{claim UI}\}$	
^b $b(w_t, n_t) = \exp[\theta_0 + \theta_1 n_t] \times \frac{(w_t + K)^{(1-\sigma)\kappa}}{1-\sigma}$	

3 Estimation

We estimate the processes for survival, health, medical expenses, wages, spousal income and equivalence scales separately outside the retirement model and focus here on the estimation of preference parameters. Preferences are determined by the 10 parameters collected and explained in Table 2. 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 to data by means of a method of moments objective function. The MSM estimator is the parameter vector that minimizes the difference between simulations and 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). While simulated annealing starts from a single initial parameter vector, it achieves robustness

by sampling the parameter space widely, exploring areas around parameter vectors away from the current best (see Goffe et al., 1994, for a detailed discussion of simulated annealing).

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 ($2 \times 2 \times 8$). Labor supply at ages 66-70 is not targeted, since the data show that workers with occupational pensions retire by age 65. According to the model, these retirees no longer make labor supply decisions. 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 ($32 + 8 + 8 + 4 + 39$). In order to ensure a level playing field, all three versions of the model are estimated based on the same set of moments. No moments are targeted that condition on longevity, since life tables do not capture variation in expectations.

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 profound effects 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. The parameter that allows the bequest weight to depend on the equivalence scale, the only new parameter in the specification of utility, is particularly sensitive to wealth holdings at ages older than 70. Fixing it at alternative values substantially worsens the fit of wealth profiles generated by the model.

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. Standard errors are not adjusted for estimation uncertainty in the auxiliary processes, but they do reflect the noise introduced by simulation. Appendix D provides more details on the objective function and calculation of standard errors.

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 admin-

istered to the CentERpanel by CentERdata, which is affiliated with Tilburg University.¹⁸ 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 rich information on psychological and economic aspects of financial behavior.

We model the labor supply and wealth of men who are at least 50 years old. Sample selection proceeds as follows. The auxiliary processes of (subjective) survival, health,¹⁹ 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). Data moments to be targeted in estimation are computed from this sample after dropping men who were self-employed during any survey wave and those with fewer than 15 years of work experience (the self-employed are dropped because they are not covered by occupational pensions). Moreover, in order to isolate the groups with and without access to early retirement, we drop all men born after 1949 from the pre-reform sample. Similarly, we drop all men born before 1950 from the post-reform sample. We estimate the net income stream of the spouse on the partners of the men in these reduced samples.

Auxiliary processes are estimated on broader samples compared to the moments targeted in estimation, in particular with respect to birth cohort. The reason is that those processes have to be approximated for all ages in the model, up to age 100, and doing so would rely excessively on extrapolation if all individuals born before 1950 would be dropped from the post-reform sample (for which no observations would be available for ages older than 67). Nonetheless, for the pre-reform sample very similar processes are obtained when the

¹⁸See <https://www.centerdata.nl/en/databank/dhs-data-access> for more information.

¹⁹The binary measure for health is constructed by collapsing a five-point scale for subjective health into two categories. The bottom two levels, labeled ‘bad’ and ‘not very good’, are combined into ‘bad’ health, the top three into ‘good’ health.

Table 3: Overview of sample selection due to different criteria (sample sizes)

	Pre-reform sample (1993-2001)	Post-reform 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 cohorts were born before 1950 for the pre-reform sample and after 1949 for the post-reform sample.

additional restrictions on self-employment, work experience and birth cohort are imposed. Hence, the difference in samples does not lead to misspecification (auxiliary processes for restricted samples are available on request).

Table 3 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 pre-reform sample observations are lost mostly due to incomplete records for work experience. In the post-reform sample the restriction to cohorts born after 1949 results in the largest loss of data.

4.1 Survival expectations

Figure 3 shows mortality estimated from life tables using the approach described in section 2.2.1 and the mean and distribution of the probabilities constructed from survey data according to the methods explained in section 2.2.2. The figure plots the probability that an individual dies within one year conditional on attaining a given age. Looking at the top panels, the life tables approach yields much higher probabilities of dying when in poor health than do average subjective expectations. The actuarial method results in a probability of death close to 0.7 for ages above 90, while subjective expectations only increase to 0.3 in the pre-reform sample. Probabilities in good health, on the other hand, increase to around 0.2 for both approaches, with slightly higher probabilities derived from subjective data. Condi-

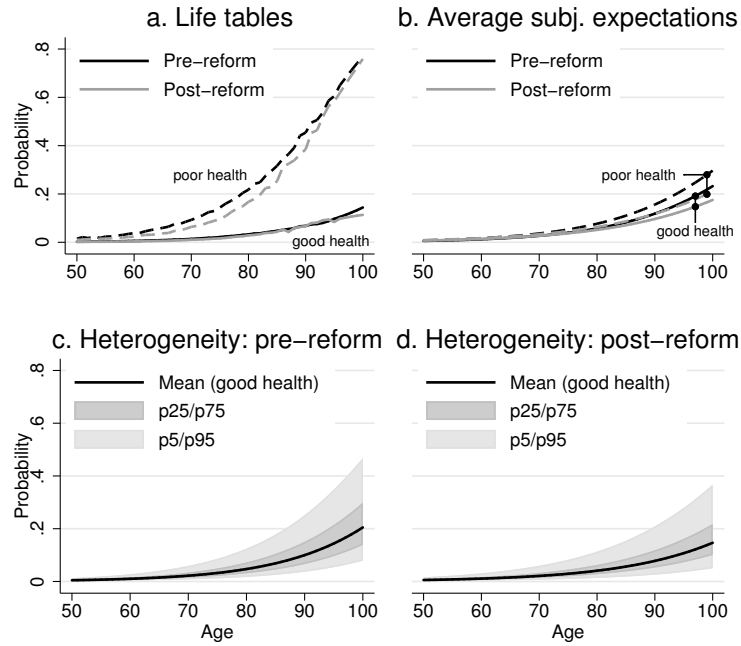


Figure 3: Estimated mortality expectations (probability of dying at a given age conditional on surviving to that age)

tional mortality given health can be combined with the estimated health process to simulate lifespans. Such simulations based on the probabilities in panel a. yield relatively short lives: the mean age of death across 5,000 simulations is 76.2, with a median of 77. As a comparison, Statistics Netherlands reports a life expectancy of 78.9 for a man aged 50 in 2000.²⁰ The subjective probabilities of panel b. give a higher, if still a bit low, mean lifespan of 78.1 years and the average across the heterogeneous expectations in panel c. is 80.2 years. The fact that average survey expectations differ from life tables implies that preference estimates may be sensitive to the modeling of survival, even when expectations are modeled as homogeneous.²¹

²⁰Retrieved from [statline website](#).

²¹Average subjective expectations differ between the pre and post-reform samples. However, linear models show that within the post-reform sample the pension and tax reforms did not change average reported survival probabilities (estimates available on request). Hence, the shift in expectations was not induced by the policy reform.

In addition to its potential to better approximate the *level* of survival expectations, subjective information also allows one to capture *variation*. This variation is reflected in the posterior mean individual effect from the measurement model for subjective survival, conditional on covariates and reported probabilities. It is illustrated in the bottom panels of Figure 3. 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 (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 pre-reform period, panel d. illustrates that the same pattern holds for the post-reform regime. 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. Those counterfactual simulations are carried out at the mortality probabilities obtained from the post-reform sample.

Appendix C reports the estimates of the measurement model for subjective longevity and analyzes the variation in posterior means documented in panel c. of Figure 3. It shows that expectations as captured by the measurement model can easily be understood in terms of the underlying data and relate intuitively to background variables. Regarding interpretability, heterogeneity in expectations corresponds to variation in reported probabilities that cannot be explained by age, target age and health. Hence, heterogeneity in expectations can be related to the raw data in a transparent and intuitive way. Furthermore, this variation is correlated with variables that measure health and health behaviors. In the pre-reform period both smoking and drinking are associated with shorter subjective longevity, while drinking stands out in the post-reform period. A more persistent measure of health, averaging across all surveys in which an individual participated, also has predictive power in the pre-reform

period. Expectations do not proxy variation in preferences, since qualitative measures of risk aversion and patience do not correlate significantly with subjective longevity. 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 moments targeted in estimation. Removing such cohort differences is important, since estimation matches longitudinal moments simulated from initial conditions for a narrow birth year bracket to a short panel in which data at different ages correspond to different birth cohorts. If older respondents, for example, hold less wealth than younger ones due to historical variation in returns, the model would rationalize this difference in initial conditions as dissaving. In particular, we estimate the following linear FE models for wealth (shown), hours worked, participation and benefit claiming:²²

$$w_{it} = \sum_{j=1}^J [\beta_j age_{it}^j] + \beta_{J+1} n_{it} + \alpha_i + \varepsilon_{it} \quad (16)$$

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} :

$$\tilde{w}_{it} = \sum_{j=1}^J [b_j age_{it}^j] + b_{J+1} 2^{0.7} + a_i + \bar{a}_{1940} - \bar{a}_i + e_{it} \quad (17)$$

²²Simulations indicate that this procedure is reliable. It works best if cohort effects are relatively constant across ages. If the difference between cohorts varies with age, the adjustment is most accurate around the ages at which cohorts overlap. Simulations are available on request.

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 transformed in a similar way.

To ensure that this procedure does not remove differences between cohorts that are due to the policy reform, separate adjustments are applied to the pre- and post-reform samples. Hence, cohort differences are removed from samples in which all cohorts are subject to the same policy regime (all respondents are either born before or after January 1st 1950). Remaining differences between cohorts do not reflect the effect of the reform. Data for the post-reform 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 4. These are the moments described in Section 3. Panel a. of Figure 4 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 4 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

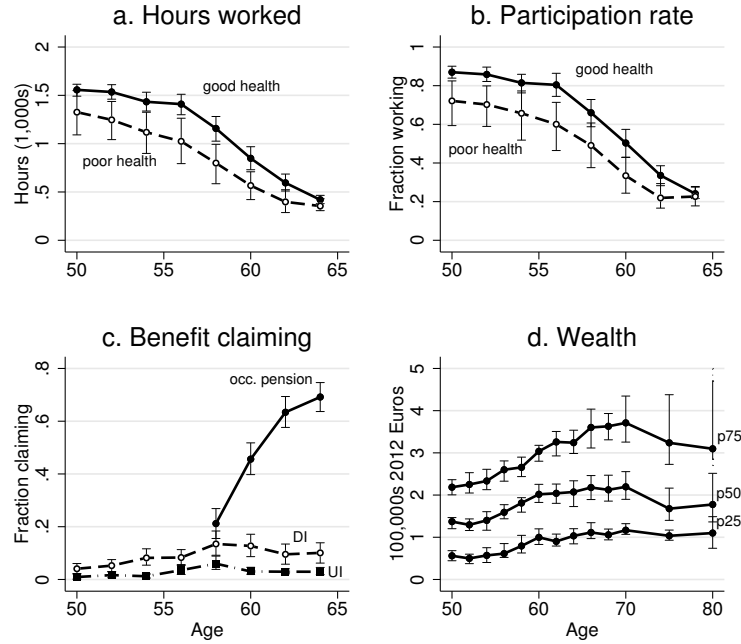


Figure 4: Moments targeted in estimation (averages for panels a.-c., quartiles for panel d.; 95% confidence intervals obtained by clustered bootstrap)

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 4 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.

Table 4: Initial conditions

	a. Pre-reform sample			b. Post-reform 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
# simulations		5,000			5,000	

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 post-reform than in the pre-reform 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 post-reform sample (43,000 compared to 53,000). Wealth, on the other hand, was higher in the post-reform 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 post-reform period, which is 7pp higher than before the reform. Unemployment insurance entitlements are 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.

5.1 Parameter estimates

Table 5 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 4.64, which is in the ballpark established by previous literature. The consumption weight of 0.68, in line with the estimates around 0.60 reported in French (2005), indicates that consumption is valued more highly than leisure. At 1.04 the estimated discount factor implies a high degree of patience.²³ Nonetheless, it is in line with previous work, in particular French (2005). Working any positive number of hours is associated with a fixed leisure cost of 1085 hours per year, which is in the range of estimates for the U.S. (French (2005)'s estimates are between 240 and 1313 hours per year). The estimated time cost of poor

²³Because 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.

Table 5: Estimates for preference parameters

	(1) Life tables	(2) Average subj. exp.	(3) Heterogeneous exp.
Utility function^a			
σ (concavity)	4.64 (0.000023)	4.64 (0.000028)	5.18 (0.000038)
κ (consumption weight)	0.68 (0.000061)	0.65 (0.000063)	0.42 (0.000065)
β (discount factor)	1.04 (0.000026)	1.04 (0.000035)	0.98 (0.000028)
ψ (fixed cost of work; hrs/yr)	1085 (27.5)	1013 (4.5)	1101 (8.0)
δ (time cost of bad health; hrs/yr)	131 (13.9)	249 (34.6)	251 (6.9)
ϕ (stigma cost DI; hrs/yr)	1681 (441.9)	2029 (345.5)	2140 (145.4)
ξ (stigma cost UI; hrs/yr)	3620 (8.9)	3509 (1.9)	3568 (5.9)
Bequests^b			
θ_0 (intercept bequest weight)	-2.45 (0.00014)	-3.91 (0.00015)	-9.27 (0.00010)
θ_1 (slope bequest weight)	4.73 (0.00012)	5.56 (0.00014)	2.51 (0.000082)
K (concavity bequests)	675,282 (29.0)	913,915 (34.3)	441,061 (6.0)
Function value	473.27	473.21	483.89

Standard errors in parentheses.

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

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

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

health is lower in the Netherlands at 131 hours per year compared with around 200 hours for the U.S.. As a result, Dutch workers will reduce labor supply less 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 sizable stigma costs of 1681 and 3620 hours of leisure respectively. These costs strongly discourage claiming. As for bequests, the estimates suggest that they are directed primarily 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 $\exp[-2.45 + 4.73 \times 2^{0.7}] \approx 188$ which declines to $\exp[-2.45 + 4.73 \times 1^{0.7}] \approx 10$ for a widower living alone. French (2005) estimates a constant bequest weight of 0.037, so bequests are much more important in the

Netherlands. Individuals face less financial risk in the Dutch institutional context, which means the model needs a strong bequest motive to rationalize limited dissaving even at advanced age. Moreover, the bequest curvature parameter is far from zero at 675,282 euro, 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. Such small standard errors are in line with previous work (e.g. Gustman et al., 1986; Palumbo, 1999; Cagetti, 2003; French, 2005; French and Jones, 2011). One explanation is that the approximation of the GMM objective function as quadratic around the minimum may be poor for non-linear models (French, 2005). Furthermore, the errors reported in Table 5 do not reflect estimation uncertainty in the auxiliary processes. Finally, the present research uses more variation to fit the models compared to other studies. Estimation uses moments that reflect wealth, labor supply and benefit claiming, whereas the literature mostly uses wealth and/or labor supply.

The middle column of Table 5 shows that the estimated concavity, consumption weight and discount factor are similar when expectations are fixed at the average level in the subjective data. The estimated time costs of work, poor health and UI benefit claiming change by many standard errors when we model survival expectations based on subjective data, but only for the cost of poor health is the change large relative to the estimates (it increases from 131 to 249 hours). The most striking impact is on the estimates for the bequest motive. In particular, the dependence of the bequest weight on household size becomes stronger, which makes bequests relatively less important at older ages. Moreover, K increases from 675,282 to 913,915 euro, reducing the curvature in bequest utility. The net effect of these changes in expectations and parameters on model fit is negligible: the criterion values are almost identical at 473 for both the specification with expectations modeled by life tables and average subjective expectations without heterogeneity.

The final column of Table 5 introduces variation in mortality expectations, acknowledging

that each individual has a unique view of his own longevity. Heterogeneity in expectations changes the estimates for most parameters substantially relative to estimation uncertainty (the only exceptions are the time costs of poor health and claiming DI). 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 5.18 which remains consistent with existing evidence. The consumption weight drops below fifty percent, but at 0.42 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 0.98. 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 models in two ways. Firstly, the weight attached to bequests is much reduced, declining from 0.006 for a man living with his spouse to 0.001 for a widower. Secondly, the K parameter that governs the curvature of the bequest utility function drops from values above 600,000 euro for models (1) and (2) to 441,061 euro for the model with variation in beliefs. The implication is that the curvature of bequest utility is stronger than for the other two models: the marginal utility of leaving a bequest is higher at low levels and then drops as more wealth is bequeathed. These changes in preference parameters in combination with heterogeneous expectations result in a model fit that is similar to the other two models: the function value is 484. The next section shows exactly how the fit differs between the three models.

5.2 Model fit

While values of the objective function summarize fit in one number, more can be learned by looking at the separate moments targeted in estimation. Figure 5 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 200 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: only at age 58-59 does it deviate 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 10pp or 14%. 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, 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

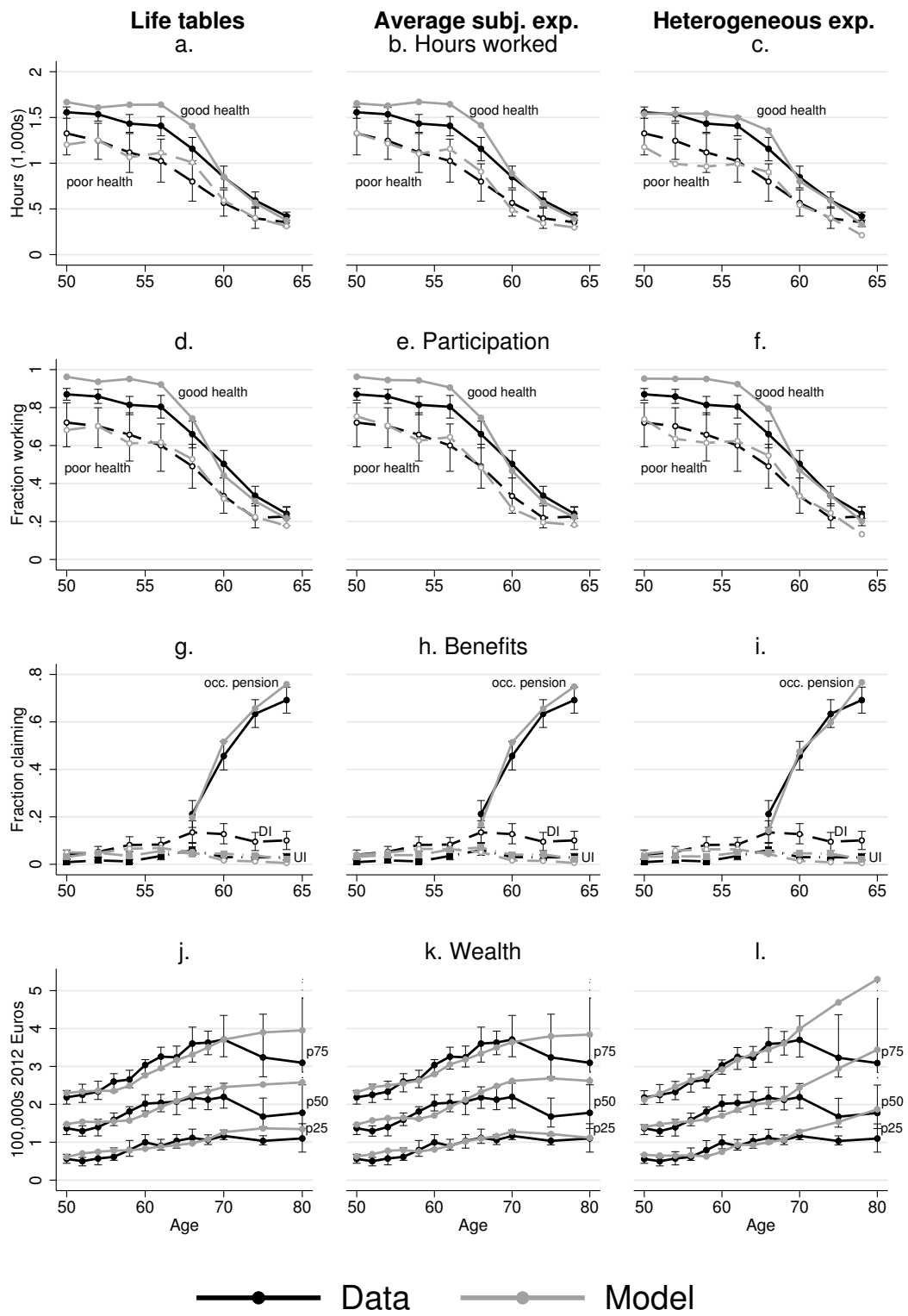


Figure 5: Model fit – target moments and simulations

age 74. As simulated wealth levels off 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 does not affect model fit. Figure 5 confirms that both models fit the data roughly equally well in all respects. Both models provide a better fit for labor supply in poor health than in good health. Panels g. and h. show that the models without variation in expectations are indistinguishable when it comes to claiming of pensions and benefits. Differences in terms of wealth are also small at all quartiles.

The model with heterogeneity in subjective expectations fits the data slightly worse than the other two models. This is the net result of a better fit for some moments and a worse fit for others. Panel c. in the third column of Figure 5 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 similar to the other models, but participation in poor health is not matched as well. Panel i. shows that there are no large changes in fit for pension claiming, or for disability and unemployment insurance. However, simulated wealth does look markedly different at ages 75-79 and 80-84. The simulated wealth quartiles for heterogeneous expectations continue to increase at those ages, while for the other two models wealth levels off. As a result, the third model fits the data less well in this respect. However, the deterioration in function value is small, 484 compared with 473, which reflects that the objective function places relatively little weight on wealth at advanced age. While the absolute discrepancy in panel l. is larger than for the other models, it remains limited when expressed relative to estimation uncertainty in the wealth moments.

Overall we find that all three models fit the data adequately and similarly. The variation between the objective function values shown in Table 5 reflects modest differences in fit. These differences are related to labor supply in good health prior to age 60, which all models overestimate to slightly different extents. 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. The model based on heterogeneous subjective expectations fits less well than the other two at those ages. However, wealth moments are estimated imprecisely at those age brackets. While there are statistically significant and economically meaningful differences between estimated preferences, the fit of all three models shows similar 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 pre-reform period (1993-2001) and the post-reform 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 6 reports the average retirement age 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

Table 6: Average retirement age with ('93-'01) and without ('06-'16) generous early retirement schemes

a. Data			
	Pre-reform ('93-'01)	Post-reform ('06-'16)	Difference (yrs)
Average pension age ^{a,b}	61.2 (60.97 – 61.41)	63.8 (63.49 – 64.03)	2.6 (2.23 – 2.92)
N	756	467	1223
b. Simulations			
	Pre-reform ('93-'01)	Post-reform ('06-'16)	Difference (yrs)
Life tables	60.7	65.9	5.2
Average subj. exp.	60.9	65.1	4.2
Heterogeneous exp.	61.0	63.7	2.7

^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.

^b Retirement ages below 50 or above 70 are set to missing; less than 1% of observations are dropped this way.
95% confidence intervals based on robust standard errors in parentheses.

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 post-reform period substantially once we account for cohort differences between individuals born in the 1950s and 1960s.²⁴ 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 6 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. These models predict an average retirement age in the pre-

²⁴Adding 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).

reform sample of 60.7 and 60.9 respectively. These averages are only just below the 95% confidence interval for the population mean and are 4-6 months below the sample average of 61.2. However, 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 first claimed at age 65.1-65.9 on average after the introduction of actuarial adjustments. These differences of 4-5 years relative to the pre-reform period are almost twice the 2.6 years observed in the data.

Allowing for variation in subjective mortality leads to a similar in-sample predicted retirement age as the other models: on average people are predicted to retire at 61.0. However, with heterogeneous expectations the model produces much more reasonable simulations out-of-sample. It predicts that the average retirement age would increase to 63.7, which is in line with the data. The resulting difference between estimation and simulation samples is 2.7 years, which is close to the observed difference of 2.6 years. Based on these simulations we conclude that while all three models fit the data similarly well within-sample, variation in expectations generates substantially better out-of-sample forecasts.

Figure 6 provides more detailed information regarding observed and simulated labor supply under the two regimes. The figure contains results regarding the average hours worked by two-year age bins (pooling individuals in good and poor health). Columns correspond 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 reiterate the discussion of model fit within the estimation sample: all models fit the data similarly well and over-estimate labor supply prior to age 60, though the model with heterogeneous expectations does slightly better than the other two. The dashed lines correspond to the new policy regime and indicate that the models with homogeneous 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 post-reform

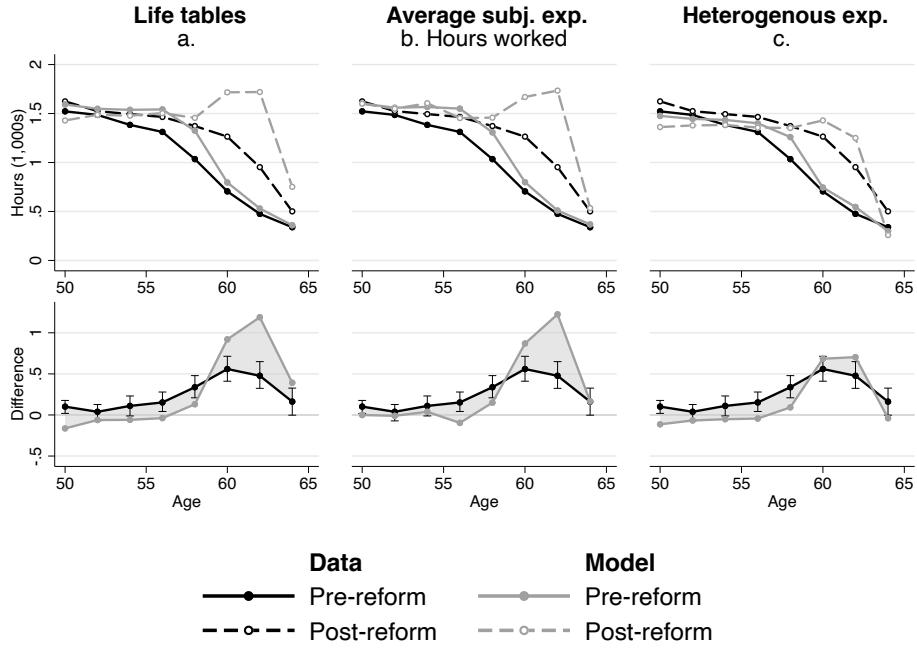


Figure 6: Simulated and observed labor supply in the pre-reform ('93-'01) and post-reform ('06-'16) samples

sample than before, these models produce differences of 800-1200 hrs/year. For the model based on life tables, the simulated difference between the regimes at ages 64-65 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 6.

Panel c. of Figure 6 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 less than half as large as that observed in panels a. and b. The bottom graph illustrates that the differences between both sets of simulations at the key ages 60-65 are much closer to the differences observed in the data than is the case for the other two models. While the results in Figure 6 combine adjustments to the extensive and intensive margins, similar patterns are obtained

for participation in the labor market (see Appendix F).

The simulations in Table 6 and Figure 6 provide a before/after comparison in which auxiliary processes (including survival), initial conditions, taxes and pensions are all changed to the post-reform situation. While this provides a relevant basis for external validation of the model, a policy maker running forecasts would not have access to post-reform data. Hence, it is interesting to decompose the total effect from Table 6 into the contributions of initial conditions and auxiliary processes on the one hand and tax and pension reform on the other. Appendix G provides such decomposition by changing one, and only one, aspect at a time from pre- to post-reform. It shows that all three models attribute the change in retirement to the reform of occupational pensions. Both models with homogeneous expectations predict an excessively strong behavioral response, forecasting an average retirement age around 65 after the pension reform, 13-14 months older than in the data. Changing only the institutional rules for occupational pensions, the model with heterogeneous expectations continues to predict an average retirement age of 63.7, which is off by only 1 month compared to the average age in the data. Hence, the model with heterogeneous expectations outperforms the other two with and without use of post-reform data.

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 homogeneous beliefs. This section shows how the interaction between preferences and expectations leads to better forecasts. We first analyze behavior conditional on expectations in the pre-reform sample and then shift attention to the post-reform period.

Pre-reform sample ('93-'01)

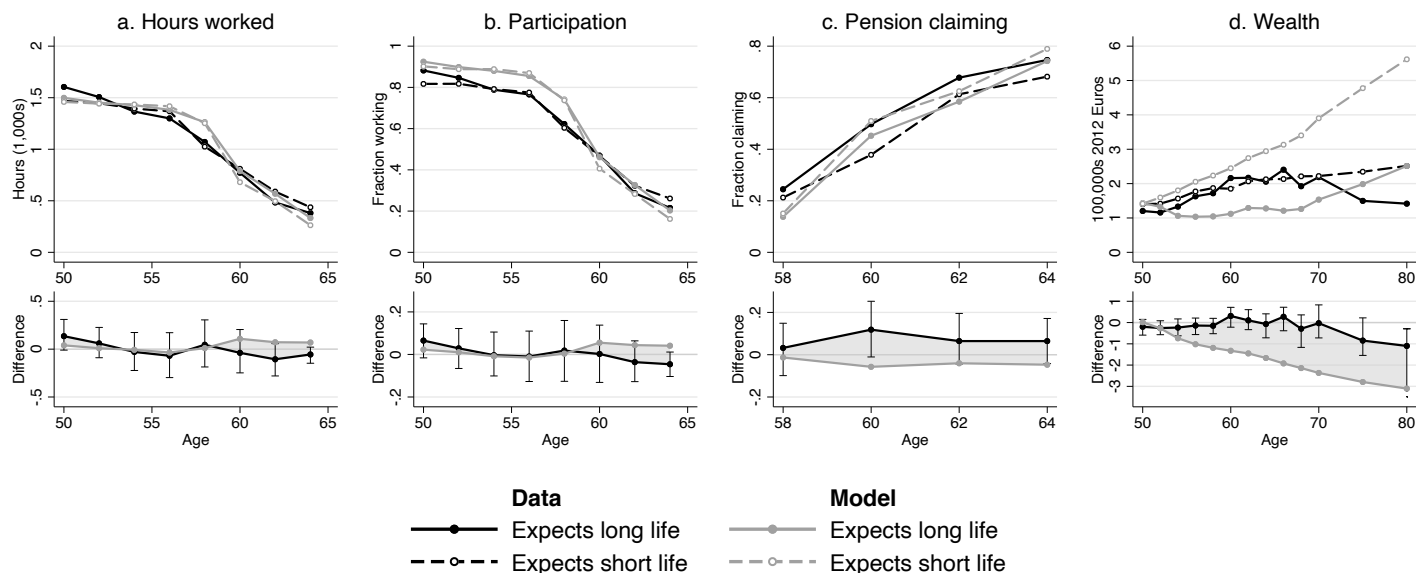


Figure 7: Behavior conditional on survival expectations for the pre-reform sample (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)

Preferences, expectations and behavior in the pre-reform period

Figure 7 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%).²⁵ Black lines correspond to data moments and grey lines represent model simulations. The data do not reveal large or statistically significant differences along survival expectations for hours worked, participation rates or pension claiming.²⁶ However, small sample sizes result in wide confidence intervals for the differences between

²⁵The 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. Grouping based on linear individual effects as explained in Appendix C results in very similar high and low survival groups (92% of respondents are allocated to the same group).

²⁶The absence of substantial variation in labor supply and pension claiming across the distribution of subjective longevity is not an artifact of data processing, since it is confirmed for raw claiming and labor supply (estimates available on request).

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 does yield substantial covariation with life expectancy: 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 300,000 euro. This divergence is driven by the bequest motive, which is the primary driver of saving in the model. The Dutch institutional setting with automatic annuitization and generous social security limits the importance of precautionary and life-cycle considerations. Bequests are more important for those who expect a short life for two reasons. Firstly, they are expected to happen sooner and hence receive more weight in the decision process for those individuals who expect to die shortly. This is the primary reason to hold more wealth relative to those for whom death, and bequests, is a more distant prospect. Secondly, the importance of bequests is a function of household size, which means they are relatively more important at younger ages. This motivates those with short life expectancy not only to accumulate more wealth, but also to start that accumulation early. 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 200,000 euro. Thus, while the model over-estimates differences in wealth holdings, it is broadly in line with the qualitative pattern observed at the oldest ages.

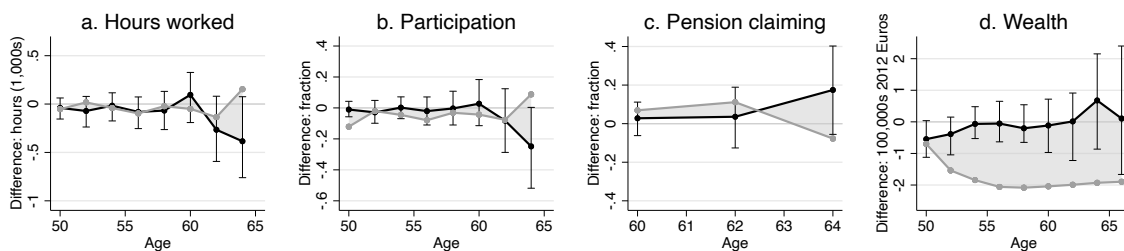
In order to understand why preference estimates change when heterogeneity in expectations is introduced, Appendix H compares simulations based on the model with heterogeneous

expectations for two sets of preference parameters: the estimates obtained for average subjective expectations and those for heterogeneous subjective expectations. Among the moments matched in estimation, wealth moments drive the change in preference estimates when variation in expectations is introduced around the subjective average (see Figure H1). Simulated labor supply and benefit claiming are similar for both sets of preferences. However, the model cannot rationalize positive wealth holdings after age 60 for at least one quarter of the sample at the preference estimates obtained for average subjective expectations. Figure H2 indicates that this depletion of wealth reflects variation in expectations: median wealth for the top 45% of the distribution of subjective survival approaches zero by age 62. Furthermore, contrary to the data, the model produces larger differences in labor supply and pension claiming along the distribution of survival expectations. Comparing model simulations at these two sets of estimates illustrates how preferences and expectations interact. In particular, the fact that in the data wealth does vary with survival but labor supply and pension claiming do not is itself valuable information regarding preferences.

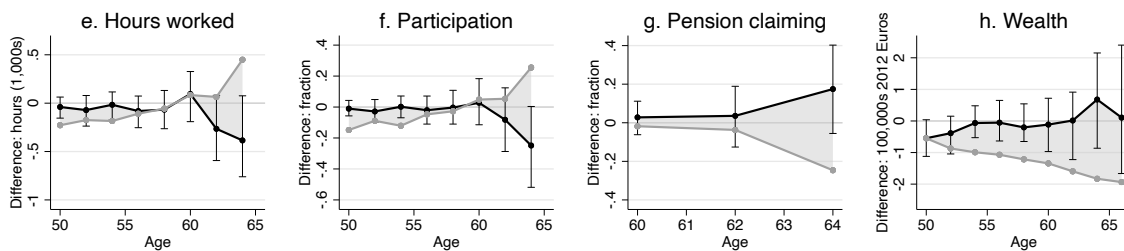
Preferences, expectations and behavior in the post-reform period

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 post-reform sample (2006-2016). Figure 8 displays differences in behavior between those who expect to live long and those who anticipate an earlier demise. At the preference estimates for average subjective expectations, the variation in simulated labor supply and pension claiming across the distribution of expectations is more pronounced for this regime, in which occupational pensions are subject to actuarial adjustments. However, at the preference estimates for heterogeneous expectations both labor supply and pension claiming vary little with survival. Estimating preferences under heterogeneous expectations thus results in behavior that is closer to the data for the counterfactual policy regime, both overall and

Preferences: heterogenous expectations



Preferences: average subjective expectations



—●— **Data: long life - short life** —●— **Model: long life - short life**

Figure 8: Differences in behavior across the distribution of survival expectations for the post-reform sample (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)

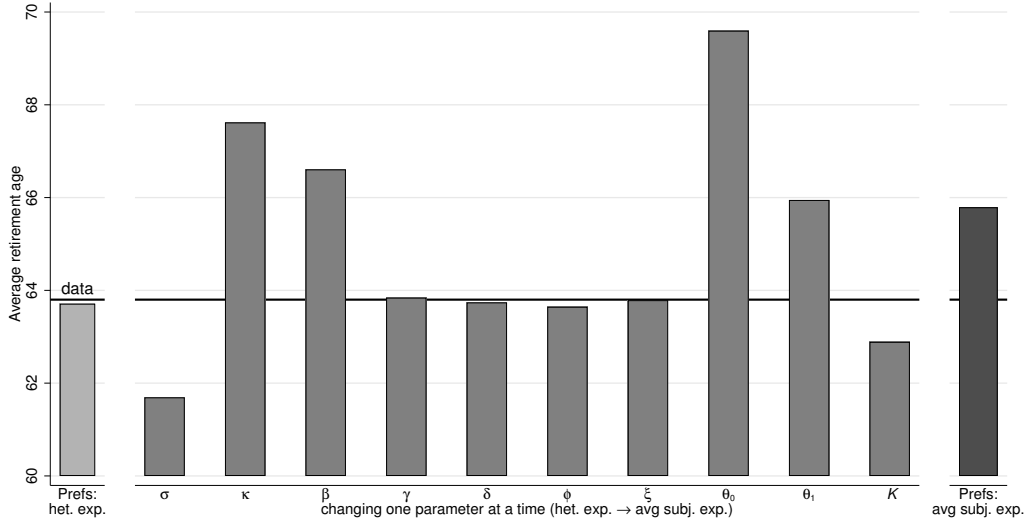


Figure 9: The impact of changing preference parameters on counterfactual simulations for the model with heterogeneous expectations

conditional on beliefs.

Figure 9 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 63.7 (compared with 63.8 in the data, see Table 6). 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 65.8, which is older than that based on the model with homogeneous expectations evaluated at the same parameters. The middle panel of Figure 9 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 both parameters in the bequest weight (θ_0 and θ_1) all have large effects on the average retirement age. These effects are intuitive: the intertemporal elasticity of substitution $1/\sigma$ is higher in the model with homogeneous expectations, which makes agents less willing to work longer to increase consumption growth. This generates a smaller response in the retirement age when early retirement is made less attractive. The estimated consumption weight κ is also higher for homogeneous expectations, which makes agents value consumption over leisure and hence leads people to retire later. Raising the discount factor or making bequests more important have the same effect of delayed retirement. 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 heterogeneous 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 preference 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 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. An analysis of the behavior implied by the models shows that all three 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. This discrepancy in wealth is larger for the model with heterogeneous survival. 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 4-5 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 2.7 year increase to an average of 63.7 (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 as it adds a single state variable to the model.

References

- Arcidiacono, P., Muralidharan, K., Shim, E.-S., and Singleton, J. D. (2016). Valuing school choice: using a randomized experiment to validate welfare evaluation of private school vouchers. Working paper.
- 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.
- Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business & Economic Statistics*, 21(3):339–353.
- CBS (2017). Statistics Netherlands (CBS) Statline. <http://statline.cbs.nl>. Accessed: 2017-01-29.
- De Bresser, J. (2019). Measuring subjective survival expectations—do response scales matter? *Journal of Economic Behavior & Organization*, 165:136–156.
- 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 Jong, P. R. and Aarts, J. M. (1992). *Economic aspects of disability behavior*. Elsevier.

- De Nardi, M., Fella, G., and Paz-Pardo, G. (2020). Nonlinear household earnings dynamics, self-insurance, and welfare. *Journal of the European Economic Association*, 18(2):890–926.
- 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.
- Deelen, A. and Euwals, R. (2014). Do wages continue increasing at older ages? evidence on the wage cushion in the netherlands. *De Economist*, 162(4):433–460.
- Delavande, A. (2008). Pill, patch, or shot? subjective expectations and birth control choice. *International Economic Review*, 49(3):999–1042.
- Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.
- 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.

- Galiani, S., Murphy, A., and Pantano, J. (2015). Estimating neighborhood choice models: Lessons from a housing assistance experiment. *American Economic Review*, 105(11):3385–3415.
- 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 Zimmer, 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. Z., Myrseth, K. O. R., and Schoenle, R. S. (2019). Yolo: Mortality beliefs and household finance puzzles. *The Journal of Finance*, 74(6):2957–2996.
- 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.
- Keane, M. and Moffitt, R. (1998). A structural model of multiple welfare program participation and labor supply. *International economic review*, pages 553–589.
- Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, 156(1):3–20.
- Keane, M. P., Todd, P. E., and Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In *Handbook of labor economics*, volume 4, pages 331–461. Elsevier.

- Keane, M. P. and Wolpin, K. I. (2007). Exploring the usefulness of a nonrandom holdout sample for model validation: Welfare effects on female behavior. *International Economic Review*, 48(4):1351–1378.
- 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.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.
- 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.
- Lise, J., Seitz, S., and Smith, J. (2005). Equilibrium policy experiments and the evaluation of social programs. Mimeo, Queens University.
- 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.
- Lumsdaine, R. L., Stock, J. H., and Wise, D. A. (1992). Three models of retirement: computational complexity versus predictive validity. In *Topics in the Economics of Aging*, pages 21–60. University of Chicago Press.
- 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.

- Moffitt, R. (1983). An economic model of welfare stigma. *The American Economic Review*, 73(5):1023–1035.
- 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.
- O’Dea, C. and Sturrock, D. (2020). Survival pessimism and the demand for annuities. Technical report, National Bureau of Economic Research.
- OECD (1999). Taxing wages. Technical report, OECD.
- OECD (2004). Taxing wages. Technical report, OECD.
- Palumbo, M. G. (1999). Uncertain medical expenses and precautionary saving near the end of the life cycle. *The Review of Economic Studies*, 66(2):395–421.
- Pantano, J. and Zheng, Y. (2013). Using subjective expectations data to allow for unobserved heterogeneity in hotz-miller estimation strategies. Mimeo University of Arizona.
- 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.

- Scholz, J. K., Seshadri, A., and Khitatrakun, S. (2006). Are americans saving ?optimally? for retirement? *Journal of political economy*, 114(4):607–643.
- Schorfheide, F. and Wolpin, K. I. (2012). On the use of holdout samples for model selection. *American Economic Review*, 102(3):477–81.
- Schorfheide, F. and Wolpin, K. I. (2016). To hold out or not to hold out. *Research in Economics*, 70(2):332–345.
- Shapira, G. (2017). How subjective beliefs about hiv infection affect life-cycle fertility evidence from rural malawi. *Journal of Human Resources*, 52(3):680–718.
- 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.
- Ter Weel, B. (2003). The structure of wages in the netherlands, 1986–98. *Labour*, 17(3):361–382.
- Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American economic review*, 96(5):1384–1417.
- Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge.
- Trommel, W. and De Vroom, B. (1994). The Netherlands: The Loreley-effect of early exit. In Naschold, F. and De Vroom, B., editors, *Regulating Employment and Welfare: Company and National Policies of Labour Force Participation at the End of Worklife in Industrial Countries*, chapter 3, pages 51–115. De Gruyter.

- 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.
- Wise, D. A. (1985). A behavioral model versus experimentation: The effects of housing subsidies on rent. *Methods of Operations Research*, 50:441–89.
- 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 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. In our model, agents automatically receive UI if they are eligible based on age and working history and choose not to work or claim DI.

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 the pre-reform 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:

$$occpens_t = 0.0175 \times yrswrk_t \times (prevearn_t - 19,000) \quad (18)$$

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

The tax function in the model is based on the 1999 edition of *Taxing Wages*, a publication by the OECD (OECD, 1999). As can be seen in Figure A1, 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 for 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 B gives the exact functional form for the tax function.

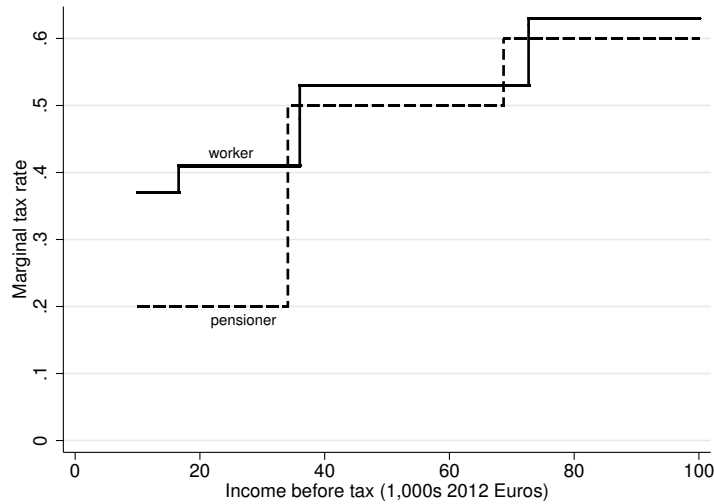


Figure A1: Marginal tax rates for workers and pensioners (estimation sample: '93-'01)

Policy reform: taxes and pensions

The Dutch income tax system was revised completely on January 1st 2001. The model for the new tax regime is based on *Taxing Wages 2004* (OECD, 2004). Figure A2 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 B.

While marginal tax rates changed markedly in 2001, the reforms of occupational pensions effective from the first of January 2006 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

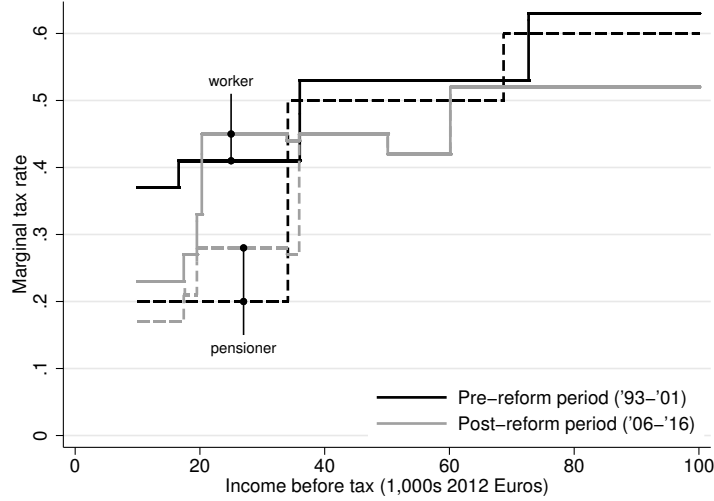


Figure A2: Marginal tax rates for the pre and post-reform periods

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:

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

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 pre-reform sample, with early retirement option, to men born prior to 1950 observed in the 1993-2001 survey waves. The post-reform 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.

B Tax functions

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

Pre-reform period: 1993-2001

We take the tax function for the 1990s from the OECD report *Taxing Wages 1999* (OECD, 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:

$$netinc_t = \tau(earn_t, DIinc_t, UIinc_t, pubpens_t, occpens_t, w_t, t) \quad (20)$$

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):

$$workcredit_t = \begin{cases} 0 & \text{if } earn_t = 0 \\ 154 & \text{if } 0 < earn_t < 1,283 \\ 0.12 \times earn_t & \text{if } 1,283 \leq earn_t < 15,825 \\ 0.12 \times 15,825 & \text{if } earn_t \geq 15,825 \end{cases} \quad (21)$$

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

$$UIcontr_t = \begin{cases} 0 & \text{if } earn_t < 16,588 \\ 0.0645 \times (earn_t - 16,588) & \text{if } 16,588 \leq earn_t < 744,357 \\ 0.0645 \times 744,357 & \text{if } earn_t \geq 744,357 \end{cases} \quad (22)$$

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

$$pubmed_t = \begin{cases} 0.012 \times earn_t + 132 \times n_t^{1/0.7} & \text{if } earn_t < 32,539 \\ 0.012 \times 32,539 + 132 \times n_t^{1/0.7} & \text{if } 32,539 \leq earn_t < 37,890 \\ 0.012 \times 32,539 & \text{if } earn_t \geq 37,890 \end{cases} \quad (23)$$

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

$$taxable_t = earn_t + DIinc_t + UIinc_t + pubpens_t + occpens_t \quad (24)$$

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

$$taxable_t = \max\{0, taxable_t - 5,266 - workcredit_t - UIcontr_t - pubmed_t\} \quad (25)$$

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:

$$scontr_t = \begin{cases} 0.275 \times taxable_t & \text{if } taxable_t < 28,835 \text{ \& } t < 65 \\ 0.275 \times 28,835 & \text{if } taxable_t \geq 28,835 \text{ \& } t < 65 \\ 0.11 \times taxable_t & \text{if } taxable_t < 28,835 \text{ \& } t \geq 65 \\ 0.11 \times 28,835 & \text{if } taxable_t \geq 28,835 \text{ \& } t \geq 65 \end{cases} \quad (26)$$

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

$$inctax_t = \begin{cases} 0.0885 \times taxable_t & \text{if } taxable_t < 28,835 \\ 0.0885 \times 28,835 \\ \quad + 0.5 \times (taxable_t - 28,835) & \text{if } 28,835 \leq taxable_t < 63,419 \\ 0.0885 \times 28,835 \\ \quad + 0.5 \times (63,419 - 28,835) \\ \quad + 0.6 \times (taxable_t - 63,419) & \text{if } taxable_t \geq 63,419 \end{cases} \quad (27)$$

Wealth w_t is taxed at a rate of 0.7%:

$$wtax_t = 0.007 \times w_t \quad (28)$$

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

$$grossinc_t = earn_t + DIinc_t + UIinc_t + pubpens_t + occpens_t + capinc_t \quad (29)$$

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:

$$netinc_t = grossinc_t - UIcontr_t - pubmed_t - scontr_t - inctax_t - wtax_t \quad (30)$$

Post-reform period: 2006-2016

The Dutch tax system was revised drastically on January 1st 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* (OECD, 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:

$$netinc_t = \tau(earn_t, DIinc_t, UIinc_t, pubpens_t, occpens_t, w_t, t) \quad (31)$$

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):

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

Unemployment insurance contributions $UIcontr_t$:

$$UIcontr_t = \begin{cases} 0 & \text{if } grossinc_t < 17,409 \\ 0.058 \times (grossinc_t - 17,409) & \text{if } 17,409 \leq grossinc_t < 50,115 \\ 0.058 \times (50,115 - 17,409) & \text{if } grossinc_t \geq 50,115 \end{cases} \quad (33)$$

Contributions to public health insurance $pubmed_t$:

$$pubmed_t = \begin{cases} 0.0125 \times grossinc_t + 446 & \text{if } grossinc_t < 33,917 \\ 0.0125 \times 33,917 + 446 & \text{if } 33,917 \leq grossinc_t < 37,490 \\ 0 & \text{if } grossinc_t \geq 37,490 \end{cases} \quad (34)$$

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

$$taxable_t = \max \{0, grossinc_t - UIcontr_t - pubmed_t\} \quad (35)$$

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

$$scontr_t = \begin{cases} 0.324 \times taxable_t & \text{if } taxable < 33,974 \text{ \& } t < 65 \\ 0.324 \times 33,974 & \text{if } taxable \geq 33,974 \text{ \& } t < 65 \\ 0.145 \times taxable_t & \text{if } taxable < 33,974 \text{ \& } t \geq 65 \\ 0.145 \times 33,974 & \text{if } taxable \geq 33,974 \text{ \& } t \geq 65 \end{cases} \quad (36)$$

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

$$inctax_t = \begin{cases} 0.01 \times taxable_t & \text{if } taxable < 18,705 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (taxable_t - 18,705) & \text{if } 18,705 \leq taxable_t < 33,974 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (33,974 - 18,705) \\ \quad + 0.42 \times (taxable_t - 33,974) & \text{if } 33,974 \leq taxable_t < 58,250 \\ 0.01 \times 18,705 \\ \quad + 0.0795 \times (33,974 - 18,705) \\ \quad + 0.42 \times (58,250 - 33,974) \\ \quad + 0.52 \times (taxable_t - 58,250) & \text{if } taxable_t \geq 58,250 \end{cases} \quad (37)$$

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:

$$workcredit_t = \begin{cases} 0.0175 \times earn_t & \text{if } earn_t < 9,298 \\ 0.0175 \times 9,298 & \text{if } 9,298 \leq earn_t < 9,316 \\ 0.0175 \times 9,298 \\ \quad + 0.112 \times (earn_t - 9,316) & \text{if } 9,316 \leq earn_t < 19,583 \\ 0.0175 \times 9,298 \\ \quad + 0.112 \times (19,583 - 9,316) & \text{if } earn_t \geq 19,583 \end{cases} \quad (38)$$

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

$$inctax_t = \max \{0, inctax_t - 2,099 - workcredit_t\} \quad (39)$$

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:

$$wtax_t = \begin{cases} 0 & \text{if } w_t < 44,280 \\ 0.012 \times (w_t - 44,280) & \text{if } w_t \geq 44,280 \end{cases} \quad (40)$$

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

$$netinc_t = grossinc_t + capinc_t - UIcontr_t - pubmed_t - scontr_t - inctax_t - wtax_t \quad (41)$$

C Measurement model for subjective expectations

Likelihood

In the model expectations follow Gompertz distributions, which implies that the true subjective probability of surviving to target age ta_k conditional on having survived to current age t is given by:

$$S_{isk}(ta_k, t) = \Pr(T \geq ta_k | T \geq t) = \frac{\exp\left(-\frac{\gamma_{is}}{\alpha} (\exp(\alpha (ta_k/100)) - 1)\right)}{\exp\left(-\frac{\gamma_{is}}{\alpha} (\exp(\alpha (t/100)) - 1)\right)} \quad (42)$$

for individual i in survey wave s and target age k . Heterogeneity in survival expectations is captured by γ_{is} , which depends on observed variables \mathbf{x}_{is} as well as unobserved individual (ζ_i) and survey-wave effects (η_{is}): $\gamma_{is} = \exp(\mathbf{x}'_{is}\boldsymbol{\beta}_1 + \zeta_i + \eta_{is})$.

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

$$P_{isk}^* = S_{isk} + \varepsilon_{isk} \quad (43)$$

where $\varepsilon_{isk} \sim \mathcal{N}(0, \sigma_{is}^2)$, independent of all covariates and across target ages, survey waves and individuals.

Latent probabilities P_{isk}^* are censored between zero and 100 and rounded prior to being reported. The 11-point answer scale allows for rounding to multiples of 100, 50 and 10. The rounding model is ordinal:

$$R_{isk} = r \iff \mu_{r-1} \leq y_{is}^* = \zeta_i^r + \eta_{is}^r + \varepsilon_{isk}^r < \mu_r \quad (44)$$

where $r \in \{10, 50, 100\}$, $\mu_0 = -\infty$ and $\mu_{100} = \infty$. The rounding equation includes individual and question sequence effects, allowing rounding to be correlated across repeated observa-

tions 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 tilts survival curves. We assume that the idiosyncratic rounding shocks ε_{isk}^r follow a standard normal distribution and are independent from covariates and all other errors, so the conditional probabilities of each category of rounding $\Pr(R_{isk} = r | \zeta_i^r, \eta_{is}^r)$ take the shape of an ordered probit, e.g. $\Pr(R_{isk} = 50 | \zeta_i^r, \eta_{is}^r) = \Phi(\mu_{50} - \zeta_i^r - \eta_{is}^r) - \Phi(\mu_{10} - \zeta_i^r - \eta_{is}^r)$ where $\Phi(\cdot)$ denotes the standard normal CDF.

A reported probability in combination with a particular level of rounding implies an interval for the perturbed probability $P_{isk}^* \in [LB_{isk}^r, UB_{isk}^r]$. For instance, a reported probability of 50% that is rounded to a multiple of 10 yields the interval $P_{isk}^* \in [45, 55]$. The conditional probability of that event given observed and unobserved heterogeneity, so given S_{isk} , is easy to calculate, since $P_{isk}^* \sim \mathcal{N}(S_{isk}, \sigma_{is}^2)$: $\Pr(45 \leq P_{isk}^* < 55) = \Phi\left(\frac{55 - S_{isk}}{\sigma_{is}}\right) - \Phi\left(\frac{45 - S_{isk}}{\sigma_{is}}\right)$.

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_{isk} the set Ω_{isk} that consists of all types of rounding that are consistent with that probability. $\Omega_{isk} = \{10\}$ for all reported probabilities $P_{isk} \in \{1, 2, 3, 4, 6, 7, 8, 9\}$; $\Omega_{isk} = \{10, 50\}$ for $P_{isk} = 5$; and $\Omega_{isk} = \{10, 50, 100\}$ for $P_{isk} \in \{0, 10\}$. We obtain the conditional density as:

$$f(P_{isk} | \mathbf{x}_{is}, \zeta_i, \eta_{is}) = \sum_{r \in \Omega_{isk}} \Pr(R_{isk} = r | \zeta_i^r, \eta_{is}^r) \times \Pr(LB_{isk}^r \leq P_{isk}^* < UB_{isk}^r | \mathbf{x}_{is}, \zeta_i, \eta_{is}) \quad (45)$$

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

$$\Pr(LB_{isk}^r \leq P_{isk}^* < UB_{isk}^r | \cdot) = \begin{cases} \Pr(LB_{isk}^r \leq P_{isk}^* | \mathbf{x}_{is}, \zeta_i, \eta_{is}); & \text{if } P_{isk} \geq P_{is,k-1} - 0.5r \\ LB_{isk}^r = P_{isk} - 0.5r & \\ \Pr(LB_{isk}^r \leq P_{isk}^* < UB_{isk}^r | \mathbf{x}_{is}, \zeta_i, \eta_{is}); & \text{if } 0.5r \leq P_{isk} \\ LB_{isk}^r = P_{isk} - 0.5r & \text{and } P_{isk} < P_{is,k-1} - 0.5r \\ UB_{isk}^r = P_{isk} + 0.5r & \\ \Pr(P_{isk}^* < UB_{isk}^r | \mathbf{x}_{is}, \zeta_i, \eta_{is}); & \text{if } P_{isk} < 0.5r \\ UB_{isk}^r = 0.5r & \end{cases} \quad (46)$$

where $r \in \{10, 50, 100\}$ as before. The first case occurs when P_{isk} 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 censored from below at zero. All probabilities in equation 46 are calculated from univariate normal distributions and are therefore easy to obtain. As an example, consider two reported probabilities equal to 60% and 50%. The former can only be rounded to a multiple of 10, so $f(60|.) = \Pr(R = 10|.) \times \Pr(55 \leq P_1^* < 65|.)$. The likelihood takes into account that the second probability can never be larger than the first:

$$f(50|.) = \Pr(R = 10|.) \times \Pr(45 \leq P_2^* < 55|.) + \Pr(R = 50|.) \times \Pr(P_2^* \geq 25|.) \quad (47)$$

Whether a given probability is censored or not depends on the degree of rounding and on the preceding reported probability.

The extent of rounding is identified from the relative frequencies of reported probabilities in the three groups that reflect different possible degrees of rounding. If most reported

Table C1: Estimates for the model of subjective expectations

	a. Pre-reform sample (1993-2001)				b. Post-reform 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

probabilities are in $\{1, 2, 3, 4, 6, 7, 8, 9\}$, without disproportionate bunching at 0, 5 or 10, that indicates that probabilities are rounded to multiples of 10. If instead there is substantial bunching at 0, 5 and 10, that suggests rounding to multiples of 50 is prevalent. Similarly, many 0s and 10s relative to the frequency of 5s point towards even coarser rounding to multiples of 100.

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 error components are modeled 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).

Estimates

Model estimates are reported in Table C1. People in bad health have a higher hazard of death than those in good health: their hazard is 36% higher in the pre-reform sample and 24% higher in the post-reform 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% and 99% of probabilities in the pre- and post-reform samples respectively 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 is are obtained by the following algorithm (more information can be found in Train, 2003):

1. Draw 500 pairs of vectors $\zeta_i^d = (\zeta_i^d, \zeta_i^{r,d})$ and $\eta_{is}^d = (\eta_{is}^d, \eta_{is}^{r,d})$ from their bivariate normal distributions ($d = 1, \dots, 500$).
2. For each pair of vectors d calculate the conditional likelihood contribution $\mathcal{L}_i^d(\mathbf{P}_i | \zeta_i^d, \eta_i^d; \mathbf{x}_i, \mathbf{b})$ 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 $\mathcal{L}_{is}^d(\mathbf{P}_{is} | \zeta_i^d, \eta_{is}^d; \mathbf{x}_{is}, \mathbf{b})$: the likelihood contribution of probability sequence is . Note that the likelihood for individual i is the product of

the likelihoods for all his sequences: $\mathcal{L}_i^d = \prod_{s=1}^{S_i} \mathcal{L}_{is}^d$. 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 $w_i^d = \frac{\mathcal{L}_i^d}{\sum_{d=1}^{500} \mathcal{L}_i^d}$. For the sequence effects we calculate weights for each question sequence ($w_{is}^d = \frac{\mathcal{L}_{is}^d}{\sum_{d=1}^{500} \mathcal{L}_{is}^d}$).
4. Simulate the posterior mean individual effect as $\tilde{\boldsymbol{\zeta}}_i = (\tilde{\zeta}_i, \tilde{\zeta}_i^r) = \sum_{d=1}^{500} w_i^d \boldsymbol{\zeta}_i^d$. The mean of the posterior distribution of the sequence effect is $\tilde{\boldsymbol{\eta}}_{is} = (\tilde{\eta}_{is}, \tilde{\eta}_{is}^r) = \sum_{d=1}^{500} w_{is}^d \boldsymbol{\eta}_{is}^d$.

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

$$\tilde{\gamma}_{is} = \exp\left(\mathbf{x}'_{is} \mathbf{b}_1 + \tilde{\zeta}_i + \tilde{\eta}_{is}\right) \quad (48)$$

where $\tilde{\zeta}_i$ is the posterior mean for ζ_i and $\tilde{\eta}_{is}$ is the mean for η_{is} . We leave covariates \mathbf{x}_{is} at the levels observed in the data, except for health which we set to ‘good’ for everybody when calculating the expectations index $\tilde{\gamma}_{is}$. 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

$$\tilde{\gamma}_i = \exp\left(\mathbf{x}'_{is} \mathbf{b}_1 + \tilde{\zeta}_i + \tilde{\eta}_i\right) \quad (49)$$

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}_{is} is constant over time for each individual.

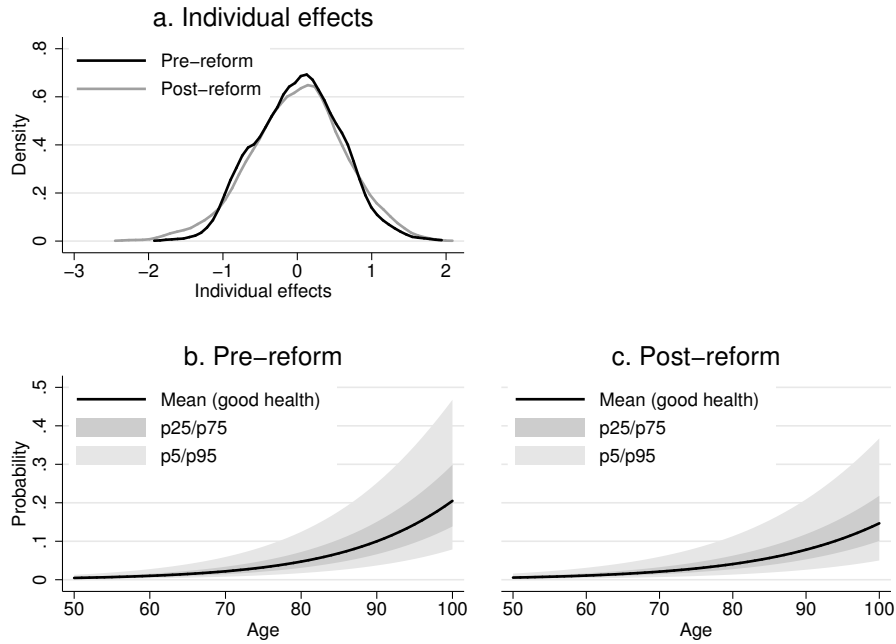


Figure C1: Heterogeneous expectations: distributions of individual effects and corresponding variation in survival probabilities

Figure C1 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 post-reform 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.

Variation in posterior means

The complex nature of the measurement model and the importance of heterogeneity in expectations for the retirement model underline the need to understand what drives the variation

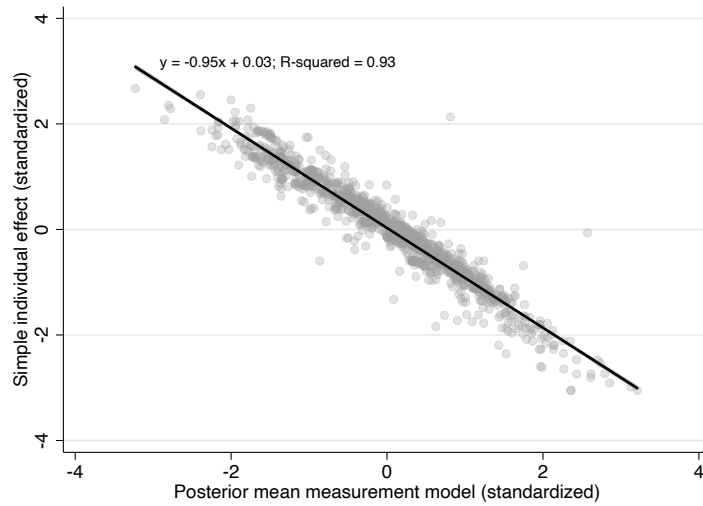


Figure C2: Subjective longevity: posterior means and a more direct measure of subjective survival

in expectations captured by the posterior means. In order to aid such transparency and interpretability, Figure C2 relates the standardized posterior means to a simpler measure of subjective expectations for the pre-reform sample. This simpler measure is the estimated individual effect from a linear Random Effects model that explains reported probabilities by cubic polynomials in current and target age and a dummy for being in poor health. Like the posterior means from the measurement model, these individual effects reflect subjective longevity controlling for current and target age as well as current health. The correlation between the two measures is close to perfect and the posterior means explain 93% of variation in linear individual effects. The negative sign of the relationship reflects the fact that higher posterior means imply a higher hazard of death and hence shorter life, while higher survival probabilities relate positively to longevity. Figure C2 indicates that the heterogeneity in expectations analyzed in this paper reflects variation in reported probabilities that cannot be explained by age, target age or current health.

While it is reassuring to find that the individual means relate to the raw data in an intuitive way, more can be learned by looking at the covariates of expectations. Table C2

shows estimates for linear models that regress posterior means on a set of regressors proposed in the literature (e.g. in Hurd, 2009). For the pre-reform period, there is some variation in expectations across birth cohorts with those born after 1939 expecting a longer life than those born prior to 1930. Education, income and wealth do not predict subjective longevity significantly.²⁷ Health and health behaviors do stand out as relevant predictors. In the pre-reform period, both smoking and drinking are associated with lower longevity. In the post-reform period smoking is no longer significant, but excessive alcohol consumption, more than 4 alcoholic beverages per day, remains predictive of a shorter expected life. Health, the individual average of the binary measure, strongly predicts longevity in the pre-reform period but does not enter significantly for the post-reform period. The reason why *average* health correlates with longevity in the pre-reform sample even though the measurement model controls for *current* health is that the measurement model is estimated based on probabilities reported during 1993-2001 and average health uses all observations for a given individual (also those after the pension and tax reforms). The significant association reflects the fact that those who report lower probabilities of survival during 1993-2001 tend to end up in worse health during subsequent survey waves. For the post-reform sample, on the other hand, more survey waves are used in the estimation of the measurement model. As a result, the association between expectations and average health is much attenuated.

Models 2 and 4 in Table C2 add controls for qualitative measures of preferences. Risk aversion is captured by the individual average agreement on a 7-point scale with the statement “I’m willing to take the risk that I will lose money when there is also a possibility to make money” (reverse-coded so stronger agreement means more risk aversion). Patience is measured as an individual-specific index based on agreement with nine statements that measure time preferences, such as “I am willing to sacrifice my current welfare in order to reach

²⁷Education does not vary over time, since the sample is restricted to individuals aged 50 and older. Income and wealth dummies, in 2012 euros, are defined based on the individual mean across all waves in which that respondent participated.

Table C2: Linear models explaining individual-specific survival expectations (posterior means)

	Dependent variable: individual-specific hazard of death ^a			
	Pre-reform period		Post-reform period	
	(1)	(2)	(3)	(4)
Coh. 1930-1934	-0.0867 (0.0626)	-0.0768 (0.0707)		
Coh. 1935-1939	-0.0888 (0.0618)	-0.0625 (0.0709)		
Coh. 1940-1944	-0.174*** (0.0612)	-0.121* (0.0718)		
Coh. 1944-1949	-0.244*** (0.0659)	-0.203*** (0.0753)		
Coh. 1955-1959			-0.0525 (0.102)	-0.0658 (0.105)
Coh. 1960-1966			-0.0553 (0.105)	-0.0745 (0.109)
Educ. middle	-0.00415 (0.0566)	0.00349 (0.0661)	-0.0443 (0.107)	-0.0752 (0.107)
Educ. high	0.0329 (0.0544)	0.0277 (0.0636)	-0.0233 (0.107)	-0.0789 (0.107)
Net inc. 20-30,000	0.123* (0.0674)	0.112 (0.0763)	-0.168 (0.125)	-0.165 (0.128)
Net inc. 30-40,000	0.106 (0.0700)	0.114 (0.0804)	-0.120 (0.132)	-0.107 (0.135)
Net inc. \geq 40,000	0.0936 (0.0734)	0.102 (0.0836)	-0.0697 (0.137)	-0.0850 (0.143)
Wealth 25-120,000	0.0933 (0.0662)	0.126 (0.0787)	0.104 (0.149)	0.0929 (0.152)
Wealth 120-250,000	0.122* (0.0704)	0.169** (0.0829)	0.141 (0.153)	0.129 (0.155)
Wealth \geq 250,000	0.0691 (0.0766)	0.112 (0.0888)	0.242 (0.154)	0.231 (0.158)
Ever smoker	0.127*** (0.0453)	0.126** (0.0513)	0.0710 (0.0951)	0.0987 (0.0951)
Ever alcoholic	0.123** (0.0479)	0.122** (0.0532)	0.282** (0.135)	0.309** (0.135)
Avg. health	-0.326*** (0.0573)	-0.380*** (0.0664)	-0.163 (0.117)	-0.162 (0.117)
Risk aversion		0.0108 (0.0190)		-0.0352 (0.0299)
Patience		0.0107 (0.0434)		0.0203 (0.0865)
Constant	-4.236*** (0.112)	-4.275*** (0.176)	-3.729*** (0.242)	-3.508*** (0.315)
Observations	1,016	798	293	286
R-squared	0.068	0.075	0.057	0.070

^a Survival as captured by the posterior mean $\tilde{\gamma}_i$ in equation 20 in the paper. A higher value of the posterior mean corresponds to a higher hazard of death and hence *lower* life expectancy.

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

certain results in the future". The patience index is constructed according to the method proposed in Kling et al. (2007). Neither risk aversion nor patience correlate significantly with survival expectations.

D Objective function and standard errors

Having estimated the auxiliary processes for survival, health, medical expenses, wages, income of the spouse and equivalence scales, one can use the model described in section 2.1 to simulate labor supply, benefit claiming and wealth accumulation. Preferences are estimated by minimizing the difference between simulated behavior and data:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{1 + \frac{N_{\text{ind}}}{N_{\text{sim}}}} \bar{\mathbf{g}}(\mathbf{Z}, \boldsymbol{\theta})' \widehat{\mathbf{W}} \bar{\mathbf{g}}(\mathbf{Z}, \boldsymbol{\theta}) \quad (50)$$

where $\boldsymbol{\theta}$ is the vector of 10 preference parameters; $\bar{\mathbf{g}}(\mathbf{Z}, \boldsymbol{\theta})$ is a vector of 91 simulated moments that are functions of data \mathbf{Z} and preferences; and $\widehat{\mathbf{W}}$ is a (diagonal) weighting matrix. N_{ind} is the number of individuals in the sample from which moments are computed and N_{sim} is the number of individuals for whom behavior is simulated, which equals 5,000.

For hours worked at age t and health $M_t = m$ the elements of $\bar{\mathbf{g}}(\cdot)$ are

$$\bar{g}_{m,t}^{\text{work}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_{m,t}} hrs_{is} - \check{hrs}_{m,t}(\boldsymbol{\theta}) \quad (51)$$

$\mathcal{N}_{m,t}$ is the set of person-year observations for age bin t who have health m , hrs_{is} is the observed hours worked in the data for individual i in survey wave s who is in the set $\mathcal{N}_{m,t}$, and $\check{hrs}_{m,t}(\boldsymbol{\theta})$ is the average simulated hours worked at health m and age t . N is the total number of observations on which the moment would be based if data were available from all respondents that appear in at least one moment. Since labor supply moments are based on two-year age bins, this hypothetical total is $N = 2 \times N_{\text{ind}}$, where N_{ind} is the number of individuals in the sample. Dividing by this hypothetical number of observations shrinks moments for which there are few observations to zero, thereby correcting for missing data (French, 2005).

The moments for participation in the labor market given age t and health m are similar:

$$\bar{g}_{m,t}^{\text{part}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_{m,t}} \text{part}_{is} - \check{\text{part}}_{m,t}(\boldsymbol{\theta}) \quad (52)$$

where part_{is} is one if individual i works in survey wave s according to self-reported labor market status and zero otherwise and $\check{\text{part}}_{m,t}(\boldsymbol{\theta})$ is the simulated rate of positive labor supply at age t and health m .

DI, UI and occupational pension claiming are captured by moments that do not condition on health:

$$\bar{g}_t^{\text{claim}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_t} \text{claim}_{is} - \check{\text{claim}}_t(\boldsymbol{\theta}) \quad (53)$$

Here claim_{is} is a claiming indicator for a certain kind of benefits and $\check{\text{claim}}_t(\boldsymbol{\theta})$ is the simulated rate of benefit claiming at age t . As before, \mathcal{N}_t is the set of person-year observations in the data for age bin t and N is the hypothetical sample size equal to twice the number of individuals in the sample.

The final set of moments matches wealth quartiles. These are defined as

$$\bar{g}_t^{\text{p25}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_t} \mathbb{I}\{w_{is} \leq \check{w}_t^{\text{p25}}(\boldsymbol{\theta})\} \times 0.75 + \mathbb{I}\{w_{is} > \check{w}_t^{\text{p25}}(\boldsymbol{\theta})\} \times -0.25 \quad (54)$$

$$\bar{g}_t^{\text{p50}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_t} \mathbb{I}\{w_{is} \leq \check{w}_t^{\text{p50}}(\boldsymbol{\theta})\} \times 0.50 + \mathbb{I}\{w_{is} > \check{w}_t^{\text{p50}}(\boldsymbol{\theta})\} \times -0.50 \quad (55)$$

$$\bar{g}_t^{\text{p75}}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i,s \in \mathcal{N}_t} \mathbb{I}\{w_{is} \leq \check{w}_t^{\text{p75}}(\boldsymbol{\theta})\} \times 0.25 + \mathbb{I}\{w_{is} > \check{w}_t^{\text{p75}}(\boldsymbol{\theta})\} \times -0.75 \quad (56)$$

where $\mathbb{I}\{w_{is} \leq \check{w}_t^q(\boldsymbol{\theta})\}$ is an indicator equal to one if the wealth of observation i, s in age bin t is smaller than or equal to the q th percentile of simulated wealth at age t and zero otherwise. As mentioned in the text, age bins have a width of two years for the range 50-69

and five years afterwards ($N = 2 \times N_{\text{ind}}$ for $t < 70$ and $N = 5 \times N_{\text{ind}}$ for $t \geq 70$).

The weight matrix $\widehat{\mathbf{W}}$ is the inverse of a diagonal matrix that contains the estimated variances of the data, $\widehat{\mathbf{W}} = \widehat{\mathbf{S}}^{-1}$:

$$\widehat{\mathbf{S}} = \begin{pmatrix} \widehat{\text{var}}_{2,50} [hrs_{is}] & 0 & 0 & \dots & 0 \\ 0 & \widehat{\text{var}}_{2,52} [hrs_{is}] & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \widehat{\text{var}}_{80} [w_{is}^{p75}] \end{pmatrix} \quad (57)$$

where w_{is}^{p75} is equal to 0.25 if wealth is smaller than or equal to the 75th percentile for a given age bin and equal to -0.75 otherwise (analogously to equation 56). The variances are adjusted for missing data by division by the hypothetical sample size N :

$$\widehat{\text{var}}_{2,50} [hrs_{is}] = \frac{1}{N} \sum_{i,s \in \mathcal{N}_{2,50}} (hrs_{is} - \bar{hrs}_{2,50})^2 \quad (58)$$

$\bar{hrs}_{2,50}$ is the sample average number of hours worked by people in good health, $M_{50} = 2$, in age bin 50-51.

The MSM estimator is asymptotically normal: $\sqrt{N} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. We use the following estimator for the asymptotic covariance matrix:

$$\widehat{\boldsymbol{\Sigma}} = \left(1 + \frac{N_{\text{ind}}}{N_{\text{sim}}}\right) (\widehat{\mathbf{D}}' \widehat{\mathbf{W}} \widehat{\mathbf{D}})^{-1} \widehat{\mathbf{D}}' \widehat{\mathbf{W}} \widehat{\mathbf{V}} \widehat{\mathbf{W}} \widehat{\mathbf{D}} (\widehat{\mathbf{D}}' \widehat{\mathbf{W}} \widehat{\mathbf{D}})^{-1} \quad (59)$$

where $\widehat{\mathbf{D}}$ is the matrix of first derivatives of moments evaluated at the estimator: $\widehat{\mathbf{D}} = \frac{\partial \bar{\mathbf{g}}(\mathbf{Z}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\widehat{\boldsymbol{\theta}}}$. This matrix is approximated by numerical differentiation using two points around $\widehat{\boldsymbol{\theta}}$. The matrix $\widehat{\mathbf{V}}$ is the empirical covariance matrix of the moments evaluated at the estimated preferences $\mathbf{g}(\mathbf{z}_{is}, \widehat{\boldsymbol{\theta}})$. $\widehat{\mathbf{V}}$ is approximated by means of a clustered bootstrap, in which preferences are fixed at $\widehat{\boldsymbol{\theta}}$ and $\bar{\mathbf{g}}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}})$ is calculated for 500 bootstrap samples b of

N_{ind} individuals drawn with replacement. For any combination of moments $m1$ and $m2$ the corresponding element of $\widehat{\mathbf{V}}$ is

$$\widehat{\mathbf{V}}_{m1,m2} = \frac{N_{m1,m2}}{500 - 1} \sum_{b=1}^{500} \left(\bar{g}_{m1}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}}) - \bar{g}_{m1} \right) \left(\bar{g}_{m2}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}}) - \bar{g}_{m2} \right) \quad (60)$$

$\bar{g}_{m1}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}})$ and $\bar{g}_{m2}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}})$ are sample-average moments across the bootstrap sample \mathbf{Z}_b . \bar{g}_{m1} and \bar{g}_{m2} are the averages of those sample moments across the 500 bootstrap samples. Note that the elements of $\widehat{\mathbf{V}}$ correspond to (co-)variances of $\mathbf{g}(\mathbf{z}_{is}, \widehat{\boldsymbol{\theta}})$, which are computed from the covariances of *average moments* $\bar{\mathbf{g}}(\mathbf{Z}_b, \widehat{\boldsymbol{\theta}})$ by multiplication with $N_{m1,m2}$, the hypothetical sample size for the covariance between $m1$ and $m2$. In line with the discussion above, $N_{m1,m2} = 2 \times N_{\text{ind}}$ if at least one moment refers to an age bin younger than 70 and $N_{m1,m2} = 5 \times N_{\text{ind}}$ otherwise.

E Auxiliary processes

In addition to survival we estimate processes for health, average equivalence scales by age, medical expenses, wages and the exogenous income of the spouse. The estimates are shown in Figure E1. Black lines refer to the pre-reform sample and grey lines to the post-reform 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 pre- and post-reform periods 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 post-reform 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 post-reform sample is extrapolated from observations of men aged 50-66. Hence, the steady decline observed for the post-reform sample may be an artifact of extrapolation of the downward pattern between ages 50 and 60 also observed for the estimation sample.

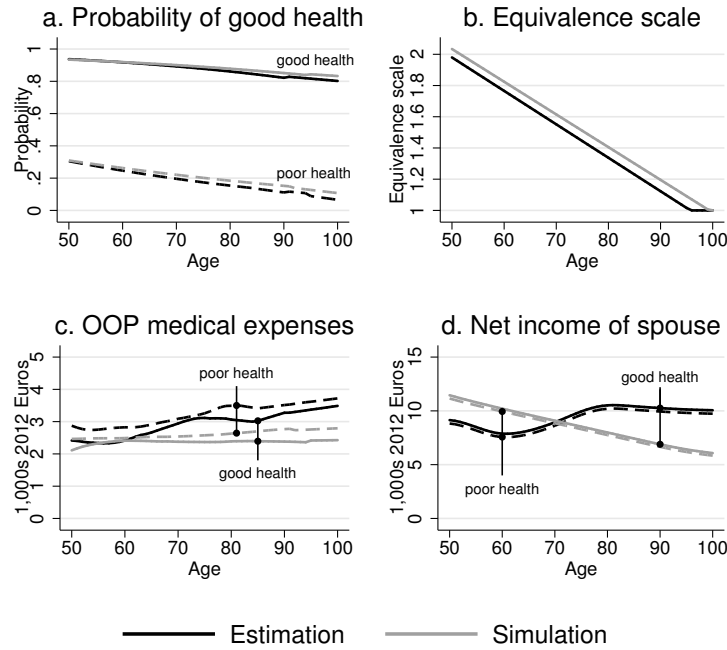


Figure E1: Auxiliary processes (estimated from DHS dataset)

However, robustness analysis indicates that none of the results for the new regime 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. We estimate the auto-regressive coefficient at 0.95 and the standard deviation of the

Table E1: Wage processes

	Pre-reform ('93-'01)	Post-reform ('06-'16)
Auto-correlation	0.971*** (0.0454)	0.989*** (0.00408)
Error standard deviation	0.00418 (0.00689)	0.00516* (0.00301)
N	3,087	2,480

Standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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

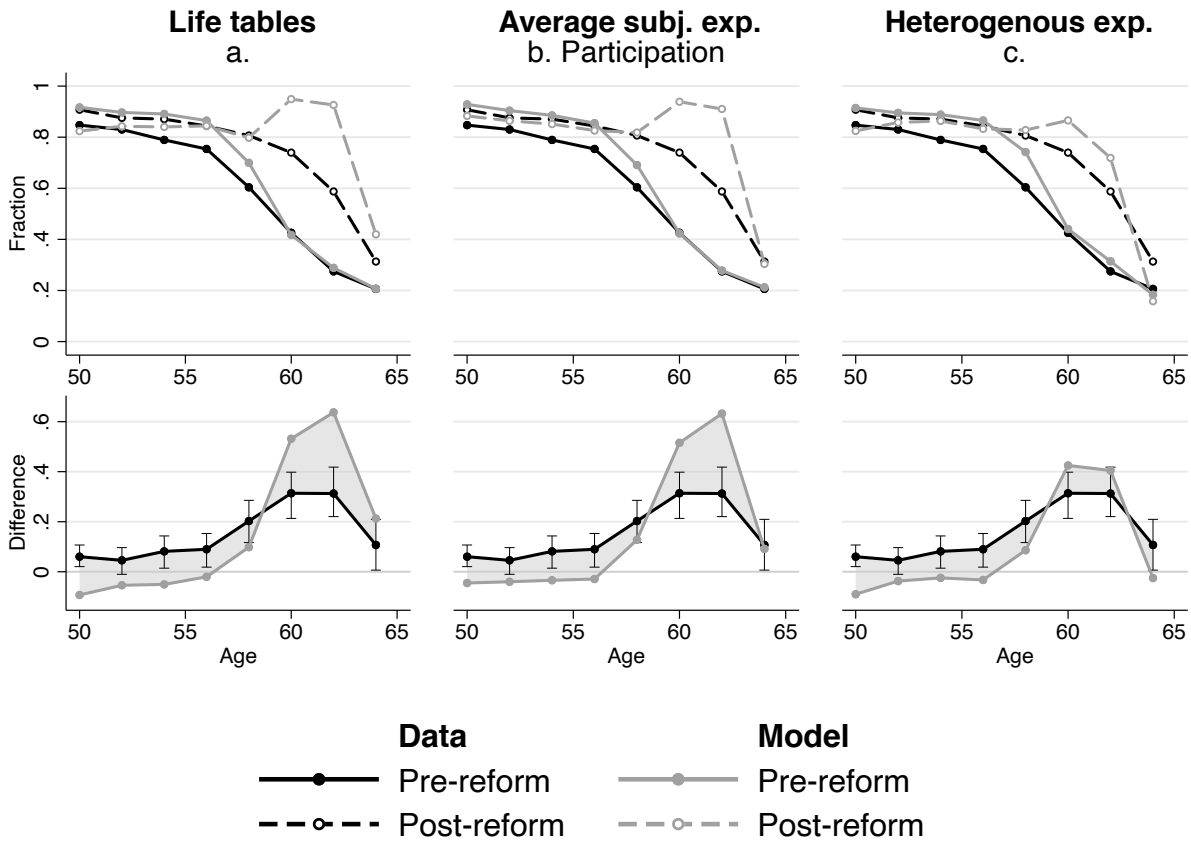


Figure F1: Simulated and observed labor supply in the pre-reform ('93-'01) and post-reform ('06-'16) samples

G Decomposition of simulated difference between pre- and post-reform periods

Table G1: Decomposition of difference in average retirement age between pre- and post-reform simulations

a. Data^{a,b}			
	Pre-reform ('93-'01)	Post-reform ('06-'16)	Difference (yrs)
Average pension age	61.2 (60.97 – 61.41)	63.8 (63.49 – 64.03)	2.6 (2.23 – 2.92)
N	756	467	1223
b. Simulations^c			
	Life tables	Average subj. exp.	Heterogeneous exp.
All at pre-reform	60.7	60.9	61.0
Initial conditions and survival	61.8	61.0	61.1
All aux. except survival	60.2	60.3	60.7
Taxes	61.0	61.5	61.4
Occ. pensions	65.0	64.9	63.7
All at post-reform	65.9	65.1	63.7

^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.

^b Retirement ages below 50 or above 70 are set to missing; less than 1% of observations are dropped this way.

^c These changes are not cumulative: only the one process or institution mentioned in the leftmost column is set to post-reform situation, the rest is at pre-reform values.

95% confidence intervals based on robust standard errors in parentheses.

H Model predictions for different preferences (pre-reform sample: '93-'01)

Targeted moments for model with heterogeneous expectations
 Top: heterogeneous expectations; Bottom: average subj. expectations

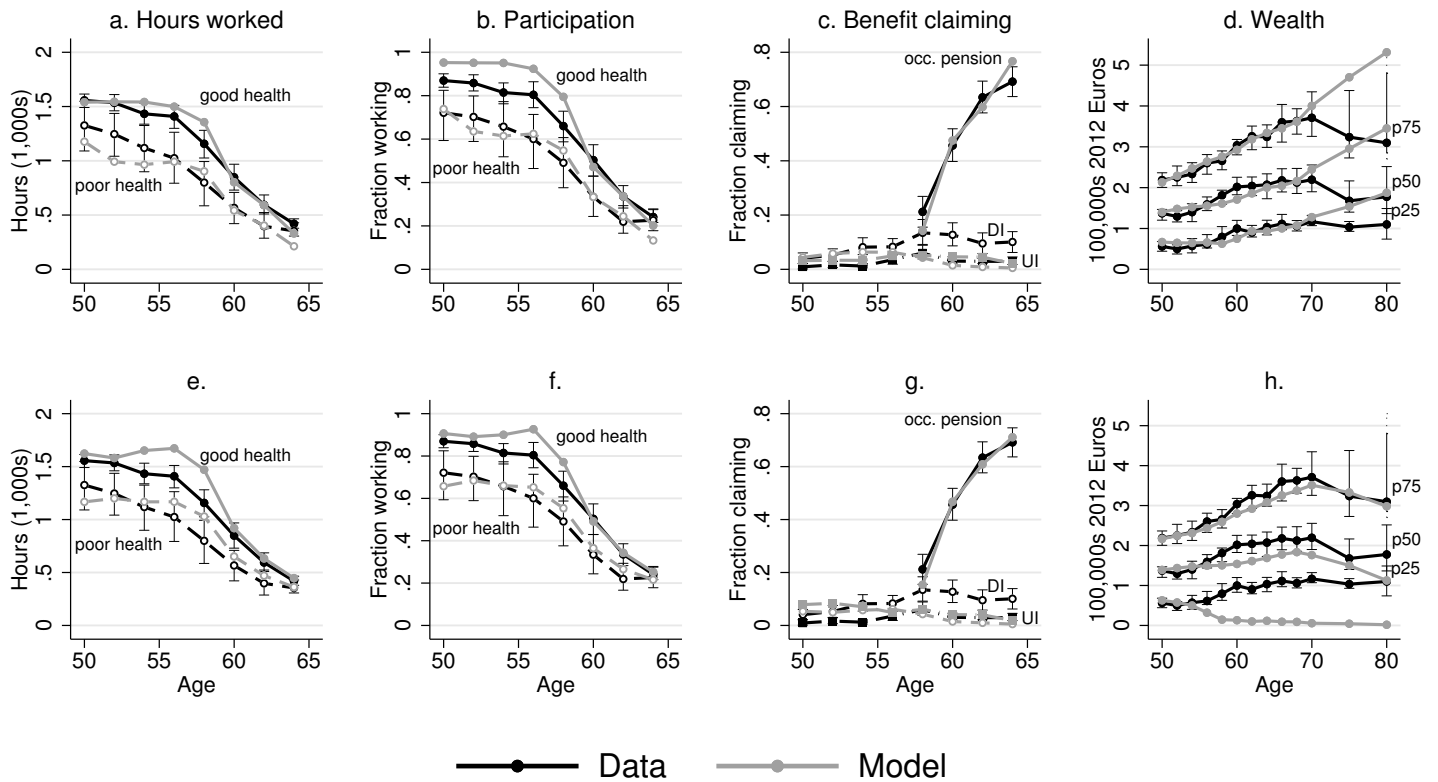
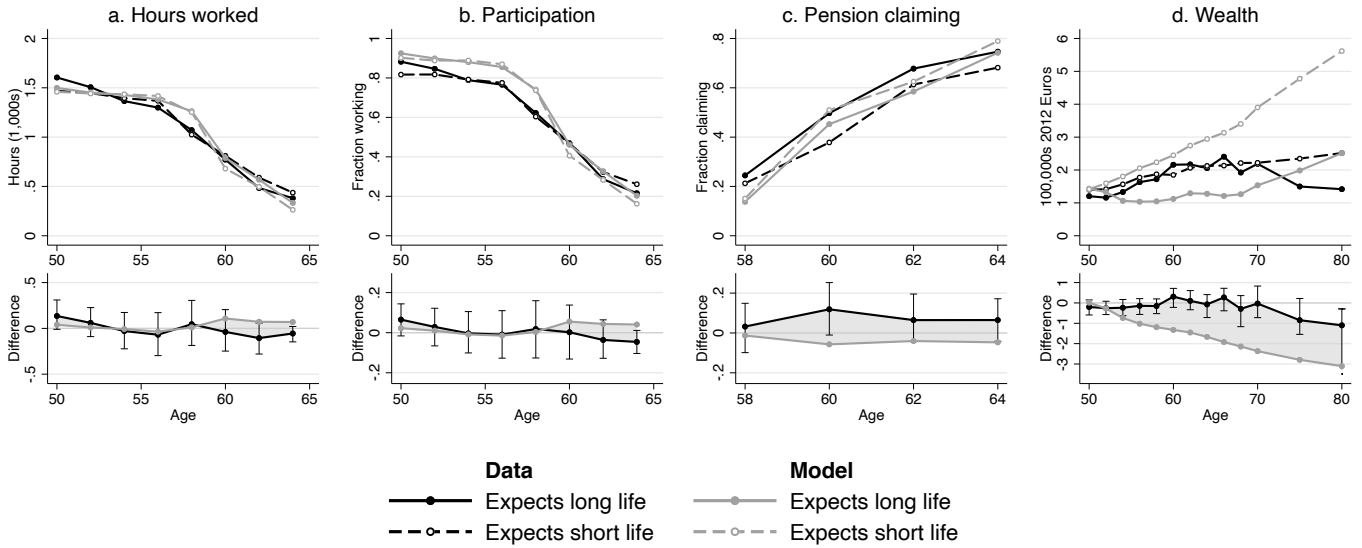


Figure H1: Model fit for model with heterogeneity in survival expectations evaluated at preference estimates for that model (top panels) and at preference estimates for model with survival fixed at average subjective expectations (bottom panels)

Preferences: estimates for heterogeneous expectations



Preferences: estimates for average subj. expectations

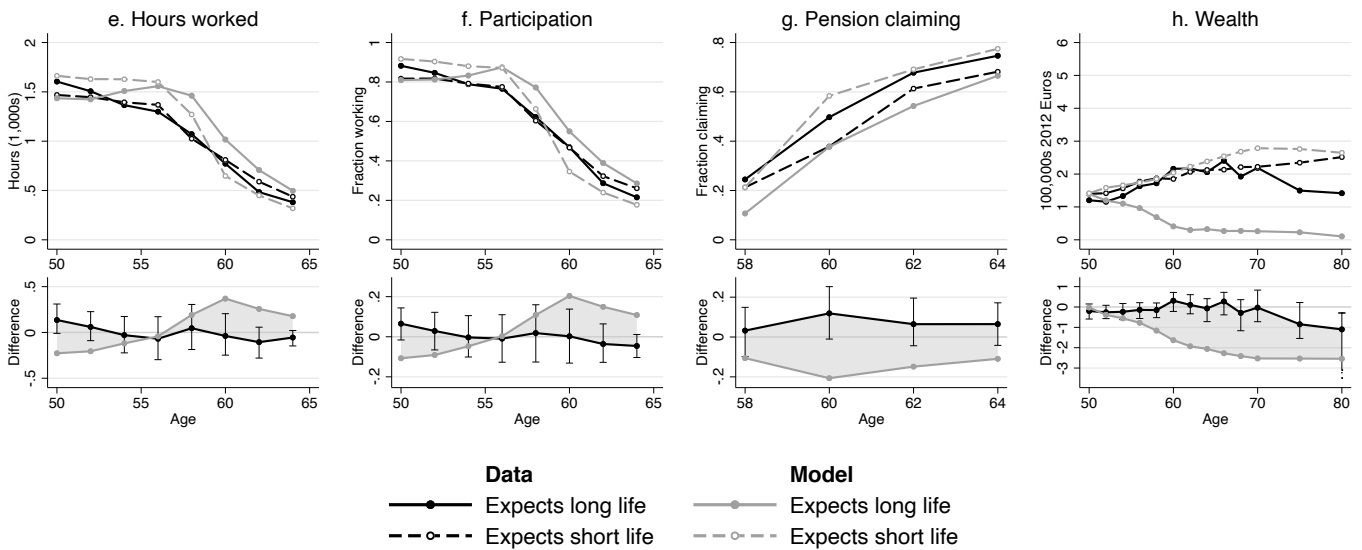


Figure H2: Behavior conditional on survival expectations for the *pre-reform* sample for different sets of preference estimates (“long life” means top 45% of distribution of subjective survival, “short life” means bottom 45%)