The Impact of Measurement Error in Health on Health-Related Counterfactuals*

Martín García-Vázquez† Luis Pérez†

This version: August 15, 2022

PRELIMINARY AND INCOMPLETE

Abstract

Health is typically imperfectly measured. How important is this imperfect observability to evaluate the costs of bad health? We estimate a dynamic, structural life-cycle model of savings and labor supply with health risk under two assumptions on the observability of health. The first one, which is prevalent in much of the literature, is that health is perfectly observable. The second one is that, while health is not observable, a battery of noisy measures of health is available to the researcher. We find that ignoring measurement error in health leads to substantially underestimating both the persistence of health and the time costs of being unhealthy. Ultimately, measurement error has an effect on the estimated lifetime costs of bad health—as measured by labor earnings, hours worked, consumption, and assets—leading to underestimate these by as much as 300%. A key message of our paper is that estimating the lifetime costs of bad health using structural economic models requires researchers to worry about measurement error in health.

*We thank our advisors V. V. Chari, Mariacristina De Nardi, Loukas Karabarbounis, Jeremy Lise, and Joseph Mullins for useful feedback and stimulating discussions. We are also grateful to Cormac O'Dea for providing us with simulations from the model in his job-market paper. We are also thankful to Jacob Adenbaum and Zhixiu Yu for helpful comments. Errors are our own.

†Affiliation: University of Minnesota. Emails: garca001@umn.edu and perez766@umn.edu.

1 Introduction

An increasing body of research recognizes the importance of health in shaping economic decisions and outcomes.[1,2] Special emphasis has been placed in understanding how health affects labor-supply decisions, savings, retirement, and inequality in consumption, income, and wealth.[3] The vast majority of studies, however, treat health as perfectly observed. In this paper, we ask whether failing to account for measurement error in health leads structural models to estimate substantially-biased costs of bad health.

To answer this question, we estimate a structural, dynamic life-cycle model of savings and labor supply under two assumptions on the observability of health. The first one is that health is perfectly observable from survey data. The second one is that, while health is not observable, a battery of noisy measures of health is available to the researcher.

We adopt a canonical model of labor supply and saving behavior which includes health- and labor-productivity risk. As in many other papers, health is exogenous and affects pecuniary resources, the time endowment, and future health. Individuals choose how much to work, consume and save. The government taxes income, gives mean-tested transfers, and provides social security. Because bad-health shocks are more prevalent at old ages, and because there are excellent longitudinal data sets starting at age fifty, so does our model.

To evaluate whether it is important to explicitly acknowledge that health is measured with error, we estimate our structural model twice. First, by ignoring the measurement error, and then by explicitly modeling it. When health is measured with error, we

[1]A non-exhaustive list of papers is: Bound (1991), Smith Jr (1993), Smith (1999), Wu (2003), French (2005), De Nardi, French and Jones (2010), French and Jones (2011), Gustman and Steinmeier (2014), Capatina (2015), Poterba, Venti and Wise (2017), De Nardi, Pashchenko and Porapakkarm (2018), Bueren (2021), Costa-Dias, Blundell, Britton and French (2021), and Amengual, Bueren and Crego (2021).

[2]It is now well understood that health not only drives life expectancy, but also has the potential to affect the productivity of workers, the non-pecuniary costs of work, medical expenditures, and long-term needs, among others.

[3]In the reduced-form literature, Poterba et al. (2017), Smith Jr (1993), Smith (1999) and Wu (2003) examine the nexus between adverse health shocks and wealth accumulation; Bound (1991) and Costa-Dias et al. (2021) study the labor-supply effects of adverse health conditions. On the structural side, papers like Gustman and Steinmeier (2014), Capatina (2015), and De Nardi, Pashchenko and Porapakkarm (2018) have looked at the impact of health shocks on different economic outcomes.

compute health transitions using self-reported health status. To correct for measurement error in health, we assume that health follows a non-stationary hidden Markov model (NSHMM). Estimating the NSHMM for health yields the dynamics of true latent health and the relationship between observed health measures and true health. Estimating the structural model under the two assumptions on the observability of health produces two sets of estimated costs of bad health, as measured by labor earnings, hours worked, consumption, and assets. We find that ignoring measurement error in health leads to considerably underestimating the costs of bad health, especially for people who are already unhealthy at age fifty. More specifically, our exercise suggests that not taking into account measurement error could lead to underestimating the lifetime costs of bad health by 50–300%.

Ignoring measurement error in health leads to underestimating the costs of bad health because it changes two key estimated parameters: the time costs of bad health and the persistence of health. A lower time cost of bad health translates into a smaller increase in resources when moving all individuals to the good health state. To see this, note that lower time costs of bad health mean that individuals have more time to work and hence can earn more and accumulate more assets when unhealthy. We expect this channel to be the first-order force behind the higher estimates for the costs of bad health when taking into account measurement error for the overall population.

Because health is persistent, the costs of bad health are higher for those who are initially unhealthy. This is true both when taking into account and when ignoring measurement error in health. However, when ignoring measurement error, we estimate a lower health persistence. This exacerbates the downward bias in the estimated lifetime costs of bad health for the initially unhealthy. To see this, note that if health was i.i.d., the initially healthy and the initially unhealthy would spend approximately the same amount of time in the bad health state.

The rest of the paper is organized as follows. Section 2 discusses how our paper relates and contributes to the existing literature. Section 3 develops our structural model of labor supply and saving behavior with health risk, which is designed to fit into the institutional context of the United Kingdom. Section 4 discusses our estimation strategy and the model fit. Section 5 lists data sources and data restrictions. Section 6 reports our main results, which compare the estimated lifetime costs of bad health when health is measured with and without error. Section 7 concludes.

2 Related Literature

Our paper contributes to two strands of the literature. First, we contribute to the literature that uses life-cycle structural models with health risk to ask a variety of substantive questions (Bueren, 2021; De Nardi, French and Jones, 2010; French, 2005; French and Jones, 2011). Most papers in this literature ignore measurement error in health. Two notable exceptions are French (2005) and Bueren (2021), which partially address the presence of measurement error in health. Both papers, however, impose restrictive parametric assumptions on the transition matrices for health. Also, they ignore measurement error in health when estimating the initial distribution of states, which includes health. On top of that, identification of the measurement error model for health in French (2005) is done under unnecessarily restrictive assumptions, and is not discussed in Bueren (2021). Our contribution to this literature is to estimate a structural model with health risk, guaranteeing the identification of the dynamics of health and its measurement system, and taking into account measurement error in each stage of the estimation procedure. In particular, we show how to estimate the initial distribution of states and the spousal earnings function taking into account measurement error in health. Moreover, we do so without imposing restrictive parametric assumptions on the evolution of health. We also extend results from the literature on identification and estimation of non-stationary hidden Markov models to secure identification of our measurement-error model, and to estimate it in a computationally tractable way.

Second, our paper speaks to the literature that examines the impact of health on different economic outcomes. Contributions to this literature come from both reduced-form and structural exercises. On the reduced-form side, Poterba, Venti and Wise (2017), Smith Jr (1993), Smith (1999), and Wu (2003), examine the nexus between adverse health shocks and wealth accumulation. Some other papers in the reduced-form literature examine the effects of adverse health conditions on labor supply, such as for example Bound (1991) and Costa-Dias, Blundell, Britton and French (2021). On the structural side, Capatina (2015) and De Nardi, Pashchenko and Porapakkarm (2018) estimate the impact of health in various economics outcomes, Gustman and Steinmeier (2014) estimate the role of health on retirement behaviour, and Amengual, Bueren and Crego (2021) estimate the assets costs of bad health using the traditional health measure and a novel measure proposed by them.

Our paper is most related to Capatina (2015), De Nardi, Pashchenko and Porapakkarm (2018), and Amengual, Bueren and Crego (2021). Like Capatina and De Nardi, Pashchenko and Porapakkarm, we look at the costs of bad health in a given outcome by comparing the average individual in the economy with a counterfactual individual that is lucky enough to receive always the best possible health shock. We follow Capatina in focusing on the costs of bad health as measured by earnings, assets, and participation. Some differences with respect to these papers are worth pointing out. The most important is that while these two papers ignore measurement error in health, we document that doing so leads to severely underestimating the costs of bad health as measured by these outcomes.

In principle, the additional flexibility when modeling health by De Nardi et al. (2018) makes their estimates more robust to deviations from the first-order Markov assumption for health than in what we and Capatina use. De Nardi, Pashchenko and Porapakkarm relax the first-order Markov assumption and add unobserved heterogeneity in health motivated by the observed state dependence in transition probabilities—the fact that individuals that have been longer in bad health have a lower probability of exiting that state. However, a hidden Markov model can generate state dependence in *measured* health.[4] Hence, while the additional flexibility in modelling health by De Nardi et al. (2018) makes a step forward with respect to Capatina (2015), it is not clear that their assumption is preferable to ours. Moreover, there is a sense in which the hidden Markov assumption is better for a structural model than the higher-order Markov assumption. If both assumptions can generate the same observable state dependence, the higher-order Markov assumption is less desirable from a computational standpoint since it adds an additional state to the structural model for each lag of health that is needed to predict health tomorrow.

We note that the magnitudes for the estimated costs of bad health between these studies and ours are not directly comparable for two reasons. First, we cover different parts of the life-cycle. Second, there are economic forces at play in the frameworks of the aforementioned studies that are not relevant in our institutional context (e.g., medical expenses and private health insurance.) The institutional context in our model is that of the United Kingdom. The reason why we focus on the United Kingdom

[4]For instance, hidden Markov models have been used in labor economics to explain the state-dependence of unemployment. See Shibata (2019).

rather than in, say, the United States is because this allows us to abstract from unnecessary complications in the environment, such as modeling employer-provided health insurance and out-of-pocket medical expenditures.[5]

Finally, Amengual, Bueren and Crego (2021), like us, estimate a non-stationary hidden Markov model for health and calculate the asset costs of bad health, both when taking into account and when ignoring measurement error. By estimating a hidden Markov model for health using a battery of many noisy measures, they account for measurement error when estimating health dynamics. However, they do not estimate structural preference parameters, and instead take the ones from De Nardi, French and Jones (2010), which are estimated ignoring measurement error in health. In this paper, we argue that re-estimating all structural parameters is key in to assess the bias in the estimated costs of bad health. In particular, we find that the estimated time cost of bad health increases substantially when we take measurement error into account.

3 Structural Model

In this section, we describe the individual intertemporal optimization problem of core household members.[6] For the sake of comparability with the literature, our model closely follows French (2005), a canonical model of labor supply and saving behavior with health risk. Since we are writing a model for the UK, we follow O'Dea (2018) in modelling institutional aspects that are relevant for our purposes.

3.1 Preferences

In the model, household heads make consumption (C_t), savings (a_{t+1}), leisure (L_t), and labor supply decisions (N_t) in each period to maximize their discounted sum of lifetime

[5]While the US does not provide universal health care, many European countries, including the UK, do. In the UK, the National Health System (NHS) provides free treatment at the point of use to the majority of the UK population—technically speaking, to UK ordinarily residents, which are people living in the UK lawfully, voluntarily, and for settled purposes. Private health care in the UK in 2015 was used by less than 11% of the population according to estimates of The Commonwealth Fund, and generally as a top-up to NHS services. For our purposes, this means that we can avoid modeling private health insurance and OOP medical expenditures if we focus in an institutional setting like that of the UK.

[6]See section 5 for a definition.

expected utility, which is given by

$$\mathbb{E}_0 \sum_{t=0}^{T} \beta^t \left[s_t(H_t) U(C_t, L_t) + \left[1 - s_t(H_t)\right] b(a_{t+1}) \right].$$

Conditional on being alive in period $t-1$, a household head with health H_t stays alive at time t with probability $s_t(H_t)$, and derives flow utility from consumption C_t and leisure L_t. With probability $1-s_t(H_t)$, the household head dies and derives utility from leaving behind a bequest a_{t+1}. This utility is captured by the warm-glow bequest motive utility function $b(a_{t+1})$.

We parametrize the flow utility that the household head derives from consumption and leisure when alive as:

$$U(C_t, L_t) = \frac{1}{1-\gamma} \left(C_t^{\nu} L_t^{1-\nu} \right)^{1-\gamma},$$

where $\nu \in [0, 1]$ measures the importance of consumption in the instantaneous utility, and the parameter γ governs both relative risk aversion and the degree of inter-temporal substitution of consumption and leisure.

The amount of leisure time that the household head enjoys each period is given by $\overline{L} - \phi_H \mathbf{1}(H = 2) - \phi_P \mathbf{1}(N > 0) - N$. In this expression, $\overline{L}$ is the total endowment of hours, N is the number of hours worked, and $\phi_P \mathbf{1}(N > 0)$ is a fixed cost of participation in the labor force. This term captures the time costs associated to working (such as commuting or dressing for work) and helps to rationalize that, conditional on working, hours of work are likely to be large. Finally, the term $\phi_H \mathbf{1}(H = 2)$ is the time cost associated to bad health, and it helps to rationalize that individuals that appear to be unhealthy work less hours.[7] As we will see below, $H = 2$ will correspond to the bad health state. The warm-glow bequest motive is parametrized following De Nardi (2004) as:

$$b(a) = \theta_B \frac{(\kappa_B + a)^{(1-\gamma)\nu}}{1-\gamma},$$

where θ_B is the weight on bequests, and κ_B governs the curvature of bequests. Note that if $\kappa_B > 0$, the marginal utility of leaving a positive bequest versus leaving no bequest is finite. This parameter can also be interpreted as governing the extent to which bequests are a luxury good.

[7]The choice of words "appear to be unhealthy" is deliberate. In our benchmark framework, which presupposes that health is imperfectly measured, health is unobserved by the econometrician.

3.2 Government and Private Pensions

In this model, the government taxes income, gives transfers, and administers public pensions, which start paying at the retirement age R_a.

The government taxes the personal income of the household head. Taxes are captured by the post-tax income function $y_t(\cdot)$, which takes as argument the gross income of the household head. The dependence of this function on t captures the fact that the tax scheme in the UK changes with age (see O'Dea, 2018, Appendix D.7).

The government also gives transfers tr_t to households to ensure a minimum level of consumption:

$$tr_t = \begin{cases} \max\{0, C_{min,t} - (W_t N_t + r_t a_t + a_t + ys_t)\}, & \text{if } t < R_a \\ \max\{0, C_{min,t} - (W_t N_t + r_t a_t + a_t + ys(t, H_t) + pbb_t + privben_t)\}, & \text{if } t \geqslant R_a \end{cases},$$

where ys_t denotes spousal income, and pbb_t and $privben_t$ denote public and private pension payments, respectively. In order to capture the fact that retirees in the UK face different means-tested programs than non-retirees, we let the consumption floor vary with age.[8] In particular, $C_{min,t}$ changes at R_a according to:

$$C_{min,t} = \begin{cases} C^y_{min}, & \text{if } t < R_a \\ C^o_{min}, & \text{if } t \geqslant R_a \end{cases}.$$

Public pension payments are a function of average lifetime earnings at R_a:

$$pbb_t = \begin{cases} 0, & \text{if } t < R_a \\ g(ae_{R_a}), & \text{if } t \geqslant R_a \end{cases}.$$

In an analogous fashion, private pension payments are also a function of average lifetime earnings at R_a:

$$privben_t = \begin{cases} 0, & \text{if } t < R_a \\ privben(ae_{R_a}), & \text{if } t \geqslant R_a \end{cases}.$$

Average lifetime earnings evolve according to:

$$ae_{t+1} = \frac{W_t N_t + (12 + t - 1) ae_t}{12 + t}.$$

Note that this specification assumes that individuals start working at age 26 (hence the number 12 in the expression, since 12 bi-annual periods have passed since labor market entry when the model starts at age 50.)

[8]See Appendix D.7 in O'Dea (2018) for a discussion of Pension Credit in the UK.

3.3 Wages, Assets, and Spousal Income

At each time period t, log wages are given by:

$$\log W_t(H_t, t) = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}(H_t = 1) + u_t.$$

Hence, wages have a deterministic age component captured by $a_0 + a_1 t + a_2 t^2$ that is common to all individuals, and an individual stochastic component captured by u_t. On top of that, the term $a_H \mathbf{1}(H = 1)$ allows for the possibility of healthy individuals receiving higher wages. We assume that the stochastic component of wages follows an autoregressive process:

$$u_t = \rho u_{t-1} + \xi_t, \qquad u_0 = \xi_0, \quad \xi_t \sim \begin{cases} \mathcal{N}(0, \sigma^2_{\xi,0}) & \text{if } t = 0 \\ \mathcal{N}(0, \sigma^2_{\xi,t}) & \text{if } t \geqslant 1 \end{cases}.$$

Since we are modelling households (as opposed to single individuals), it is important to include spousal income as part of the total amount of resources. We capture post-tax spousal income in a reduced form way—that is, we do not explicitly model the labor supply decision of the spouse. We model spousal income as in French and Jones (2011). That is, in our model, post-tax spousal income is given by:

$$ys_t = \begin{cases} ys(t, H), & \text{if } \ t \leqslant R_a + 1, \\ 0, & \text{if } \ t > R_a + 1, \end{cases}$$

where the choice of setting spousal income to zero starting from the period 68–69 $(= R_a + 2)$ is justified by the fact that, in the data, approximately 85% of individuals older than 68 have spouses with zero non-pension income .

Individuals can save on a safe asset a_t with rate of return r. Assets evolve according to:

$$c_t + a_{t+1} = y_t(r_t a_t + w_t N_t) + ys(t, H) + tr_t + a_t, \qquad \text{if } t < R_a,$$
$$c_t + a_{t+1} = y_t(r_t a_t + w_t N_t + privben_t + pbb_t) + ys(t, H) + tr_t + a_t, \qquad \text{if } t \geqslant R_a.$$

Moreover, individuals face a no-borrowing constraint:

$$a_{t+1} \geqslant 0.$$

3.4 Health

Since health affects survival probabilities, the total amount of available hours, and wages, it is important to specify how health evolves. In this paper, we assume that health takes two values when alive ($H_t \in \{1, 2\}$) and specify the "dead" state as the third health category ($H_t = 3$). Moreover, we assume that health evolves according to a non-stationary Markov process:

$$\mathbb{P}(H_{t+1} = i | H_t = j) = K_t(j, i),$$

for $i, j = 1, 2, 3$, where K_t denotes the transition matrix for health at age t. The non-stationarity of health transitions is to account for the fact that health declines with age, as according to most measures of health and consistently with the empirical literature.

Survival probabilities $s_t(H_t)$ can be obtained from transition matrices K as:

$$s_t(H_t) = 1 - K_t(H_t, 3),$$

for $H_t = 1, 2$.

3.5 Recursive Formulation and Model Solution

The solution of this model is a sequence of value functions $\{V_t\}_{t=0}^{T}$ and policy functions for assets, consumption, hours worked, and leisure $\{g_a, g_c, g_N, g_L\}_{t=0}^{T}$ that depend on the state vector:

$$\mathbf{X}_t = (H_t, a_t, ae_t, \eta_t).$$

At each t, the value function V_t solves:

$$\begin{aligned}
V_t(H, a, ae, \eta) = \max_{a', N} \quad & u\big(c, \bar{L} - \phi_P 1(N > 0) - N - \phi_H 1(H = \text{Bad})\big) \\
& + \big(1 - s(H, t)\big) b(a') + \beta s(H, t) \mathbb{E} V_{t+1}(\cdot, a', ae, \cdot) \\
\text{s.t} \quad & c + a' = y(ra + wN) + ys(t, H) + tr_t + a, \qquad \text{if } t < R_a \\
& c + a' = y(ra + wN + privben + pbb) + ys(t, H) + tr_t + a, \qquad \text{if } t \geqslant R_a, \\
& a' \geqslant 0, \\
& ae' = \frac{wN + (12 + t - 1) ae}{12 + t}.
\end{aligned}$$

The policy functions are the argmax of the RHS of this expression at each t. As usual, no analytic expression exists for the solution of this problem. Hence, value functions and policy functions have to be found numerically. We give more details on the numerical procedure used to solve this problem in Appendix D.

4 Estimation

We follow the large literature that estimates structural life-cycle models using a two-step procedure to estimate the parameters of our model (see, for example, De Nardi, French and Jones, 2010; French, 2005; Gourinchas and Parker, 2002). In the first step, we estimate some parameters outside the model and set some others to values taken from the literature. In the second step, we use an indirect-inference procedure to estimate the remaining parameters.

Because we have to estimate the model both taking into account measurement error and ignoring it in order to perform the comparison in counterfactuals, we explain how we conduct the estimation in each case.

4.1 Parameters Estimated or Calibrated Outside the Model

4.1.1 Process for Health and Measurement-Error Model

The stochastic process for health is the non-stationary Markov model described in the model section. The estimation of this process differs between the case in which we account for measurement error in health and the case in which we ignore it.

When we ignore measurement error, we follow the literature and assume that the observable health state can be obtained from collapsing self-reported health status into two categories. Identification and estimation of the stochastic process for health is straightforward when we assume that health is observable. In this case, the transition matrix for health and mortality can be identified and estimated using the empirical transition probabilities from the data between health states.

Assuming that health can only be measured with error complicates identification and estimation of the health process, because now transition probabilities are not directly observable. In order to deal with this complication, we adopt methods developed by García-Vázquez (2021) to identify non-stationary hidden Markov models. We adapt slightly those methods to allow for attrition due to mortality. See Appendix A for a formal identification argument.

Importantly, we place no parametric restrictions on the transition probabilities for health, both when we account for measurement error and when we ignore it. An alternative to doing this is to impose such parametric restrictions, by assuming for example that they are a logit as a function of age. This is precisely what Amengual, Bueren and Crego (2021) do. While this approach can improve efficiency if the parametric restrictions hold for the true data generating process, it can lead to bias if this is not the case. If we followed this approach, comparisons between counterfactuals when ignoring and when taking into account measurement error could come from measurement error itself, or from misspecification of the transition probabilities when estimating them using parametric assumptions. By leaving transition probabilities unrestricted in estimation, our procedure is robust to these kind of concerns.

Identification and estimation of the measurement-error model require three noisy measures. In this paper, we choose the following as noisy measures of health: self-reported pain, the total number of complications with ADLs and IADLs, and the number of mobility conditions.

In order to estimate the stochastic process for health, we also need to estimate the relationship between the health measures and true health status, which is captured by the *emission matrices*. More precisely, let the distribution of the noisy measure for health m, given that the true health status is c, be given by the vector p_c^m. The emission matrix for measure m is given by:

$$P^m = \left(p_1^m, \quad p_2^m \right).$$

Moreover, the emission matrices are also important because we use them to generate the noisy measure of health inside the model, which we need in order to target the relevant profiles by measured health. Tables 1–3 report the estimates for the different emission matrices.[9]

[9]Since the number of ADLs and IADLs can take many values we top-code this measure at 4. For the same reason, we top-code the number of mobility conditions at 8.

Table 1: Emission matrix for number of mobility conditions.

Mobility Conditions	Good Health	Bad Health
0	0.5200	0.0014
1	0.2143	0.0285
2	0.1385	0.0755
3	0.0738	0.1226
4	0.0350	0.1408
5	0.0155	0.1446
6	0.0030	0.1403
7	0.0003	0.1332
8+	1.3259E-06	0.2132

Table Notes. Probabilities are rounded up to four decimals.

Table 2: Emission matrix for pain

	Good Health	Bad Health
No Pain	0.7749	0.1863
Mild Pain	0.10039	0.1259
Moderate Pain	0.1051	0.4379
Severe Pain	0.0196	0.2499

Table Notes. Probabilities are rounded up to four decimals.

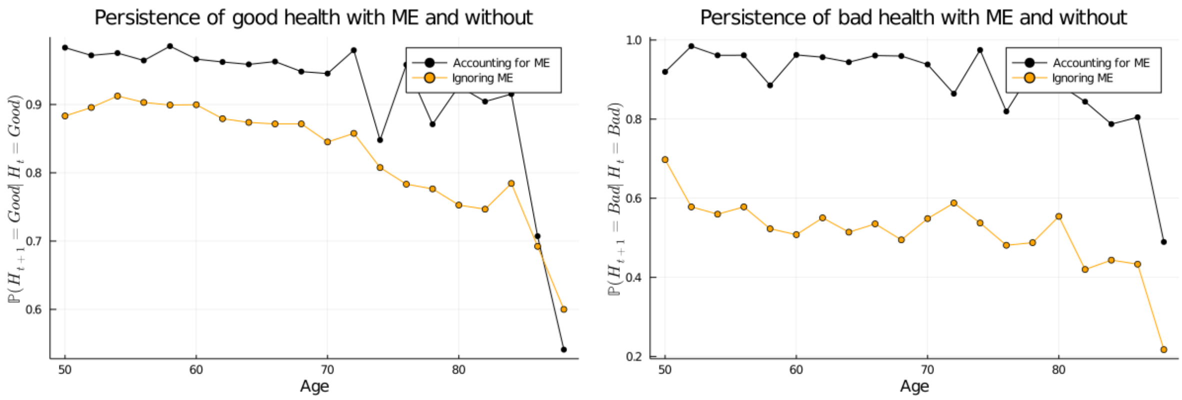
Table 3: Emission matrix number of limitations with ADL's and IADL's

Limitiations (ADL's + IADL's)	Good Health	Bad Health
0	0.8692	0.1879
1	0.0986	0.2066
2	0.0246	0.1722
3	0.00583	0.1219
4+	0.0019	0.3114

Table Notes. Probabilities are rounded up to four decimals.

In Figure 1, we present the comparison between transition probabilities for health when health is measured with and without error. The left panel depicts the probability of staying in good health for the different age groups, and the right panel does the same for those in bad health. As it is evident from these graphs, the measured persistence of health is higher when measurement error is accounted for.

Figure 1: Persistence of good and bad health accounting for and ignoring ME.



4.1.2 Spousal Income

In the model, spousal income is a deterministic function of the health status and the age of the household head. Hence, the spousal income function evaluated at age t and health H_t is simply the conditional mean of spousal income given that age is t and health status is H_t. That is:

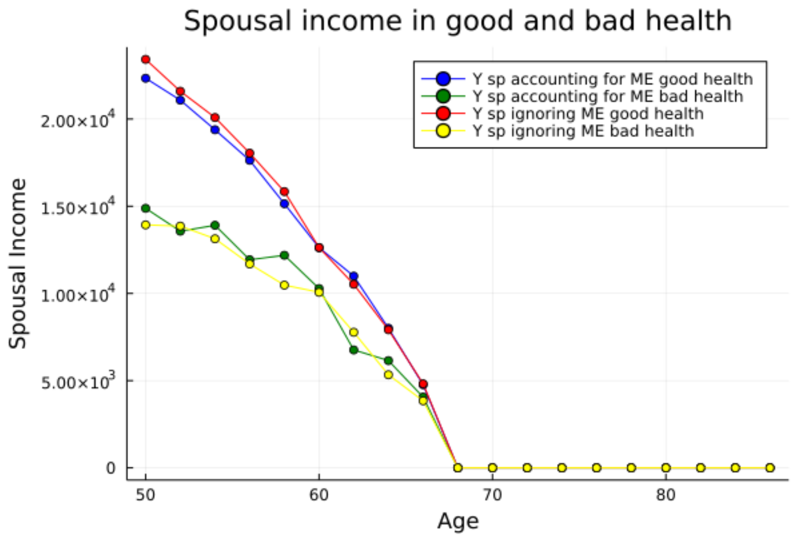
$$ys(H,t) = \mathbb{E}[ys_{i,t}|H_t].$$

Since non-pension spousal income is zero for most households with a retired household head, and since we include pensions of the spouse in the private and public pension functions, we assume that spousal income is zero for $t > R_a + 1$.

When we take into account measurement error in health, average spousal income given health status and age can still be identified from the mean spousal income by age given the noisy measures of health and the distribution of the noisy measures given true health. See Appendix B.2 for a formal identification proof. See also Appendix B.2 for a detailed explanation of the minimum distance procedure used to estimate the spousal income function when health is measured with error.

The results of estimating the spousal income function accounting for measurement error and ignoring it can be seen in Figure 2. From this figure, it is clear that the spousal income function that we estimate taking into account measurement error is remarkably similar to the one that we estimate when we ignore it. Hence, we do not expect differences in counterfactuals when accounting and ignoring measurement error to come from differences in spousal income. The similarity of the two functions is more remarkable once we remember that the measure of "observable" health is not included as a noisy measure of health when taking into account measurement error. Hence, the previous figure serves as (admittedly informal and partial) validation of our choice for noisy measures of health.

Figure 2: Spousal income function.



4.1.3 Wage-shock Parameters

Table 4 reports the estimated parameter values for the wage-shock process for two different health measures. The first measure (Health, with DLs) is a dummy variable for good health that takes the value 1 if the individual presents no limitations with ADLs or IADLs. This is the health measure that we use when estimating the model that takes into account measurement in health. The second measure (Health, with SRHS) is also a dummy variable for good health, but that takes the value 1 if the individual claims to be in “excellent,” “very good” or “good” health in questions about self-reported health status. This is the health measure that we use when we purposely ignore measurement error in health. For specific details on the estimation procedure, see Appendix B.3.

Notice that the estimated parameters reported in Table 4 are remarkably similar. The main reason for this is that there is a very high correlation between reporting to be in “excellent,” “very good” or “good” health and not experiencing any limitations with daily-living activities (DLs). Since all other variables in regression equation (7) are

Table 4: Minimum Distance estimates of wage-shock parameters.

Parameter	Health, with DLs	Health, with SRHS
ρ	0.8519	0.8589
	(0.19)	(0.19)
σ^2_η	0.1029	0.0962
	(0.07)	(0.07)
$\sigma^2_{\xi,1}$	0.1604	0.1690
	(0.09)	(0.09)
$\sigma^2_{\xi,t}$	0.0614	0.0602
	(0.04)	(0.04)
σ^2_m	0.1683	0.1696
	(0.03)	(0.02)

Table Notes. Standard errors, calculated using 100 bootstrap samples, are reported in parenthesis.

the same, the estimated residuals will also be very similar and, hence, the estimated parameters under the two health classifications very close to each other.

4.1.4 Initial Distribution of States

In order to initialize the model, we need the initial distribution of states. When we ignore measurement error, we sample 5,000 individuals with replacement from the joint distribution of states of assets, average earnings, and health. For each of these individuals, we generate their initial wage shock according to the estimated initial distribution of wage shocks. When measurement error is taken into account, we proceed as follows. First, we initialize the initial distribution by simulating initial health status for 5,000 individuals according to the estimated initial probability distribution of health. Then, we simulate average earnings and assets for those individuals sampling from the joint distribution of assets and average earnings given their health status, which we estimate. Then, we simulate the wage shock independently just as before. See Appendix D for details on the identification and estimation of the initial distribution of assets and average earnings when health is measured with error.

4.1.5 Pension Parameters

Similar to O'Dea (2018), we estimate parametrically the parameters that relate public- and private pension benefits with average earnings by running the following two OLS regressions:

$$pbb_{i,65} = ss_1 ae_{i,64} + ss_2 ae_{i,64}^2 + \varepsilon_i, \qquad \text{for } ae_{i,64} \leqslant \widehat{ae}_{ss}, \tag{1}$$

$$privben_{i,65} = pp_0 + pp_1 ae_{i,64} + pp_2 ae_{i,64}^2 + \xi_i, \tag{2}$$

where $pbb_{i,65}$ denotes the public pension benefits at the household level of individual i at age 65, $ae_{i,64}$ are average earnings of the individual at age 64, $\widehat{ae}_{ss}(= 75,000)$ is the threshold at which the quadratic relationship between public pension benefits and average earnings starts to decrease, $privben_{i,65}$ are private pension benefits at the household level, and ε_i and ξ_i are white noise.

The estimated parameters are reported in Table 5.

Table 5: Pension parameters

Parameter	Value	S.E.
ss_1	0.6518	0.0006
ss_2	−3.56E-06	1.31E-08
pp_0	5,980.80	591.35
pp_1	0.3426	0.0266
pp_2	5.23E-07	2.45E-07

4.1.6 Parameters Fixed Outside the Model

We fix several parameters outside the model. The curvature parameter for bequests κ_B is set to $650,000$ pounds following O'Dea (2018). The total endowment of hours is fixed at 8,760 bi-annual hours, corresponding to 12 daily hours. We set the interest rate on non-housing wealth at 3.23%. This is consistent with the annual rate of 1.6% used by O'Dea (2018). These parameters are displayed in Table 6.

Table 6: Parameters taken from the literature

Parameter	Value	Source
κ_B: curvature of bequests	650,000	O'Dea (2018)
$\bar{L}$: total endowment of bi-annual hours	8,760	12 daily hours
r: interest rate non-housing wealth	0.0323	O'Dea (2018)

We also set outside the model the parameters of the tax function. Since we are modeling the UK, we follow O'Dea (2018) and model individual income taxes as:

$$\text{Income taxes}(ti, t) = \begin{cases} 0, & \text{if } ti \leqslant \kappa_1^t \\ 0.2(ti - \kappa_1^t), & \text{if } \kappa_1^t < ti \leqslant \kappa_2^t, \\ 0.2(\kappa_2^t - \kappa_1^t) + 0.4(ti - \kappa_2^t), & \text{if } \kappa_2^t < ti \end{cases}$$

where ti denotes taxable income, which consists of labor earnings, asset income, social security payments and private pension payments, where the last two only pay starting at retirement age.

Table 7 lists the income tax thresholds. Notice that these are just twice as large as O'Dea (2018)'s thresholds to account for the bi-annual frequency of our model. Discrepancies in age with O'Dea are due to the bi-annual nature of our model.

Table 7: Income Tax Thresholds

	Age		
Parameter	< 64	64–73	$\geqslant 74$
κ_1	16,210	21,000	21,200
κ_2	84,940	89,740	89,940

4.2 Indirect Inference

In the second step, we estimate preference parameters $(\beta, \gamma, \nu, \phi_P, \theta_B, \kappa_B, \phi_H)$, consumption floors (C^y_{min}, C^o_{min}), and the parameters of the deterministic wage profile (a_0, a_1, a_2, a_H) using indirect inference (Gourieroux et al., 1993; Smith Jr, 1993). Indirect inference proceeds by minimizing the distance between a vector of target statistics calculated in the data and their model counterparts, which are calculated using simulation.

To be more precise, let $\theta = (\beta, \gamma, \nu, \phi_P, \theta_B, \kappa_B, \phi_H, C^y_{min}, C^o_{min}, a_0, a_1, a_2, a_H)$ be the vector of parameters to be estimated. We estimate our second-stage parameters by solving:

$$\hat{\theta} \equiv \arg\min_{\theta} \, (\psi(\theta) - \psi_n)' W_n (\psi(\theta) - \psi_n)',$$

where ψ_n is the vector of targets in the data, $\psi(\theta)$ is the vector of simulated targets, and W_n is a positive-definite weighting matrix. See Appendix F for more details on the indirect inference estimation procedure used here.

Note that the parameters that we are estimating in the second stage include the deterministic wage profile parameters. In principle, one could wonder why we need to include them in the second stage estimation, as opposed to calculate them directly from the data. The reason for that is because participation is endogenous in our model, and hence wage profiles estimated directly from the data will be subject to selection bias.

It is worth mentioning that the literature has proposed other ways of dealing with this selection problem. Notably, French (2005) proposes an iterative procedure that reaches convergence when the biased estimates for the deterministic wage profiles in simulated data are identical to the biased wage profiles estimated in the actual data. The problem with this procedure is that it is not guaranteed to converge (see footnote 19 in French, 2005). This lack of convergence turned out to be a problem for us in the implementation of this procedure. Nevertheless, targeting the wage profiles estimated in the data is conceptually similar—the distinction is essentially a "soft" matching of the profiles versus an strict matching as in the French procedure—and it allows us to use standard asymptotic Indirect Inference results for the wage profiles estimates.

The targets that we are seeking to match are assets, hours worked, and participation (all by measured health status), and the coefficients of a fixed-effect regression of log-wages on a quadratic polynomial in age and a dummy for being classified as "healthy."

At this point, it is worth clarifying what we mean by "measured health status." For the sake of calculating targets, both when we take into account measurement error and when we purposefully ignore it, we measure health status as a binary variable. When we ignore measurement error, we follow most of the literature and measure health status in the data as the binary variable resulting from collapsing self-reported health status. When calculating targets involving health in the simulated data, we use the true health status of each simulated individual.

When we take into account measurement error, we use as our binary health indicator a transformation of the number of IADL's and ADL's. More precisely, we classify as healthy individuals that have no problems with ADL's or IADL's, and as unhealthy individuals that exhibit problems with some ADL or IADL. The targets that we calculate in the data use that classification. Moreover, because we want the targets in the simulated data to be comparable to the targets in the actual data, we generate a dummy variable in the simulated data that tells us whether the individual suffers from problems with some IADL or ADL. In order to generate the binary ADL-IADL measure for each individual in the simulated data, we use the distribution of this dummy variable given true health status, which we can estimate consistently (see section 4.1.1).

Table 8 contains the parameters estimated by indirect inference, both when we take into account measurement error and when we ignore it.

Table 8: Parameters estimated by Indirect Inference.

Parameter	**Taking into account ME**	**Ignoring ME**
β: bi-annual discount factor	0.76	0.80
γ: CRRA coefficient	3.93	3.71
ν: consumption weight in utility function	0.48	0.46
ϕ_P: fixed cost of participation	1097.71	1076.12
θ_B: weight on bequest	0.066	0.076
ϕ_H: time cost of bad health	1851.190	875.05
C^y_{min}: consumption floor when young	6927.27	4573.36
C^o_{min}: consumption floor when old	13775.78	12408.93
a_0: constant term of wage profile	2.2065	1.92
a_1: linear age-term of wage profile	0.005	0.0118
a_2: quadratic age-term of wage profile	−0.005	−0.005
a_H: health coefficient of wage profile	0.0224	0.0256

Table Notes. Parameters ϕ and C_{min} should be interpreted in terms of bi-annual hours and bi-annual GBP, respectively.

We now briefly discuss identification. The bi-annual discount factor, β, is identified by the average growth of assets when individuals are young. The coefficient of relative risk aversion, γ, is identified by differences in assets between individuals in good and bad health. Recall that the coefficient of relative risk aversion equals the inverse of the inter-temporal elasticity of substitution of consumption. With a higher γ, the inter-temporal elasticity of substitution is lower and, hence, individuals are more eager to equalize consumption across periods with different health status. Hence, with a higher γ, individuals are willing to deplete assets more quickly in bad health and accumulate assets quicker in good health. Hence, γ is identified by the difference in assets in good and bad health. The weight of consumption in the utility function, ν, is identified by the level of hours that individuals work. This can be seen by noting that $(1-\nu)$ is the weight of leisure in instantaneous utility. The fixed cost of participation, Φ_P, is identified by the level of participation. The weight on bequests, θ_B, is identified by the level of assets that individuals have at the end of the life cycle. The time cost of bad health, Φ_H, is identified by differences in the level of participation and hours between those in good and bad health. The difference in consumption floors, C^y_{min} and C^o_{min}, is identified by variation in assets at retirement age. If the consumption floor for the elderly is more generous than the consumption floor for the young, assets should decline more around retirement age. Finally, the parameters pertaining to the wage process, $(\alpha_0, \alpha_1, \alpha_2, \alpha_H)$, are estimated in the data and the internal calibration procedure tries to match them correcting for the selection of workers into the labor market that the model generates. In essence, the identification argument for these parameters is similar to that in French (2005).

4.3 Assets, Participation, Hours, and Wage Profiles

Ideally, we would like to observe assets, labor-force participation, hours worked, wages, and health status for individuals in our cohort (those born between 1950 and 1957) since they enter the labor-market born until they die. This would allow us to construct profiles for the variables of interest spanning the relevant part of the life cycle. ELSA, however, starts surveying individuals at age 50 and is still today a relatively short-lived panel—it follows individuals for a maximum of eighteen years. This means that in the best case scenario we would be able to observe a large number of individuals from age 50 to age 68. In practice, however, there are not many such individuals, and many of the individuals we observe are not consistently followed throughout due to attrition.[10]

Given the data limitations that ELSA presents, the best we can do is to pool individuals from different birth cohorts and assume that differences in profiles across cohorts are driven by cohort effects, but not by different treatment of health and age attributes.

[10]Indeed, many researchers have pointed to the high attrition rates observed in ELSA.

This allows us to exploit much of the information contained in ELSA and thus construct profiles that range from ages 50–51 to 86–87.[11]

To construct profiles for non-housing wealth, hours worked, and labor-force participation by health status, we use individuals born between 1916 and 1957. To this purpose, we classify individuals in 19 two-year age groups (from ages 50–51 to 86–87), 4 cohort groups (those born before 1935, those born in between 1935–1943, those born in between 1943–1950, and those born after 1950), and 2 health groups (healthy and unhealthy). Since we estimate the model with and without taking into account measurement error, we use two binary indicators of health; one that relies on self-reported health status (SRHS), another that uses the number of limitations with ADLs and IADLs (DL). For each targeted variable $y \in$ {non-housing wealth, hours worked, labor-force participation} and health indicator Health^m, where $m \in \{\text{SRHS, DL}\}$, we run the OLS regression

$$y_{i,w} = \gamma_0^{y,m} + \eta_a^{y,m} + \eta_c^{y,m} + \gamma_a^{y,m}\left(a_{iw} \times \mathbf{1}_{\{\text{Health}_i^m = \text{Good}\}}\right) + u_{i,w}^{y,m}, \tag{3}$$

where i indexes individuals, w indexes waves, m indexes the health indicator used in the regression, η_a are age-group fixed effects, η_c are cohort fixed effects, a is the age group, and $u_{i,w}$ is white noise.

The implicit assumption behind this estimation procedure is that differences in y profiles across individuals of different cohorts are solely due to cohort effects. In other words, health and age are attributes that do not drive differences in y profiles across cohorts. In terms of economics, this amounts to saying that individuals of different cohorts may have faced different labor- and asset-market circumstances, but that these markets did not discriminate between them in terms of health and age.

Estimation of equation (3) for each (y, m)-pair provides us with estimates

$$\left\{\widehat{\gamma}_0^{y,m}, \{\widehat{\eta}_a^{y,m}, \widehat{\gamma}_a^{y,m}\}_{a=1}^{19}, \{\widehat{\eta}_c^{y,m}\}_{c=1}^{4}\right\}.$$

With these estimates, we can generate profiles by health status for each variable y and health indicator m for individuals in the cohort 1950–1957 according to:

$$\begin{aligned} y_a^{\text{good health}(m)} &= \widehat{\gamma}_0^{y,m} + \widehat{\eta}_a^{y,m} + \widehat{\eta}_{c=4}^{y,m} + \widehat{\gamma}_a^{y,m} a, \qquad \forall a \in \{1, \dots, 19\}, \\ y_a^{\text{bad health}(m)} &= \widehat{\gamma}_0^{y,m} + \widehat{\eta}_a^{y,m} + \widehat{\eta}_{c=4}^{y,m}, \qquad \forall a \in \{1, \dots, 19\}, \end{aligned}$$

[11] The choice of age 87 as an upper threshold is determined by inspection of ELSA data. There are not many individuals which are observed above this age. Moreover, given that life expectancy in the UK was 81.26 years for 2018 (as estimated by the World Bank), 87 is an age that most individuals will never reach, and an age that, for those who are able to reach it, is associated with a high probability of death.

where good health(m) is used to indicate that health was determined to be good based on variable Health^m. Similarly for bad health.

The vectors

$$y^{\text{good health}(m)} = \left(y_1^{\text{good health}(m)}, \ldots, y_{19}^{\text{good health}(m)}\right)$$

and

$$y^{\text{bad health}(m)} = \left(y_1^{\text{bad health}(m)}, \ldots, y_{19}^{\text{bad health}(m)}\right)$$

give the profiles of $y \in \{$non-housing wealth, hours worked, labor-force participation$\}$ for individuals in good and bad health, respectively, for a given health indicator.

To estimate wage profiles, we use individuals born between 1940 and 1970, who are younger than 77 years of age, and for whom we observe non-missing wages for at least two years. The restriction on wages is necessary because the estimation procedure exploits the time-series dimension of the data. The restrictions on individuals' birth years and the upper threshold for age respond to technical reasons and data limitations.[12]

We assume that the data generating process for wages is given by:

$$\log\left(\tilde{W}_{it}^{\text{data}}\right) = \tilde{a}_0^m + \tilde{a}_1^m t + \tilde{a}_2^m t^2 + \tilde{a}_H^m \mathbf{1}_{\{\text{Health}_{it}^m = \text{Good}\}} + \tilde{\eta}_i^m + \underbrace{\nu_{it}^m}_{= u_{it}^m + m_{it}^m}, \tag{4}$$

where $\tilde{W}_{it}^{\text{data}}$ is individual i's hourly wage at age t (measured with error), m indexes the health indicator used in the regression, $\tilde{a}_0$ is a constant term, $\tilde{a}_1$ is the coefficient of age, $\tilde{a}_2$ is the coefficient of age squared, $\tilde{a}_H$ is the coefficient of good health, $\tilde{\eta}_i$ is the individual fixed effect, and ν_{it} is the error term. The error term ν_{it} is the sum of the stochastic component of wages, u_{it}, and serially-uncorrelated measurement error, $m_{it} \sim \mathcal{N}(0, \sigma_m^2)$.

[12] We consider individuals born approximately 10 years before and after the individuals in our cohort (1950–1957) to cover a larger age span (50–77 vs. 50–68) and also to increase the precision of our estimates. As indicated by French (2005), using data on individuals older than the individuals of our sample helps to overcome some of the end-point problems associated with polynomial smoothing. Two reasons lead us to focus on individuals younger than 77 years of age. First, because our structural model is bi-annual, we have to respect a two-year age group classification (i.e., 50–51, 52–53,…, 76–77). In this classification, 77 is one of the natural end points. Other end points could be 79, 81, etc. However, given the small number of individuals with non-missing wages for at least two years starting at ages 78–79, we find age 77 to be an appropriate upper threshold for age.

We estimate equation (4) using the within estimator, and note that the estimated parameters for the earnings process for a given health indicator, $(\widehat{a}_0, \widehat{a}_1, \widehat{a}_2, \widehat{a}_H)$, do generally differ from the true ones, (a_0, a_1, a_2, a_H), due to the selection bias we encounter in the data—that is, we only observe wages for individuals who are currently working, but not the potential wages for all individuals in the labor force.[13] We deal with the selection problem for wages using the estimation procedure described in section 4.2.

4.4 Model Fit

The fit of the model can be assessed by looking at Figure 3. This figure displays the fit of the model for participation, hours worked, and assets when taking into account measurement error (left panels) and when ignoring it (right panels).

As we can see, the ability of the model to fit the data is similar when we take into account measurement error and when we ignore it. At the current parameter estimates, the model misses the level of the profiles, but roughly captures the trends. Looking at the graphs for participation and hours, it becomes clear that the model has a hard time capturing the smooth decline in hours and participation around retirement age. This is likely due to the way in which we model social security, which becomes available for all individuals at age R_a and then pays independently of earnings and hours worked. More precisely, given our parameter estimates, the model predicts a much steeper decline in hours and participation around retirement age than the one observed in the data. Also, notice that the model that takes into account measurement error requires a time cost of bad health that is much larger than when we purposely ignore measurement error. The time costs of bad health is a key parameter for estimating the earnings costs of bad health and, thus, ignoring measurement error in health will likely lead to underestimate significantly the earnings costs of bad health.

Table 9 reports the biased estimates for the wage profile parameters in ELSA and in the simulated data, both taking into account and ignoring measurement error in health.

[13] These include individuals who could not find and a job, and individuals who found a job but decided to turn down job offers.

Figure 3: Model fit taking into account ME (left) and ignoring ME (right).

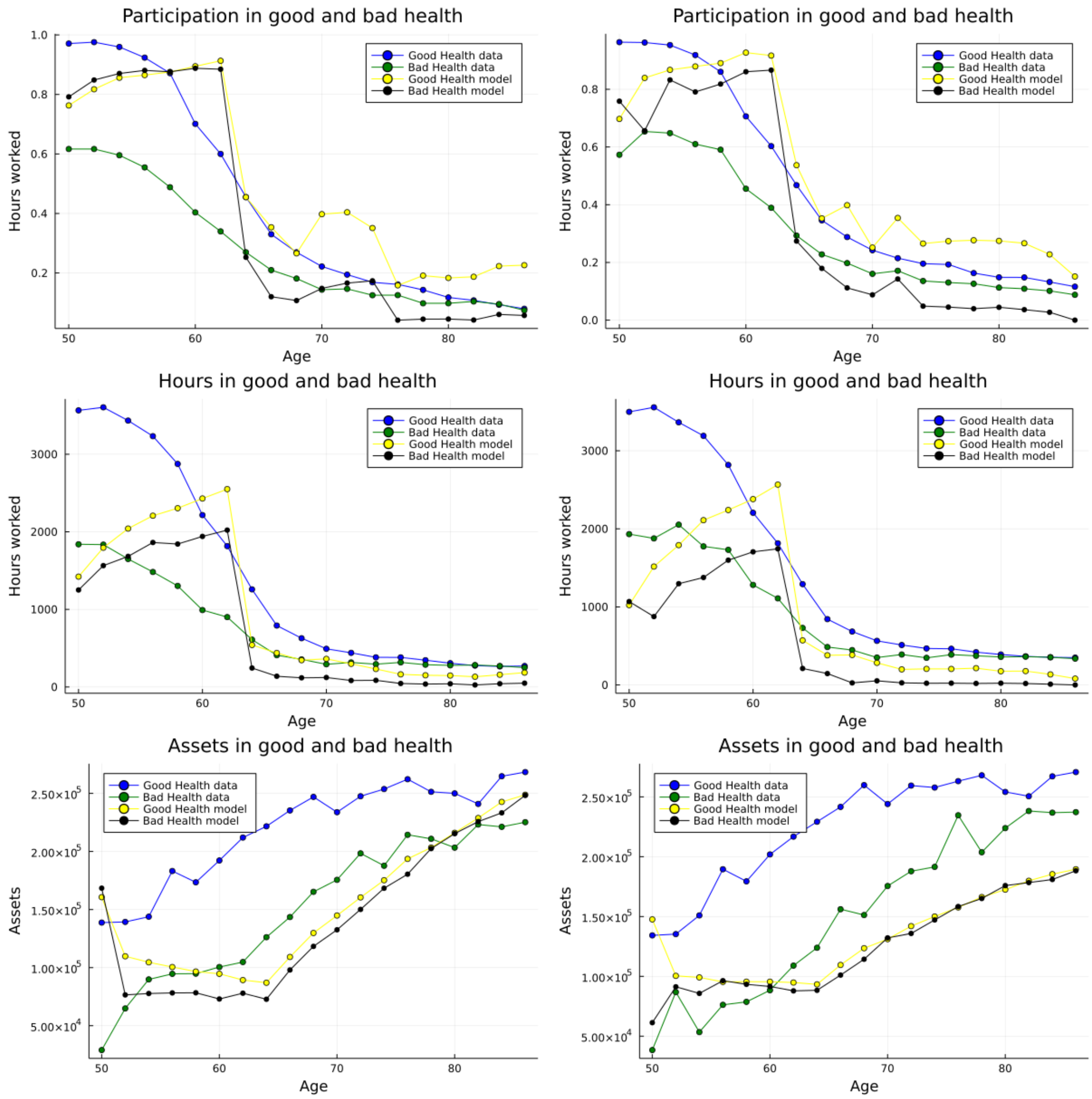


Figure Notes. Participation is in scale 0–1, hours of work are bi-annual, and assets are expressed in GBP.

Table 9: Fit estimated wage profiles

Parameter	**Accounting for ME**		**Ignoring ME**	
	Data	**Model**	**Data**	**Model**
a_0	2.09	2.24	2.1	1.94
a_1	0.058	0.003	0.06	0.013
a_2	-0.02	-0.004	-0.002	-0.005
a_H	0.024	0.002	0.009	0.013

5 Data

We use the English Longitudinal Study of Ageing (ELSA) for the time period 2002–2019. ELSA collects bi-annual survey data that is representative of the English population living in private households and which is aged 50 and over. In addition to these individuals, referred to as *core members*, ELSA also surveys their cohabiting partners in order to understand the circumstances of these individuals during the ageing process.[14] In order to maintain the representativeness of the data, we restrict attention to ELSA core members. ELSA's multidisciplinary coverage allows us to obtain the data necessary for our study. In particular, we can access a wide array of demographics, labor-market indicators, and health measures. Although we do not follow partners of core members, ELSA provides income information on these individuals for as long as they are in a formal relationship (either marriage or cohabiting) with core members. This information is essential to construct household-level variables of assets and pension income, as well as spousal income measures. Because we model the behavior of the core member of a household, we use labor supply, income- and health variables for this person, and household-level data for assets and pension benefits.

Since ELSA surveys core members starting at age 50, we need retrospective information on employment and earnings in order to construct labor-market and earnings histories prior to the ELSA survey period. This information is essential to compute average earnings—one of the state variables in our model—and also to estimate pension parameters outside the model. Ideally, we would like to have access to administrative data on National Insurance Contributions (NICs) as this would allow us to obtain reliable estimates of average earnings at different ages. Access to Social Security data from the UK is, however, restricted to researchers from UK academic institutions or government departments. For us, this makes accessing NICs data an impossible task. A potential go-around could be to exploit the information contained in ELSA Life History Interview, which enables construction of employment spells and earnings histories since an individual entered the job market.[15] This strategy is, however, not very effective in practice. Imputed average earnings do not exhibit the relationship with pension

[14]ELSA core members are individuals who fitted the age eligibility criteria, participated in the sample-origin Health Survey for England (HSE), and participated in ELSA's first wave if invited to do so. Core members remained eligible over the waves as long as they did not die or move outside of Great Britain.

[15]The ELSA Life History Interview was conducted jointly with wave-3 interviews with the objective of obtaining retrospective information on individuals in a variety of areas, including fertility, housing

benefits that many other researchers have documented using UK data, including O'Dea (2018). This points to a very noisy imputation procedure resulting from the many caveats that the data exhibit.[16] For these reasons, we decided to supplement data from ELSA with data from the simulated model of O'Dea (2018), which provides a good fit for the earnings profiles obtained from UK Social Security data for the cohort he studies. The implicit assumption that we make is, therefore, that our cohort (those born in between 1950–1957) is similar to his (those born in between 1935–1950). See Appendix C for a detailed description of the imputation procedure for average earnings.

To estimate the parameters that relate (public and private) pension payments with average earnings, we also use the data simulated by O'Dea (2018). The reason for this is, once again, twofold. First, the lack of access to administrative data on NICs. Second, the imputation procedure for average earnings that relies on data from the ELSA Life History Interview turns out to be very noisy, as indicated by the weak association between pension benefits and average earnings.

When estimating wage profiles and wage-shock parameters, we use individuals born between 1940 and 1970, who are younger than 77 years of age, and for whom we observe non-missing wages for at least two years. We drop observations for which we have missing wages, which results in a total of 16,880 person-wave observations for log wages, ages, and health dummies. Wages are computed as annual earnings divided by annual hours.

To estimate spousal income we use individuals born between 1942 and 1965, who are younger than 68 years of age, and for whom we observe spousal income and mobility conditions.[17] We define spousal income as the spouse's labor income, annuity income, and state benefit income, which includes incapacity benefits, disability allowances,

mobility, jobs and earnings, health, etc. Approximately 80% of individuals in ELSA wave 3 consented to participate in this complimentary survey.

[16]Some of the caveats the researcher has to deal with when working with the ELSA Life History Interview is that individuals only report "start" and "end" dates for each job, as well as their initial salary *at the time* of starting the job which is, in many cases, reported in pre-tax terms. This forces the researcher to make assumptions about the evolution of salaries during employment spells, to deal with the conversion of old English money to new English money, and to make simplifying assumptions on the complexity of the UK tax system.

[17]We focus on individuals younger than 68 years of age and assume, consistent with the structural model, that spousal income is zero for older ages.

war pensions, and the like. Dropping observations with missing values, this results in 34,152 person-wave observations for spousal income and health dummies.

We estimate profiles for non-housing wealth, hours worked, and labor-force participation by health status in the sample of individuals born between 1916 and 1957. Discarding observations with missing values, this results in a total of 48,578 person-wave observation for non-housing wealth, 49,406 person-wave observations for hours worked, and 49,346 person-wave observations for labor-force participation.

To estimate jointly the stochastic process for health and the measurement error model, we use data on adverse mobility conditions, limitations with ADL's and IADL's, and self-reported pain from the IFS derived variables. When ignoring measurement error, our measure of health is self-reported health status, which we take directly from the corresponding ELSA question.[18]

Since observing mortality is key for the estimation of the stochastic process for health, both when we take into account measurement error and when we ignore it, we restrict our attention to individuals that give permission to link their Social Security mortality record with ELSA. Since we only observe this measure of mortality for waves 1–5, we only use these waves when estimating the stochastic process for health. After applying these restrictions, discarding missing data, and restricting attention to ages 50–87, we end up with 45,099 and 38,944 individual-age group observations to estimate the stochastic process for health when taking into account and ignoring measurement error respectively.

6 The Lifetime Costs of Bad Health

In this section we report the results of several different health-related counterfactuals. More specifically, we study the costs of bad health in a battery of outcomes—some of these are pecuniary costs, such as foregone earnings, and others are non-pecuniary, such as the number of hours individuals do not work due to bad health.

[18]In order to construct the measure of observable health, we use for waves 1, 2, 4, and 5 a question in which the respondent is asked to describe her health as "excellent," "very good," "good," "fair" or "poor". Since this question is not available in wave 3, for this wave we use another question in which respondents are asked to rate their health as "very good," "good," "fair," "bad" or "very bad". In both cases, we collapse the three categories that correspond to the most favorable assessment of health as "Good Health" and the two worse as "Bad Health".

We follow De Nardi, Pashchenko and Porapakkarm (2018) in calculating the costs of bad health as measured by many outcomes. The exercise proceeds in three steps. First, we simulate the model imposing that all individuals are always in good health. From these simulations, we can obtain individuals histories for earnings, hours worked, consumption, and assets. Then, we simulate the model letting individuals' health evolve according to the transition matrix estimated in the data. Again, these simulations provide us with individual histories for the same objects. Finally, to find the difference in counterfactuals, we find the differences between the two scenarios in mean annual earnings before age 64, mean annual hours worked before age 64, mean annual consumption, and mean annual assets. These are the objects reported in Table 10. Notice that this table has three columns: one for when we take into account measurement error, one for when we purposely ignore it, and the difference between the two (expressed in percentage).

Table 10: Pecuniary Lifetime Costs of Bad Health (all individuals)

Outcome	**Taking into account ME**	**Ignoring ME**	**Difference** (%)
Earnings	971.86	456.42	112.93%
Hours worked	105.88	61.93	70.97%
Consumption	1,772.87	1,081.29	63.96%
Assets	21,826.64	16,632.15	31.23%

Table Notes. All variables are means expressed in annual terms. Mean earnings and hours worked are computed up to age 64 (inclusive). The units of earnings, consumption, and assets are GBP.

Our interpretation of Table 10 is the following. First, comparison between columns makes clear that ignoring measurement error in health would lead to substantially underestimate the costs of bad health for all outcomes of interest. The difference is especially noticeable in earnings, where accounting for measurement error more than doubles the estimated earnings costs of bad health. Second, looking at the differences in estimated parameters when we take into account measurement error and when we ignore it, it is clear that most of the discrepancy in the costs of bad health comes from differences in the persistence of health and the time costs of bad health. The way in which underestimating the time cost of bad health biases the estimated lifetime costs of bad health is clear. A lower estimated time cost of bad health translates into a smaller increase in resources when moving everyone to the good health state. We expect this channel to be the first-order force behind the higher estimates for the costs of bad health when taking into account measurement error.

In Table 11, we report the costs of bad health for individuals who are initially unhealthy. Because health is persistent, the costs of bad health are higher for these individuals than for the overall population. This is true both when taking into account and when ignoring measurement error. To understand the differences between the middle two columns, it is useful to keep in mind that the estimated persistence of health is higher when taking into account measurement error. Moreover, the persistence of health is the main driver of the higher costs of bad health for the initially unhealthy.[19] This helps rationalizing why the costs of bad health for the initially unhealthy are higher when we take into account measurement error than when we purposely ignore it. Table 11 confirms this intuition.

Table 11: Pecuniary Lifetime Costs of Bad Health (individuals initially in bad health)

Outcome	**Taking into account ME**	**Ignoring ME**	**Difference** (%)
Earnings	3,962.32	987.00	301.45%
Hours worked	432.84	127.37	239.83%
Consumption	3,821.79	1,618.09	136.19%
Assets	32,046.68	21,038.62	52.32%

Table Notes. All variables are means expressed in annual terms. Mean earnings and hours worked are computed up to age 64 (inclusive). The units of earnings, consumption, and assets are GBP.

Overall, our exercise suggest that exercises like the ones in Capatina (2015) and De Nardi, Pashchenko and Porapakkarm (2018) are likely to highly underestimate the lifetime costs of bad health.

7 Conclusion

In this paper, we estimate a structural model of savings and labor supply with health risk under two different assumptions on the observability of health. The first assumption, which is prevalent in much of the literature, is that health is perfectly observable (see, for instance, Capatina, 2015; De Nardi et al., 2018; French and Jones, 2011). The

[19] If health was i.i.d., the expected time that an initially healthy and an initially unhealthy individual would spend in the bad health state would be more similar than when health is persistent.

second assumption is that while health is not observable, a battery of noisy measures of health can be observed.

Estimating the structural model under the two different assumptions on the observability of health enables us to assess the impact of measurement error in the costs of bad health. We measure the costs of bad health using a battery of outcomes, such as labor earnings, hours worked, consumption, and assets. Our main finding is that ignoring measurement error in health leads to considerably underestimating the costs of bad health, especially for those who are initially unhealthy. More specifically, our exercise reveals that not taking into account measurement error could lead to underestimating the costs of bad health by as much as 300%. Not taking into account measurement error in health leads to underestimating the costs of bad health because doing so results in both lower estimated time costs of bad health and lower persistence of health.

Our paper suggests that exercises that try to measure how health contributes to lifetime inequality in income and wealth, such as De Nardi, Pashchenko and Porapakkarm (2018), are likely to highly underestimate the importance of health. Thus, a key message from our paper is that researchers seeking to address the costs of bad health using structural life-cycle models should take measurement error in health seriously.

References

Amengual, D., Bueren, J. and Crego, J. A. (2021). Endogenous health groups and heterogeneous dynamics of the elderly, *Working paper* .

Amengual, D. and Sentana, E. (2015). Is a normal copula the right copula?

Bonhomme, S., Jochmans, K., Robin, J.-M. et al. (2016). Estimating multivariate latent-structure models, *Annals of Statistics* **44**(2): 540–563.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models, *Journal of Human Resources* **26**(1): 106–138.

Bueren, J. (2021). Long-term care needs and savings in retirement, *Working paper* .

Capatina, E. (2015). Life-cycle effects of health risk, *Journal of Monetary Economics* **74**: 67–88.

Costa-Dias, M. S., Blundell, R., Britton, J. and French, E. (2021). The impact of health on labor supply near retirement, *Journal of Human Resources* .

De Nardi, M. (2004). Wealth inequality and intergenerational links, *The Review of Economic Studies* **71**(3): 743–768.

De Nardi, M., French, E. and Jones, J. B. (2010). Why do the elderly save? the role of medical expenses, *Journal of political economy* **118**(1): 39–75.

De Nardi, M., Pashchenko, S. and Porapakkarm, P. (2018). The lifetime costs of bad health, *Technical report*, National Bureau of Economic Research.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour, *The 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.

García-Vázquez, M. (2021). Identification and estimation of non-stationary hidden markov models, *Working Paper* .

Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect inference, *Journal of applied econometrics* **8**(S1): S85–S118.

Gourinchas, P.-O. and Parker, J. A. (2002). Consumption over the life cycle, *Econometrica* **70**(1): 47–89.

Gustman, A. L. and Steinmeier, T. L. (2014). The role of health in retirement, *Technical report*, National Bureau of Economic Research.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*, CRC Press.

O'Dea, C. (2018). Insurance, efficiency and the design of public pensions, *2018 Meeting Papers*, Society for Economic Dynamics.

Poterba, J. M., Venti, S. F. and Wise, D. A. (2017). The asset cost of poor health, *The Journal of the Economics of Ageing* **9**: 172–184.

Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives, *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* pp. 26–46.

Shibata, I. (2019). Labor market dynamics: A hidden markov approach.

Smith, J. P. (1999). Healthy bodies and thick wallets: the dual relation between health and economic status, *Journal of Economic perspectives* **13**(2): 145–166.

Smith Jr, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics* **8**(S1): S63–S84.

Song, P. X.-K., Fan, Y. and Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference, *Journal of the American Statistical Association* **100**(472): 1145–1158.

Wu, S. (2003). The effects of health events on the economic status of married couples, *Journal of Human Resources* **38**(1): 219–230.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension, *Communications in Statistics—Theory and Methods* **34**(9-10): 1867–1879.

Appendices

A Identification of the Measurement-Error Model for Health

In this Appendix we provide two identification results for the measurement-error model for health. Both of these results show identification of a non-stationary Markov model with attrition (due to mortality) and refreshment samples. The difference between the two results lies in the set of sufficient conditions needed for identification. The reason why we provide two identification results is because there exist situations in which the researcher cannot establish identification with one result but could do so with the other.

A.1 Identification with Mortality

In this appendix, we show that the stochastic process for health, and the relation between the observable health measures and the unobservable health state are identified. Identification is complicated by the fact that individuals can die during the sample period, and also enter the sample at different ages. To make more transparent the identification challenge we face, we are going to distinguish in this section between t (the ELSA wave) and age a. Note that this distinction is not relevant in the model, since we are only modelling a cohort. However, this distinction is relevant for the estimation and identification of the health process, since we are using data from different cohorts. Assume that mortality from age a to age $a+1$ is perfectly observable. Let the number of ages the model is composed of be A and the number of sample periods be T.[20] Let $T < A$ (that is, using only people that enters the sample at each $a = 0$ and at time $t = 0$ will not be enough for identification). For simplicity, we discuss the case in which there are three underlying health states: Bad Health, Good Health, and Death. That is,

$$H_{iat} \in \{1, 2, 3\}.$$

The underlying health state H is indexed by individual (i), age (a), and sample period (t). Let a_i^0 be the age at which individual i enters our sample. We have that

$$\mathbb{P}(H_{i,a_i^0,t} = 3) = 0.$$

That is, individuals cannot enter the sample when they are already death.

[20]In our particular application, T corresponds to the number of waves that we use from ELSA.

As we said before, we observe three noisy measures Y^m of the underlying state (self-reported pain, number of mobility conditions and total number of ADL and IADL conditions). Each noisy measure can take $\kappa_m + 1$ values; that is, $Y^m \in \{1, 2, \ldots, \kappa_m, \kappa_m + 1\}$. The first κ_m values of Y^m are the different values that the noisy measure of health Y^m can take, while the last value tells us if the individual is dead. That is, if an individual is alive, $Y^m = \kappa_m + 1$ with probability zero.

Let the m-th emission matrix be given by Q^m, where:

$$Q^m = \begin{pmatrix} P^m & 0_{\kappa_m \times 1} \\ 0 & 1 \end{pmatrix}.$$

As in the body of the text, the emission matrix for measure m while the individual is alive is given by:

$$P^m = \begin{pmatrix} p_1^m, & p_2^m \end{pmatrix}.$$

where p_c^m is a vector that contains the distribution of measure m given that $H = c$. Then we can prove the following identification result:

Proposition 1 (Identification with mortality, refreshment samples, and $T < A$). *Suppose that for each $m = 1, 2, 3$, P^m is full column rank. Moreover, assume that for each age $\pi_{ac} > 0$ for $c = 1, 2$. Finally, assume that for each age a we can observe a subset of individuals at a and $a + 1$. Then, the measurement-error model for health is identified.*

Proof. For each age a we can observe the distribution of Y^m for $m = 1, 2, 3$ conditional on $S \neq 3$. Hence, the data at each age is generated by a finite mixture model with parameters P^m, $m = 1, 2, 3$, and $\tilde{\pi}_{ac}$, $c = 1, 2$, where $\tilde{\pi}_{ac} = \mathbb{P}(H_a = c | H_a \neq 3)$. Since $\{P^m\}_{m=1,2,3}$ are full rank and $\tilde{\pi}_{ac} > 0$ for all a and $c = 1, 2$, by application of Theorems 2–3 in Bonhomme et al. (2016), it follows that P^m and $\tilde{\pi}_{ac}$, $c = 1, 2$, are identified.

Since P^m is identified for $m = 1, 2, 3$, Q^m is identified (once we have P^m, we only have to complete the matrix with known entries to get Q^m).

Now, note that since we observe $\mathbb{P}(Y_a^m, Y_{a+1}^{m'} | H_a \neq 3)$ for $m, m' = 1, 2, 3$, we can find π_{a3} for each a according to:

$$\pi_{03} = 0, \qquad \text{(we are interested in individuals that are alive at 0)}$$

$$\pi_{a3} = \pi_{a-1,3} + \mathbb{P}(H_a = 3 | H_{a-1} \neq 3)(1 - \pi_{a-1,3}),$$

where

$$\mathbb{P}(H_a = 3 | H_{a-1} \neq 3) = \mathbb{P}(Y^m_{a-1} \neq \kappa_m + 1, Y^{m'}_a = \kappa_m + 1 | H_{a-1} \neq 3).$$

From this, we can identify π_a from $\tilde{\pi}_{ac}$, $c = 1, 2$, and π_{a3} as:

$$\pi_a = \big(\tilde{\pi}_{a1}(1 - \pi_{a3}), \tilde{\pi}_{a2}(1 - \pi_{a3}), \pi_{a3}\big).$$

Finally, since we know Q^m and π_a, we can identify K as:

$$K_a = (\Omega'_{0a}\Omega_{0a})^{-1}\Omega'_{0a}\mathbb{P}(Y_{1a}, Y_{1a+1})\Omega'_{1a}(\Omega'_{1a}\Omega_{1a})^{-1},$$

where $\Omega_{a0} = Q^1\Pi_a$, $\Omega_{a1} = \Pi^{-1}_{a+1}(Q_1\Pi_{a+1})'$ and $\Pi_a = \mathrm{diagm}(\pi_a)$ [21]. Finally notice that $\mathbb{P}(Y_{1a}, Y_{1a+1})$ can be recovered from:

$$\mathbb{P}(Y_{1a}, Y_{1a+1}) = \underbrace{\mathbb{P}(Y_{1a}, Y_{1a+1} | H_a \neq 3)}_{\text{Observed}}(1 - \pi_{a3}) + \underbrace{\mathbb{P}(Y_{1a}, Y_{1a+1} | H_a = 3)}_{\text{Known}}\pi_{a3}.$$

This completes the proof. □

B Estimation of Parameters Outside the Model

In this Appendix, we explain the different estimation procedures for parameters that are estimated outside the model.

B.1 Emission Matrices for Health and Transition Probabilities

The measurement-error model for health is the non-stationary hidden Markov model described in sections 3 and 4.1.1.

The use of Maximum likelihood and the Baum–Welch algorithm to estimate this model is challenging because of the joint presence of non-stationarities and refreshment samples.[22] Hence, we adapt and apply the two-step, constrained Baum–Welch algorithm presented in García-Vázquez (2021). Adaptations are needed because of the need to deal with mortality.

The two-step constrained Baum–Welch algorithm adapted to our application is:

Step 1.

[21] $\mathrm{diagm}(x)$ denotes a matrix that contains the vector x in its diagonal and 0's outside of it.

[22] The Baum-Welch algorithm is a particular case of the EM algorithm, used in Maximum Likelihood Estimation of Hidden Markov Models.

(a) At each $t = 1, \ldots, T$ restrict the sample to observations that are not missing or death. Use Maximum Likelihood and the EM algorithm to get $\sqrt{N}$-consistent and asymptotically-normal estimates of P^m and $\tilde{\pi}_t = (\mathbb{P}(H_t = s | H_t \neq r + 1))_{s=1,\ldots,r}$.

(b) For each t, calculate the proportion of people that dies between t and t+1 as:

$$\text{prop death}_{t,t+1} = \hat{\mathbb{P}}(Y^1_{t+1} = \kappa_1 + 1 | Y^1_t \neq -7, \kappa_1 + 1).$$

(c) Let

$$\pi^{H_t \neq r+1}_{t+1} = (\mathbb{P}(H_{t+1} = s | H_t \neq r + 1))_{s=1,\ldots,r+1}.$$

A consistent estimate of this object is given by:

$$\hat{\pi}^{S_t \neq r+1}_{t+1} = \left(\hat{\tilde{\pi}}_{t+1}(1 - \text{prop death}_{t,t+1}), \text{prop death}_{t,t+1}\right),$$

where $\hat{\tilde{\pi}}_{t+1}$ denotes the estimate for $\tilde{\pi}_{t+1}$ from step 1.

Step 2. For each $t = 1, \ldots, T-1$ restrict the sample to observations at t and t+1 that are non-missing in t and t+1 and non-death in t. Estimate K_t iterating between the following two steps until convergence:

- **E step:** Given estimates for $\hat{\tilde{\pi}}_t, \hat{\pi}^{H_t \neq r+1}_{t+1}, Q^1$, $\{Y^1_{i,\tau}\}_{\tau=t,t+1}$ and a guess for $K^{(h)}_t$, obtain the filtered probabilities

$$\hat{\nu}_{i,k,j} := \mathbb{P}(H_{i,t+1} = j, H_{i,t} = k | Y^1_{i,t}, Y^1_{i,t+1}, \{\hat{\pi}_\tau\}_{\tau=t,t+1}, \hat{Q}^1, K^h_t).$$

These filtered probabilities can be calculated as follows:

$$\hat{\nu}_{i,k,j} = \frac{\hat{Q}^1(y^1_{i,t}, k)\hat{\pi}_t(k)K^{(h)}(k,j)\hat{Q}^1(y^1_{i,t+1}, j)}{\sum_{j=1}^r \sum_{k=1}^r \hat{Q}^1(y^1_{i,t}, k)\hat{\pi}_t(k)K^{(h)}(k,j)\hat{Q}^1(y^1_{i,t+1}, j)}.$$

- **M step:** Calculate the new guess $K^{(h+1)}_t$ as:

$$\begin{aligned} K^{(h+1)}_t = \arg\max_K \quad & \sum_{i=1}^N \left\{ \sum_{k=1}^r \sum_{j=1}^r \nu_{ikj} \log(K(k,j)) \right\} \\ \text{s.t.} \quad & \sum_{j=1}^r K_t(k,j) = 1 \text{ for all } k, \\ & \sum_{j=1}^r K(j,c)\hat{\tilde{\pi}}_t(j) = \hat{\pi}^{H_t \neq r+1}_{t+1}(c). \end{aligned}$$

B.2 Spousal Income

We are interested in estimating $\mathbb{E}[ys_t|H_t = c]$ for $c = 1, \ldots, r$. The identification challenge here is that obviously S_t is not observed. In order to overcome this challenge, we need the following exclusion restriction:

Assumption 1 (Exclusion Restriction). $\mathbb{E}[ys_t|Y_t^1, H_t = c] = \mathbb{E}[ys_t|H_t = c]$.

Assumption 1 says that the measurement error in Y_t^1 does not predict spousal income. In the absence of economically-driven measurement error, such as for example justification bias, this is a reasonable assumption. Under this exclusion restriction, we can write $\mathbb{E}[ys_t|Y_t^1 = y]$ as:

$$\mathbb{E}[ys_t|Y_t^1 = y] = \sum_{c=1}^{r} \mathbb{E}[ys_t|H_t = c]\mathbb{P}(H_t = c|Y_t^1 = y), \qquad y = 1, \ldots, \kappa.$$

This can be written as the following linear system:

$$\mathbb{P}(Y_t^1)^{-1} P^1 \Pi_t \mathbb{E}[ys_t|H_t] = \mathbb{E}[ys_t|Y_t^1],$$

where $\mathbb{E}[ys_t|H_t]$ is a column vector whose i-th element is given by $\mathbb{E}[ys_t|H_t = i]$ and $\mathbb{E}[ys_t|Y_t^1]$ is a column vector whose i-th element is given by $\mathbb{E}[ys_t|Y_t^1 = i]$. Moreover, $\mathbb{P}(Y_t^1)$ is a diagonal matrix that contains the cross-sectional distribution of Y_t^1 at t in the diagonal.

Note that the fact that all values of Y_t^1 have positive probability implies that under the full-column-rank assumption and Π_t having non-zero diagonal elements (all needed for the identification of the health process), parameters for spousal income are identified.

Estimating the parameters for spousal income can be done by using Minimum Distance to impose in the sample the restrictions implied by the population model—that is, the linear system seen above. Note that if Y^1 is chosen to be such that its cardinality is equal to r, estimation can simply be done by matrix inversion.

B.3 Wages

In the model, we assume that log wages are given by the sum of a deterministic component (a quadratic polynomial in age, a coefficient for good health, and the individual fixed effect) and a stochastic component (u). Hence,

$$\log\left(w_{it}^{\text{model}}\right) = a_0 + a_1 t + a_2 t + a_H \mathbf{1}_{\{\text{Health}_{it} = \text{Good}\}} + \eta_i + u_{it}. \tag{5}$$

The evolution of the stochastic component of wages follows an AR(1) process with normally-distributed innovations whose variance differs in the first and in subsequent periods. Formally,

$$\begin{aligned} u_{it} &= \rho u_{it-1} + \xi_t, \qquad \rho \in (0,1), \\ \xi_1 &\sim \mathcal{N}(0, \sigma^2_{\xi,1}), \\ \xi_t &\sim \mathcal{N}(0, \sigma^2_{\xi,t}), \qquad \forall t > 1. \end{aligned} \tag{6}$$

It is assumed that $\eta_i \perp \xi_t$ for $i = 1, \ldots, N$, and $t = 1, \ldots, T$.

In the data, estimate, for a given health indicator, in the sub-sample of labor market participants:

$$\log\left(\tilde{w}_{it}^{data}\right) = \tilde{\alpha}_0 + \tilde{\alpha}_1 t + \tilde{\alpha}_2 t^2 + \tilde{\alpha}_H \mathbf{1}_{\{Health_{it}=Good\}} + \tilde{\eta}_i + \underbrace{\nu_{it},}_{=u_{it}+m_{it}} \tag{7}$$

where $\tilde{w}_{it}^{data}$ is the individual hourly wage (measured with error), $\tilde{\alpha}_0$ is a constant term, $\tilde{\alpha}_1$ is the coefficient of age, $\tilde{\alpha}_2$ is the coefficient of age squared, $\tilde{\alpha}_H$ is the coefficient of good health, $\tilde{\eta}_i$ is the individual fixed effect, and ν_{it} is the error term. The error term ν_{it} is the sum of the stochastic component of wages, u_{it}, and serially-uncorrelated measurement error, $m_{it} \sim \mathcal{N}(0, \sigma^2_m)$.

The parameters of the data generating process for the stochastic component of wages and the fixed effects, $(\rho, \sigma^2_{\xi,1}, \sigma^2_{\xi,t}, \sigma^2_m, \sigma^2_\eta)$, are estimated outside the model by Minimum Distance. That is, we choose the parameter values that minimize the distance between a vector of target statistics calculated in the data and a vector of their theoretical counterparts.

At this point, it is worth mentioning that the estimated wage-shock parameters will be biased due to the selection of workers in the labor market that was discussed in section 4.3. Estimation of biased wage-shock parameters is common practice in the literature.[23,24] Despite the presence of bias, there are ways to know how large this bias is.

[23] French (2005) and O'Dea (2018) are just two examples of papers in the literature that estimate wage-shock parameters with bias.

[24] The likely reason for this is that, in order to keep the second stage of the estimation procedure tractable, wage-shock parameters must be estimated outside the model. Since there seems to be no known procedure to estimate consistent wage profiles when the error term in the value of participating enters non-linearly and is correlated with the error term in the wage equation, this leaves researchers with the non-ideal solution of using contaminated wage profiles to form the residuals that are used to estimate wage-shock parameters.

For instance, if the bias in wage-profile parameters is large, then this means that there is large bias in the residuals, which translates into large biases in wage-shock parameters. An assessment of the bias in wage-profiles parameters is possible by comparing the parameters estimated directly from the data with those estimated inside the model.

B.3.1 Identification of Wage Parameters

Let ε_{it} denote the residuals from the equation in levels (7), where

$$\varepsilon_{it} = \eta_i + u_{it} + m_{it}, \qquad t = 1, \ldots, T. \tag{8}$$

Then note that we can find a general expression for u_{it} as a function of parameters by recursion:

$$\begin{aligned}
u_{i1} &= \rho u_{i0} + \xi_1 \\
&= \xi_1, && \text{(since } u_{i0} = 0 \text{ by assumption)} \\
u_{i2} &= \rho u_{i1} + \xi_2 \\
&= \rho\xi_1 + \xi_2, \\
u_{i3} &= \rho u_{i2} + \xi_3 \\
&= \rho(\rho\xi_1 + \xi_2) + \xi_3 \\
&= \rho^2\xi_1 + \rho\xi_2 + \xi_3, \\
&\vdots \\
u_{it} &= \sum_{\tau=1}^{t} \rho^{t-\tau}\xi_\tau. && (9)
\end{aligned}$$

Substituting for u_{it} in equation (8), we have:

$$\varepsilon_{it} = \eta_i + m_{it} + \sum_{\tau=1}^{t} \rho^{t-\tau}\xi_\tau, \qquad t = 1, \ldots, T. \tag{10}$$

We can use this expression for ε_{it} to calculate:

$$\begin{aligned}
\varepsilon_{i2} - \varepsilon_{i1} &= m_{i2} - m_{i1} + \rho\xi_1 + \xi_2 - \xi_1 \\
&= m_{i2} - m_{i1} + (\rho - 1)\xi_1 + \xi_2, \\
\varepsilon_{i3} - \varepsilon_{i2} &= m_{i3} - m_{i2} + \rho^2\xi_1 + \rho\xi_2 + \xi_1 - (\rho\xi_1 + \xi_2) \\
&= m_{i3} - m_{i2} + \rho(\rho - 1)\xi_1 + (\rho - 1)\xi_2, \\
\varepsilon_{i4} - \varepsilon_{i2} &= m_{i4} - m_{i2} + \rho^3\xi_1 + \rho^2\xi_2 + \rho\xi_3 + \xi_4 - (\rho\xi_1 + \xi_2) \\
&= m_{i4} - m_{i2} + \rho(\rho^2 - 1)\xi_1 + (\rho^2 - 1)\xi_2 + \rho\xi_3 + \xi_4.
\end{aligned}$$

We can then obtain the expressions

$$\mathrm{Cov}(\varepsilon_{i4}-\varepsilon_{i2},\varepsilon_{i1}) = \rho(\rho^2-1)\sigma^2_{\xi,1} = \rho(\rho-1)(\rho+1)\sigma^2_{\xi,1} \tag{11}$$

and

$$\mathrm{Cov}(\varepsilon_{i3}-\varepsilon_{i2},\varepsilon_{i1}) = \rho(\rho-1)\sigma^2_{\xi,1}, \tag{12}$$

which, when combined, identify ρ:

$$1+\rho = \frac{\mathrm{Cov}(\varepsilon_{i4}-\varepsilon_{i2},\varepsilon_{i1})}{\mathrm{Cov}(\varepsilon_{i3}-\varepsilon_{i2},\varepsilon_{i1})}.$$

Using the expression

$$\mathrm{Cov}(\varepsilon_{i4}-\varepsilon_{i2},\varepsilon_{i1}) = \rho(\rho^2-1)\sigma^2_{\xi,1} \tag{13}$$

we can back out $\sigma^2_{\xi,1}$. We can then use the expression

$$\mathrm{Cov}(\varepsilon_{i2}-\varepsilon_{i1},\varepsilon_{i1}) = (\rho-1)\sigma^2_{\xi,1} - \sigma^2_m \tag{14}$$

to obtain σ^2_m.

It remains to show that $\sigma^2_{\xi,t}$ and σ^2_η are both identified. To do this, we first use the expression

$$\mathrm{Var}(\varepsilon_{i2}-\varepsilon_{i1}) = 2\sigma^2_m + (\rho-1)^2\sigma^2_{\xi,1} + \sigma^2_{\xi,t}, \tag{15}$$

which gives us $\sigma^2_{\xi,t}$. We then use

$$\mathrm{Var}(\varepsilon_{i2}) = \rho^2\sigma^2_{\xi,1} + \sigma^2_m + \sigma^2_\eta + \sigma^2_{\xi,t} \tag{16}$$

to obtain σ^2_η.

This proves that the parameters of the wage-shock process are identified when $T \geqslant 4$.

B.3.2 Overidentifying Restrictions

To increase the precision of the estimation by Minimum Distance, we can target additional statistics of the wage-shock process. In principle, we could obtain the following additional statistics:

$$\mathrm{Var}(\varepsilon_{it}) = \sigma_\eta^2 + \sigma_m^2 + \rho^{2(t-1)}\sigma_{\xi,1}^2 + \sum_{\tau=1}^{t-1}\rho^{2(\tau-1)}\sigma_{\xi,t}^2, \qquad t \in \{1,\dots,14\}\setminus\{2\}, \tag{17}$$

$$\mathrm{Cov}(\varepsilon_{it},\varepsilon_{i1}) = \sigma_\eta^2 + \rho^{t-1}\sigma_{\xi,1}^2, \qquad t \in \{2,\dots,14\}, \tag{18}$$

$$\mathrm{Cov}(\varepsilon_{it},\varepsilon_{i2}) = \sigma_\eta^2 + \rho^{t}\sigma_{\xi,1}^2 + \rho^{t-2}\sigma_{\xi,t}^2, \qquad t \in \{3,\dots,14\}, \tag{19}$$

$$\mathrm{Cov}(\varepsilon_{it},\varepsilon_{i3}) = \sigma_\eta^2 + \rho^{t+1}\sigma_{\xi,1}^2 + \rho^{t-3}(1+\rho^2)\sigma_{\xi,t}^2, \qquad t \in \{4,\dots,14\}, \tag{20}$$

$$\mathrm{Cov}(\varepsilon_{it+1},\varepsilon_{it}) = \sigma_\eta^2 + \rho^{2t-1}\sigma_{\xi,1}^2 + \rho\left(\sum_{\tau=1}^{t-1}\rho^{2(\tau-1)}\right)\sigma_{\xi,t}^2, \qquad t \in \{4,\dots,13\}, \tag{21}$$

$$\mathrm{Cov}(\varepsilon_{it+1}-\varepsilon_{it},\varepsilon_{it}) = \rho^{t-1}(\rho-1)\sigma_{\xi,1}^2 + \sum_{\tau=1}^{t-1}\rho^{2(\tau-1)}(\rho-1)\sigma_{\xi,t}^2 - \sigma_m^2, \qquad t \in \{2,\dots,13\}, \tag{22}$$

$$\mathrm{Cov}(\varepsilon_{it+2}-\varepsilon_{it},\varepsilon_{it}) = \rho^{2(t-1)}(\rho^2-1)\sigma_{\xi,1}^2 + \sum_{\tau=1}^{t-1}\rho^{2(\tau-1)}(\rho^2-1)\sigma_{\xi,t}^2 - \sigma_m^2, \qquad t \in \{1,\dots,12\}, \tag{23}$$

$$\mathrm{Cov}(\varepsilon_{i3}-\varepsilon_{i2},\varepsilon_{i3}) = \sigma_m^2 + \rho^3(\rho-1)\sigma_{\xi,1}^2 + \rho(\rho-1)\sigma_{\xi,t}^2, \tag{24}$$

$$\mathrm{Cov}(\varepsilon_{i2}-\varepsilon_{i1},\varepsilon_{i2}) = \sigma_m^2 + \rho(\rho-1)\sigma_{\xi,1}^2 + \sigma_{\xi,t}^2. \tag{25}$$

In total, these are 85 $(= 2\times 13 + 12 + 11 + 10 + 2\times 12 + 2)$ statistics, which together with the 5 identifying restrictions, make a total of 90 statistics. In practice, however, we can only compute 81 statistics. This is due to the fact that age groups which cluster older individuals do not have enough observations as to compute some covariances.[25]

B.3.3 Estimation of Wage Parameters by Minimum Distance

The Minimum Distance estimator is defined as:

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\ Q_n(\theta,\widehat{\pi}_n), \tag{26}$$

where

$$Q_n(\theta,\widehat{\pi}_n) = \frac{1}{n}\psi(\theta,\widehat{\pi}_n)'\widehat{\boldsymbol{W}}\psi(\theta,\widehat{\pi}_n) \tag{27}$$

is the objective function, n is the number of data targets, $\theta \in \Theta$ is the vector of parameters to be estimated, $\psi(\theta,\widehat{\pi}_n)$ is a $n\times 1$ vector that relates data targets $\widehat{\pi}_n$ to parameters, and $\widehat{\boldsymbol{W}}$ is a $n\times n$ weighting matrix.

[25]The covariances that could not be computed are: $\mathrm{Cov}(\varepsilon_{i11},\varepsilon_{i1})$, $\mathrm{Cov}(\varepsilon_{i12},\varepsilon_{i1})$, $\mathrm{Cov}(\varepsilon_{i13},\varepsilon_{i1})$, $\mathrm{Cov}(\varepsilon_{i14},\varepsilon_{i1})$, $\mathrm{Cov}(\varepsilon_{i12},\varepsilon_{i2})$, $\mathrm{Cov}(\varepsilon_{i13},\varepsilon_{i2})$, $\mathrm{Cov}(\varepsilon_{i14},\varepsilon_{i2})$, $\mathrm{Cov}(\varepsilon_{i13},\varepsilon_{i3})$, and $\mathrm{Cov}(\varepsilon_{i14},\varepsilon_{i3})$.

We estimate wage parameters $\theta = (\rho, \sigma^2_\eta, \sigma^2_{\xi,1}, \sigma^2_{\xi,t}, \sigma^2_m)$ using 81 data targets, a 81×81 identity matrix $\widehat{\boldsymbol{W}}$, and a 81×1 vector of moment conditions $\psi(\theta, \widehat{\pi}_{81})$ which exploits the identifying and over-identifying restrictions derived in sections B.3.1 and B.3.2. More specifically,

$$\psi(\theta, \widehat{\pi}_{81}) = \begin{pmatrix} \psi_1(\theta, \widehat{\pi}_{81}) \\ \vdots \\ \psi_{81}(\theta, \widehat{\pi}_{81}) \end{pmatrix} = \begin{pmatrix} \widehat{\pi}_{81,1} - h_1(\theta) \\ \vdots \\ \widehat{\pi}_{81,81} - h_{81}(\theta) \end{pmatrix},$$

where $\widehat{\pi}_{81}$ is the 81×1 vector of data targets and $\boldsymbol{h}$ is a 81×1 vector, which lists the right-hand side of expressions (11)–(25).

C Imputation of Average Earnings

To impute average earnings for individuals in our sample, we use data from the simulated model of O'Dea (2018). The implicit assumption here is that our cohort (1950–1957) is similar to his (1935–1950). The reason why we focus on individuals born between 1950 and 1957 rather than in any other cohort is that these are the individuals that we can observe in ELSA at age 50–51 and, thus, the set of individuals for which we can recover initial distributions with health indicators.

The imputation procedure is as follows. We first run the following regressions in the data simulated by O'Dea for each household type $j = 1, 2, 3, 4$:[26]

$$ae^j_{i,64} = \beta^j_0 (1 - \mathbf{1}_{\{pbb^j_{i,65} \geqslant 29.13k\}}) pbb^j_{i,65} + \beta^j_1 \mathbf{1}_{\{pbb^j_{i,65} \geqslant 29.13k\}} + \beta^j_2 \mathbf{1}_{\{pbb^j_{i,65} \geqslant 29.13k\}} privben^j_{i,65} + \varepsilon^j_i, \quad (28)$$

where $ae^j_{i,64}$ are the average earnings of individual i and type j at age 64, $\mathbf{1}_{\{pbb^j_{i,65} \geqslant 29.13k\}}$ is an indicator function which takes the value one if the public pension benefits of individual i, measured at the household-level, are greater than or equal to 29,130GBP, $privben^j_{i,65}$ are the private pension benefits of the individual, and ε^j_i is white noise.

Estimation of these regressions provides us with OLS estimates of the parameters $\{\widehat{\beta}^j_0, \widehat{\beta}^j_1, \widehat{\beta}^j_2\}_j$. We use these estimates, together with a similar household classification,

[26] O'Dea (2018) classifies households in four different types, attending to education (low or high) and whether they have access to a defined-benefit pension scheme. O'Dea considers an individual to be highly educated if he/she continued to pursue education passed age 15, the compulsory schooling age for his cohort.

to generate average earnings at age 64 according to equation (28) for the individuals in our sample.

Next, we recover average earnings at age 50. This is relatively an easy task since we can construct employment spells from the moment individuals enter the labor market up to the last year in which they are observed in ELSA.[27] Having the full history of employment spells, average earnings at age 64, and the cumulative sum of labor income during the years an individual is observed in ELSA, it is straightforward to back out average earnings at age 50 from the following identity:

$$ae^j_{i,64} = \frac{\text{Emp years}^j_{i,50} \cdot ae^j_{i,50}}{\text{Emp years}^j_{i,50} + \text{ELSA emp years}^j_i} + \frac{\text{Earnings ELSA}^j_i}{\text{Emp years}^j_{i,50} + \text{ELSA emp years}^j_i}, \quad \forall i, j, \tag{29}$$

where $\text{Emp years}^j_{i,50}$ is the number of employment years for individual i at age 50, $\text{ELSA emp years}^j_i$ is the number of employment years the individual has been employed while observed in ELSA, and Earnings ELSA^j_i is the cumulative sum of labor income while the individual is observed in ELSA.

D Identifying and Estimating Initial Distribution of States

D.1 Assumptions and Identification

Finding the initial distribution of states to initialize the model is not straightforward given that health is measured with error. In order to deal with this complication, we make the following assumptions:

1. *The initial wage shock is independent of the rest of states* $(\log a_0, \log ae_0, H_0)$.

2. $(\log a_0, \log ae_0)$ *follow some known parametric distribution given true health:*

$$\begin{pmatrix} \log a_0 \\ \log ae_0 \end{pmatrix} \sim F(\cdot, \cdot; \theta_H).$$

where θ_H is the vector of parameters on which the joint distribution of initial assets and average earnings, given health H, depends.

[27] To construct employment spells for time periods prior to the ELSA survey period, we leverage on the data collected in the ELSA Life History Interview, which asks individuals for "start" and "end" dates for each job they have had since they entered the labor market until ELSA wave 3.

Under these assumptions, we can identify the initial distribution of the states from data on wages, health measures, average earnings, and non-housing wealth at age 50. A few comments on identification are in order. First, the variance of the initial wage shock is identified from the identification proof outlined before, and can be estimated using the Minimum Distance procedure described in Appendix B.3.2. Second, the conditional distribution of initial average earnings and initial wealth is identified from data on health measures, average earnings, and health measures. To see why, note that the random vector composed by average earnings and wealth is independent of noisy health measures given true health. Then, from Bonhomme et al. (2016), its conditional distribution, given health, is non-parametrically identified. It follows that under the stronger parametric assumption of joint log-normality it is also identified. Finally, the marginal distribution of health is identified (see Appendix A).

D.2 Parametric Assumptions

We make the following parametric assumptions:

1. $\log a_0 | H = 1 \sim \mathcal{N}(\mu_{a,1}, \sigma_{a,1})$.

2. $\log a_0 | H = 2 \sim ALD(\mu_{a,2}, \sigma_{a,2}, p_{a,2})$, where $ALD()$ denotes the assymetric Laplace distribution proposed by Yu and Zhang (2005).

3. $\log ae_0 | H = h \sim \mathcal{N}(\mu_{ae,h}, \sigma_{ae,h})$ for $h = 1, 2$.

4. The joint dependence of $\log a_0$ and $\log ae_0$ given $H = h$ is given by a Gaussian Copula with rank-correlation parameter ρ_h.

D.3 Estimation Algorithm

In order to estimate the parameters of the joint distribution of initial assets and average earnings we use a two-step procedure. In the first step, we estimate separately the parameters of the marginal distribution of assets given true health and those of the marginal distribution of average earnings, also given true health. In the second step, we estimate the parameters of the copula given true health using the parameters of the marginals as inputs.

This idea of estimating separately parameters of the marginals and parameters of the copula is not new in statistics. In particular, the Inference-Functions-for-Margins (see Joe, 1997) and the Maximization-by-Parts algorithms (see Song et al., 2005) make use of this idea. However, the algorithm used here is not a particular case of any of those two algorithms because we are estimating the marginal distribution of log-assets and log-average earnings conditional on an unobserved variable (health).

D.3.1 Estimating the Marginal of Assets given Health

Let $f_{a,H}(\log a_0; \theta_{a,H})$ be the density of assets given that health is H. Given our previous assumptions, $f_{a,1}(\log a_0; \theta_{a,1})$, where $\theta_{a,1} = (\mu_{a,1} \sigma_{a,1})$, is the density of a normal with mean $\mu_{a,1}$ and variance $\sigma_{a,1}$. Similarly, $f_{a,2}(\log a_0; \theta_{a,2})$ is the distribution of initial log-assets given $H = 2$. Moreover, $\theta_{a,2} = (\mu_{a,2}, \sigma_{a,2}, p_{a,2})$ and $f_{a,2}(\log a_0; \theta_{a,2})$ is given by:

$$f_{a,2}(\log a_0; \theta_{a,2}) = \frac{p_{a,2}(1 - p_{a,2})}{\sigma_{a,2}} \exp\left(-\frac{\log a_0 - \mu_{a,2}}{\sigma_{a,2}} [p_{a,2} - 1(\log a_0 \leqslant \mu_{a,2})] \right).$$

We estimate $\theta_{a,1}$ and $\theta_{a,2}$ by Maximum Likelihood. In order to find the maximum likelihood estimates, we use the following EM procedure:

- **E step:** Given guesses $\theta_{a,1}^{(l)}, \theta_{a,2}^{(l)}$ calculate the filtered probabilities:

$$\hat{\tau}_i^1 = \mathbb{P}(H_i = 1 | a_{i0}, \{Y_i^m\}_{m=1}^3) = \frac{\hat{\pi}_0(1) f_{a,1}(a_{i0}; \theta_{a,1}^{(l)}) \hat{P}^1(y_i^1, 1) \hat{P}^2(y_i^2, 1) \hat{P}^3(y_i^3, 1)}{\text{Denominator}_i},$$

  where:

$$\begin{aligned} \text{Denominator}_i = &\hat{\pi}_0(1) f_{a,1}(a_{i0}; \theta_{a,1}^{(l)}) \hat{P}^1(y_i^1, 1) \hat{P}^2(y_i^2, 1) \hat{P}^3(y_i^3, 1) \\ &+ \hat{\pi}_0(2) f_{a,2}(a_{i0}; \theta_{a,2}^{(l)}) \hat{P}^1(y_i^1, 2) \hat{P}^2(y_i^2, 2) \hat{P}^3(y_i^3, 2). \end{aligned}$$

- **M step:** Given the distributional assumptions, $\theta_{a,1}$ and $\theta_{a,2}$ are updated differently.

  1. In good health, the updated parameter values are simply the weighted mean and variance of log-assets, where the weight of individual i is given by $\hat{\tau}_i^i$:

$$\mu_{a,1}^{(l+1)} = \frac{\sum_{i=1}^N \hat{\tau}_i^1 \log a_{i,0}}{\sum_{i=1}^N \hat{\tau}_i^1},$$
$$\sigma_{a,1}^{(l+1)} = \frac{\sum_{i=1}^N \hat{\tau}_i^1 (\log a_{i,0} - \mu_{a,1}^{(l+1)})^2}{\sum_{i=1}^N . \hat{\tau}_i^1}.$$

2. Let $\rho_p(t) = t(p - 1(t < 0))$. Find $\theta_{a,2}^{(l+1)}$. Then, given an arbitrary initial guess for $p_{a,2}$, call it $p^{(0)}$, find $\theta_{a,2}^{(l+1)}$ iterating on the following equations:

$$\mu^{r+1} = \arg\min_{\mu} \frac{\sum_{i=1}^{N} \hat{\tau}_i^2 \rho_{p^r}(\log a_{i,0} - \mu)}{\sum_{i=1}^{N} \hat{\tau}_i^2},$$
$$\sigma^r = \frac{\sum_{i=1}^{N} \hat{\tau}_i^2 \rho_{p^r}(\log a_0 - \mu^{r+1})}{\sum_{i=1}^{N} \hat{\tau}_i^2},$$
$$p^{r+1} = \frac{a + \sqrt{a^2 - (\overline{x} - \mu^r)}}{\overline{x} - \mu^r},$$

where

$$\overline{x} = \frac{\sum_{i=1}^{N} \hat{\tau}_i^2 \log a_{i,0}}{\sum_{i=1}^{N} \hat{\tau}_i^2} \quad \text{and} \quad a = \frac{\sum_{i=1}^{N} (\log a_{i,0} - \mu^r) 1(\log a_{i,0} \leqslant \mu^r)}{\sum_{i=1}^{N} \hat{\tau}_i^2}.$$

Note that this is just a weighted version of the algorithm described in Yu and Zhang (2005), where the weights are given by the $\hat{\tau}_i^2$.

D.3.2 Estimating the Marginal of Average Earnings given Health

Let $f_{ae,H}(\log ae_0, \theta_{ae,H})$ be the density of log-average earnings given that health is H. Given our previous assumptions, $f_{ae,H}(\log ae_0, \theta_{ae,H})$, where $\theta_{ae,H} = (\mu_{ae,H}, \sigma_{ae,H})$, is the density of a normal with mean $\mu_{ae,H}$ and variance $\sigma_{ae,H}$. Just as in the case of assets, we estimate the parameters of the marginal of log-average earnings given health using Maximum Likelihood. In order to find the maximum likelihood estimates, we use the following EM procedure:

- **E step:** Given guesses $\theta_{ae,1}^{(l)}, \theta_{ae,2}^{(l)}$ calculate the filtered probabilities:

$$\hat{\tau}_i^1 = \mathbb{P}(H_i = 1 | ae_{i0}, \{Y_i^m\}_{m=1}^3) = \frac{\hat{\pi}_0(1) f_{ae,1}(ae_{i0}; \theta_{ae,1}^{(l)}) \hat{P}^1(y_i^1, 1) \hat{P}^2(y_i^2, 1) \hat{P}^3(y_i^3, 1)}{\text{Denominator}_i},$$

  where:

$$\begin{aligned} \text{Denominator}_i = &\hat{\pi}_0(1) f_{ae,1}(ae_{i0}; \theta_{ae,1}^{(l)}) \hat{P}^1(y_i^1, 1) \hat{P}^2(y_i^2, 1) \hat{P}^3(y_i^3, 1) \\ &+ \hat{\pi}_0(2) f_{ae,2}(ae_{i0}; \theta_{ae,2}^{(l)}) \hat{P}^1(y_i^1, 2) \hat{P}^2(y_i^2, 2) \hat{P}^3(y_i^3, 2). \end{aligned}$$

- **M-step** The updated parameter values are simply the weighted mean and variance of log-assets, where the weight of individual i is given by $\hat{\tau}_i^i$:

$$\mu_{ae,1}^{(l+1)} = \frac{\sum_{i=1}^{N} \hat{\tau}_i^1 \log ae_{i,0}}{\sum_{i=1}^{N} \hat{\tau}_i^1},$$
$$\sigma_{ae,1}^{(l+1)} = \frac{\sum_{i=1}^{N} \hat{\tau}_i^1 (\log ae_{i,0} - \mu_{ae,1}^{(l+1)})^2}{\sum_{i=1}^{N} \cdot \hat{\tau}_i^1}.$$

D.3.3 Estimating the Copula Parameters for Good and Bad Health

In this step of the algorithm, we maximize the log-likelihood of the dataset:

$$\{\{Y_i^m\}_{m=1}^3, \log a_{i,0}, \log ae_{i,0}\}_{i=1}^N.$$

In doing so, we take as given the estimated parameters for the joint distribution of health and health measurements, $(\pi_0, \{P^m\}_{m=1}^3)$, and the parameters of the marginal distributions of initial log-assets and initial log-average earnings given health, $\{\theta_{a,H}, \theta_{ae,H}\}_{H=1,2}$. Hence, the only parameters that remain to be estimated at this stage are the rank-correlation parameters of the (Gaussian) copula for log-assets and log-average earnings given health; that is, ρ_H for $H = 1,2$. In order to maximize the log-likelihood with respect to these parameters we use the following EM algorithm:

- **E step:** Given guesses for the rank-correlation parameters $\rho_H^{(l)}$, $H = 1,2$, calculate the filtered probabilities:

$$\begin{aligned}\tau_i^1 &= \mathbb{P}(H_i = 1 | a_{i,0}, ae_{i,0}, \{Y_i\}_{m=1}^3) \\ &= \hat{\pi}_0(1)\hat{P}^1(y_i^1, 1)\hat{P}^2(y_i^2, 1)\hat{P}^3(y_i^3, 1) \\ &\quad \times c(F_{a,1}(\log a_i; \hat{\theta}_{a,1}), F_{ae,1}(\log ae_i; \hat{\theta}_{ae,1}))f_{a,1}(\log a_i; \hat{\theta}_{a,1})f_{ae,1}(\log ae_i; \hat{\theta}_{ae,1})/\text{denominator}_i,\end{aligned}$$

  where

$$\begin{aligned}\text{denominator}_i = &\hat{\pi}_0(1)\hat{P}^1(y_i^1, 1)\hat{P}^2(y_i^2, 1)\hat{P}^3(y_i^3, 1) \\ &\times c(F_{a,1}(\log a_i; \hat{\theta}_{a,1}), F_{ae,1}(\log ae_i; \hat{\theta}_{ae,1}))f_{a,1}(\log a_i; \hat{\theta}_{a,1})f_{ae,1}(\log ae_i; \hat{\theta}_{ae,1}) \\ &+ \hat{\pi}_0(2)\hat{P}^1(y_i^1, 2)\hat{P}^2(y_i^2, 2)\hat{P}^3(y_i^3, 2) \\ &\times c(F_{a,2}(\log a_i; \hat{\theta}_{a,2}), F_{ae,2}(\log ae_i; \hat{\theta}_{ae,2}))f_{a,2}(\log a_i; \hat{\theta}_{a,2})f_{ae,2}(\log ae_i; \hat{\theta}_{ae,2}).\end{aligned}$$

  Here, $c()$ denotes the density function associated to the Gaussian Copula, and $F_{a,H}(\cdot;\cdot)$ and $F_{ae,H}(\cdot;\cdot)$ denote the cdf of log-assets and log-average earnings when true health is H.

- **M step:** In this step we are seeking to maximize the expected completed log-likelihood given the data with respect to the rank-correlation parameters (ρ_1, ρ_2). Because of the gaussianity of the copula, solving the score equation of this expected log-likelihood amounts to solving the following pair of cubic equations (see, for example, Amengual and Sentana, 2015):

$$\rho_H^3 - B_H\rho_H^2 + (A_H - 1)\rho_H - B_H = 0,$$

where

$$A_H = \frac{\sum_{i=1}^{N} \tau_i^H Z_i^{a,H} Z_i^{ae,H}}{\sum_{i=1}^{N} \tau_i^H},$$

$$B_H = \frac{\sum_{i=1}^{N} \tau_i^H \left((Z_i^{a,H})^2 + (Z_i^{ae,H})^2 \right)}{\sum_{i=1}^{N} \tau_i^H},$$

and

$$Z_i^{a,H} = \Phi^{-1}(F_{a,H}(\log a)),$$
$$Z_i^{ae,H} = \Phi^{-1}(F_{ae,H}(\log ae)),$$

for $H = 1, 2$. Since these equations can have up to three real roots, we solve for all of them numerically and pick the one that yields the highest expected complete log-likelihood. The resulting pair of solutions gives us the new guess $\left(\rho_1^{(l+1)}, \rho_2^{(l+1)}\right)$.

E Computational Appendix

The model is solved by iterating the value function backwards. More concretely, the right-hand side of the functional equation (30) at each t has to be maximized choosing over consumption and hours worked:

$$V_t(H, a, ae, \eta) = \max_{a', N} \quad u\left(c,\ \bar{L} - \phi_P \mathbf{1}_{\{N>0\}} - N - \phi_H \mathbf{1}_{\{H=Bad\}}\right) + (1 - s(H,t))b(a') \qquad (30)$$
$$+ \beta s(H,t) \mathbb{E} V_{t+1}(\cdot, a', ae, \cdot)$$
$$\text{s.t} \qquad c + a' = y(ra_t + w_t N_t) + ys(t, H) + tr_t + a(1+r) \qquad (31)$$

E.1 Discretizing States

We use equally-spaced grids for the continuous states, assets, average earnings, and the wage shock. Then we solve the RHS of the value function for each state on the grid and each t.

E.2 Solving the right-hand side of the value function

In order to solve the right-hand side of the value function, we proceed as follows:

1. We focus on the share of consumption and the share of hours worked as a fraction of total resources and total available hours, respectively. We create a grid between

0 and 1 for each of these choices, and we evaluate the right-hand side of the value function (30) at each of these choices.

2. If the choice today implies an state tomorrow that is outside of the grid for the states, we use linear interpolation to evaluate the expected value function at $t+1$.

3. Once we have evaluated the right-hand side of the functional equation on each point of the choice grid, we approximate the right-hand side outside the grid as the linear interpolant of the right-hand side of the value function on the grid.

4. We maximize this object using a global optimization strategy.

E.3 Global Optimization Strategy

In order to maximize the interpolated right-hand side of the functional equation, we start by sampling pseudo-random shares of consumption and hours worked using Sobol sequences. At each point of the Sobol sequence we search for a local maximum using a Nelder–Mead algorithm. Since this only finds interior local maxima, we deal with the corners separately.

F Details on Estimation by Indirect Inference

In this appendix, we provide more details on the indirect inference estimation procedure.

F.1 Weighting matrix

The weighting matrix is given by a diagonal matrix that contains the inverse of the squares of the targets calculated in the data. More precisely, let ψ_{in} be the i-th target calculated in the data. Then, our weighting matrix is given by:

$$W_N = \begin{bmatrix} \frac{1}{\psi_{1n}^2} & & \\ & \ddots & \\ & & \frac{1}{\psi_{Jn}^2} \end{bmatrix},$$

where J is the total number of targets, which is equal to:

$$J = 6 * T + 4.$$

This number of targets comes from the fact that we are targeting assets, participation and hours for the healthy and the unhealthy (hence the 6) for T age groups (hence the term T) and 4 regression coefficients corresponding to the deterministic wage profiles (hence the +4 term).

F.2 Minimizing the Indirect Inference Objective

We use a heuristic global minimization algorithm to minimize the the indirect inference objective. Our minimization procedure consists of three steps:

1. In the first step, we take 10,000 elements from a Sobol sequence defined over the parameter space. We burn-in the first 9,000 and keep the remaining 1,000. We evaluate the value of the indirect inference objective at these 1,000 candidates in parallel across 4 processes.

2. From this initial Sobol search, we take the 24 best candidates (in terms of accomplishing a small objective). Let the j-th candidate be denoted by $\theta^{(j)}$, and let $\theta_i^{(j)}$ denote the i-th coordinate of $\theta^{(j)}$. Using these 24 candidates, we define a box of the form:
$$\bigtimes_{i=1}^{12} \left[\min_j \theta_i^{(j)}, \max_j \theta_i^{(j)} \right].$$
Following a similar procedure as before, we take 10,000 parameter values from a Sobol sequence defined inside this box, burn-in the first 9,000, and keep the remaining 1,000. Again, we evaluate the objective at these 1,000 candidates in parallel across four processes.

3. The last step of the procedure is a polishing step. We take the 4 best candidates from the Sobol search in step 2 and initialize a local minimization step of the indirect inference objective in each of them. The minimization algorithm that we use to conduct the local minimization step is the BOBYQA algotithm proposed by Powell (2009). Again, each of these local searches are conducted in parallel across four processes. Finally, our global minimizer is the one corresponding to the best local minimum from these four local searches.