Bad Health Is Worse Than It Looks: Measurement Error and Structural Inference*

Martín García-Vázquez — Washington University in St. Louis

Yuta Maeda — University of Minnesota

Luis Pérez — Southern Methodist University

November 6, 2026

Abstract

Economists often treat self-reported health as if it perfectly revealed the underlying state. We show that this makes health appear less persistent, attenuates its effects on economic behavior, and can make models seem to require persistent heterogeneity. We estimate a life-cycle model of labor supply and saving on English Longitudinal Study of Ageing data, inferring latent health from noisy indicators and accounting for measurement error throughout structural estimation. Correcting measurement error raises the earnings cost of bad health by 44 percent overall and more than doubles it for the initially unhealthy. These results reflect underestimated health persistence and attenuated labor-supply differences across health states. Without preference heterogeneity or persistent health types, the latent-health model reproduces about 80 percent of the wealth–health gradient, versus about 20 percent when self-reported health is treated as the state. Ignoring measurement error can understate health costs and overstate the apparent need for persistent heterogeneity.

Keywords: Measurement error, health, life-cycle models, hidden Markov models.
JEL codes: I14, D15, C38, E21.

*This paper previously circulated as "The Impact of Measurement Error in Health on Health-Related Counterfactuals." We thank Jacob Adenbaum, V. V. Chari, Mariacristina De Nardi, Loukas Karabarbounis, Jeremy Lise, Joseph Mullins, and Zhixiu Yu, and seminar participants at the Barcelona Summer Forum, SMU, and the University of Minnesota for helpful comments and suggestions. We are also grateful to Cormac O'Dea for providing us with simulations from the model in his job-market paper. Errors are our own. Emails: garciavazquez@wustl.edu, maeda025@umn.edu, luisperez@smu.edu.

1 Introduction

Health shapes how much people work, earn, and save, and how long they live.[1] Quantifying these effects is central to understanding inequality, retirement, social insurance, and the economic costs of poor health. Health is not directly observable, yet economists often treat self-reported health as if it revealed the underlying state without error. In this paper, we show that this seemingly innocuous shortcut can substantially change what structural models tell us about health and its economic consequences. Treating reported health as the true state makes bad health appear less persistent and leads researchers to underestimate its effects on economic behavior. These distortions can also make persistent differences across individuals appear more important than they are. In our application, accounting for measurement error more than doubles the estimated earnings cost of bad health for individuals already unhealthy at age fifty and allows a model without preference heterogeneity or persistent health types to reproduce most of the observed wealth gap between healthy and unhealthy individuals.

The reason is straightforward. Some observed movements into and out of bad health reflect measurement error rather than genuine changes in health, so treating reported health as the underlying state makes health shocks appear more transitory than they are. At the same time, misclassification mixes healthy and unhealthy individuals, making differences in their labor supply, earnings, and other economic outcomes appear smaller than they actually are. The distinction matters because these two distortions affect different aspects of structural inference. Understating persistence makes the effects of bad health appear shorter-lived, while understating differences in behavior leads the model to infer that being unhealthy is less costly. Measurement error therefore affects not only the estimated evolution of health, but also the economic parameters that govern how individuals respond to it.

We quantify these effects using a dynamic life-cycle model of labor supply and saving for individuals age fifty and older. Health takes two values, good and bad, and affects survival, labor productivity, the time available for work and leisure, and future health. Individuals choose consumption, saving, and labor supply while facing health and productivity risk. We use the English Longitudinal Study of Ageing (ELSA) and model the UK institutional environment. The UK's universal health care system allows us to study the effects of health on labor supply, earnings, saving, and longevity without the additional complications of employer-provided health insurance and substantial out-of-pocket medical expenditure risk.

[1] See, among others, French (2005), De Nardi, French and Jones (2010), French and Jones (2011), Gustman and Steinmeier (2014), Capatina (2015), Low and Pistaferri (2015), Bueren (2021), Borella et al. (2025), Russo (2026), and Hosseini, Kopecky and Zhao (2026). A related reduced-form literature examines the effects of health on wealth accumulation (Smith, 1999; Wu, 2003; Smith, 2004; Poterba, Venti and Wise, 2017) and labor supply (Bound, 1991; Blundell et al., 2023).

Our empirical strategy is to hold the economic environment and structural model fixed while changing only how health is measured and incorporated into estimation. In the first specification, as in much of the structural literature, self-reported health is taken to be the health state. In the second, health is latent and self-reported health, mobility conditions, and limitations in activities of daily living and instrumental activities of daily living (ADL/IADL) are treated as noisy measures of that state. We recover the latent health process and its measurement system using a non-stationary hidden Markov model. Crucially, we account for measurement error not only when estimating health transitions, but also when constructing initial conditions, estimating economic outcomes conditional on health, and using those outcomes to identify the model's behavioral parameters.

Our first finding is that bad health is substantially more persistent than self-reported health suggests, with important implications for wealth inequality. Accounting for measurement error raises the estimated probability of remaining in bad health over two years from 55–84 percent to more than 90 percent at every age. Averaged across age groups, the model with latent health reproduces roughly 80 percent of the observed median wealth gap between healthy and unhealthy individuals, despite having neither heterogeneity in discount factors nor persistent health types. The otherwise comparable model that treats self-reported health as the true state reproduces only about 20 percent. Much of this improvement comes from correcting health persistence: replacing only the health process estimated from self-reports with the latent-health process raises the share of the observed wealth gap explained by the model to 70 percent. Thus, a more persistent underlying health process can explain much of the wealth–health gradient without additional persistent heterogeneity.

Greater persistence in underlying health also helps explain why the longer individuals report being in bad health, the less likely they are to report a recovery. We show that a first-order latent health process observed with error generates negative duration dependence without health types. The intuition is simple: an individual who reports bad health once may simply be misclassified, but repeated reports of bad health progressively raise the probability that the individual is truly unhealthy. Because latent bad health is persistent, the probability of subsequently reporting good health falls with the duration of the reported spell. Negative duration dependence in measured health need not, by itself, imply permanent heterogeneity in health.

Our second finding is that accounting for measurement error substantially increases the estimated economic costs of bad health. In both specifications, we compare the baseline economy, where individuals face the estimated health risks, with a counterfac-

tual in which they remain in good health.[2] We define the earnings cost of bad health as the resulting difference in average annual earnings of the household head through ages 64–65. We separately examine hours worked, assets, years of life, and a broader measure of remaining-life monetary costs that also incorporates spousal income, pensions, and income lost through premature mortality.

The differences are large. For the population as a whole, the estimated earnings cost of bad health is £2,554 per year when health is treated as latent, 44 percent larger than the £1,778 implied by treating self-reported health as the true state. Among individuals already unhealthy at entry, the cost is £11,036 rather than £5,107 per year, more than twice as large. The longevity gain from remaining in good health is about one year larger under latent health for the average individual and about four years larger for those initially unhealthy. Estimated effects on hours worked and wealth accumulation are also substantially larger. Treating self-reported health as the true state therefore leads the model to substantially underestimate the economic costs of bad health.

The decomposition reveals why. Accounting for measurement error produces both a substantially more persistent process for bad health and a larger estimated direct burden of being unhealthy. In particular, the estimated two-year time cost of bad health rises by 41 percent, from 1,230 to 1,735 hours (about 17 hours per week). These two changes matter for different counterfactuals. For the population as a whole, correcting the attenuation in the health-related labor-supply differences used to estimate the model accounts for essentially the entire difference in foregone earnings. For individuals who enter the model in bad health, however, persistence plays the dominant role: replacing the health process estimated from self-reports with the latent-health process accounts for 77 percent of the difference in remaining-life monetary costs and essentially the entire difference in longevity. Most individuals enter the model in good health, whereas the costs for those already unhealthy depend especially strongly on how long their bad health persists.

These results show why correcting the health process alone is not enough. Measurement error also affects the behavioral moments used to estimate the model, and correcting them can substantially change the estimated economic consequences of bad health. Moreover, the relative importance of these corrections depends on the counterfactual being considered.

Related Literature. Our paper contributes to research on how health is measured and how its dynamics affect economic outcomes. Several papers have recognized that

[2] Capatina (2015) and De Nardi, Pashchenko and Porapakkarm (2025) also measure the costs of bad health by comparing individuals who face the estimated health process with counterfactual individuals who remain healthy.

health is measured imperfectly.[3] Bound (1991) shows that how health is measured affects estimated labor-supply responses, while French (2005) and Bueren (2021) allow for measurement error within structural life-cycle models. White (2023) shows that self-reported health understates the persistence of latent health. Relatedly, Dal Bianco and Moro (2026) use the same English data to construct a continuous health index and show that allowing health dynamics to vary with the level of health has important implications for saving, welfare, and disability insurance.

Our closest methodological connection is to Amengual, Bueren and Crego (2021), who estimate a non-stationary hidden Markov model using multiple noisy health indicators and examine the implications of measurement error for the asset costs of bad health. They recover latent health dynamics but take preference parameters from De Nardi, French and Jones (2010), whose model was estimated treating reported health as the health state. We instead account for measurement error throughout structural estimation, including the initial distribution of individuals, economic outcomes conditional on health, and the moments used to estimate behavioral parameters. This distinction matters quantitatively. Re-estimating the economic model raises the estimated time cost of bad health by 41 percent, and our decomposition shows that correcting health persistence alone is insufficient to recover the full economic consequences of bad health.

Our findings also contribute to the literature on persistent heterogeneity in health and wealth. De Nardi, Pashchenko and Porapakkarm (2025) study the United States in an environment with medical expenditure risk and show that persistent health types correlated with patience help their model reproduce the wealth–health gradient and the negative duration dependence of reported health. Our results suggest a complementary explanation: a persistent underlying health state observed with error can generate much of the same evidence without permanent health types or preference heterogeneity. Our results do not rule out permanent differences in health or preferences, but they show that these empirical patterns do not, by themselves, establish the need for such heterogeneity.

Our analysis also contributes to the identification and estimation of latent health dynamics. We establish identification of a non-stationary latent health process and its measurement system in the presence of mortality, allow health transitions to vary flexibly with age, and show how to incorporate latent health consistently into the different stages of structural estimation. A central identifying assumption is that, conditional on latent health, the health indicators contain no additional information about the economic outcomes used in estimation. An important concern is that individuals may

[3] A related literature documents reporting error in self-reported health, including justification bias and state-dependent reporting (Kerkhofs and Lindeboom, 1995; Dwyer and Mitchell, 1999; Benítez-Silva et al., 2004; Lindeboom and Kerkhofs, 2009). Lange and McKee (2012) study the evolution of latent health over the life course.

report their health differently depending on their labor-market circumstances. We therefore re-estimate the entire model replacing self-reported health with reported pain while retaining mobility and ADL/IADL limitations as the other two indicators. The main quantitative conclusions remain.

The broader lesson extends beyond health.[4] When a persistent economic state is observed only through a noisy proxy, treating the proxy as the state can make persistent shocks look transitory, understate how strongly behavior responds to the underlying state, and create the appearance that additional persistent heterogeneity is needed to account for the data. Measurement error in a persistent state does not merely add noise: it can conceal persistence and create the appearance of heterogeneity that the underlying economic environment need not contain.

The rest of the paper is organized as follows. Section 2 presents the economic model. Section 3 describes the data and estimation strategy, including how we recover latent health from noisy measures and incorporate it into structural estimation. Section 4 examines the implications of measurement error for health persistence, duration dependence, and the wealth–health gradient. Section 5 quantifies the economic costs of bad health, decomposes the role of measurement error, and presents robustness and policy exercises. Section 6 concludes.

2 Economic Model

We study the labor supply and saving decisions of older male household heads in a canonical life-cycle model with health and earnings risk, following French (2005). To facilitate comparisons with the structural literature, we retain this framework and incorporate UK taxes, transfers, and pensions following O'Dea (2018).[5] A model period spans two years, consistent with the biennial frequency of the English Longitudinal Study of Ageing (ELSA). Periods $t = 1, \ldots, T$ correspond to ages 50–51 through 90–91, with $T = 21$. Individuals enter the model at their first observed age, t_0, and cannot survive beyond the final period.

Health affects economic decisions through three channels: it determines survival probabilities and future health, affects wages, and reduces the time available for work and leisure. Individuals observe their own health, which evolves according to an age-dependent Markov process.

[4] Hidden Markov models have also been used to explain state dependence in measured unemployment (see Shibata, 2019).

[5] An advantage of studying the UK is its universal health care system, which provides most medical care free at the point of use. This allows us to focus on the economic consequences of health without explicitly modeling health insurance or substantial out-of-pocket medical expenditure risk.

2.1 Preferences

In each period, a household head chooses consumption C_t, hours worked N_t, and next-period assets a_{t+1} to maximize expected lifetime utility. For an individual entering in period t_0, expected lifetime utility is

$$\mathbb{E}_{t_0}\left[\sum_{t=t_0}^{T}\beta^{t-t_0}S_t\{U(C_t,L_t)+\varpi+[1-s_t(H_t)]b(a_{t+1})\}\right], \qquad S_t\equiv\prod_{k=t_0}^{t-1}s_k(H_k), \quad S_{t_0}=1. \tag{1}$$

Here β is the discount factor, $s_t(H_t)$ is the probability of surviving from t to $t+1$ conditional on current health, and S_t is the cumulative survival weight along a health history. An individual alive in period t obtains utility $U(C_t,L_t)$ from consumption and leisure, together with an additive value of being alive, ϖ. Death at the end of the period occurs with probability $1-s_t(H_t)$ and yields warm-glow bequest utility $b(a_{t+1})$. We impose $s_T(H_T)=0$.

Flow utility takes the form

$$U(C_t,L_t)=\frac{\left(C_t^{\nu}L_t^{1-\nu}\right)^{1-\gamma}}{1-\gamma}, \tag{2}$$

where $\nu\in(0,1)$ governs the relative weight of consumption and leisure, and γ governs utility curvature.[6] Following De Nardi, Pashchenko and Porapakkarm (2025), we include ϖ to account for the welfare value of additional years of life. Because mortality is exogenous and ϖ enters additively, it does not affect household decisions. We therefore normalize $\varpi=0$ during estimation and calibrate it afterward to match a specified value of a statistical life (Section 5.1).

Leisure satisfies

$$L_t=\bar{L}-\phi_H\mathbf{1}\{H_t=B\}-\phi_P\mathbf{1}\{N_t>0\}-N_t, \tag{3}$$

where $\bar{L}$ is the time endowment. The fixed participation cost ϕ_P captures time spent commuting or preparing for work, helping the model distinguish the extensive and intensive margins of labor supply. The time cost ϕ_H reduces the time available for work and leisure when health is bad, regardless of whether the individual works.

[6]The numerical solution uses normalized consumption–leisure and bequest utility functions that admit logarithmic limits at $\gamma=1$. With exogenous mortality, these normalizations leave optimal choices unchanged but affect utility levels and the calibration of the value of being alive. The welfare analysis uses the corresponding unnormalized expressions; see Section 5.1.

Following De Nardi (2004), warm-glow bequest utility is

$$b(a) = \theta_B \frac{(\kappa_B + a)^{(1-\gamma)\nu}}{1-\gamma}, \tag{4}$$

where θ_B governs the strength of the bequest motive and $\kappa_B > 0$ governs its curvature and ensures finite marginal bequest utility at zero assets under the maintained parameter restrictions.

2.2 Health and Labor Income

Health is exogenous and follows an age-dependent Markov process with two living states, good (G) and bad (B), and an absorbing death state (D). The transition probabilities are

$$\Pr(H_{t+1} = i \mid H_t = j) = \Pi_t(j,i), \qquad i,j \in \{G,B,D\}. \tag{5}$$

The age dependence of Π_t allows health persistence and mortality to vary over the life cycle. The probability of surviving to the next period is

$$s_t(j) = 1 - \Pi_t(j,D), \qquad j \in \{G,B\}. \tag{6}$$

Labor earnings are $W_t N_t$, where the log wage follows

$$\log W_t = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}\{H_t = G\} + u_t. \tag{7}$$

The coefficient a_H captures the good-health wage premium, while the idiosyncratic productivity component u_t follows an AR(1) process:

$$\begin{aligned} u_t &= \rho u_{t-1} + \xi_t, \qquad u_1 = \xi_1, \\ \xi_t &\sim \begin{cases} \mathcal{N}(0, \sigma^2_{\xi,1}), & t = 1, \\ \mathcal{N}(0, \sigma^2_{\xi}), & t \geqslant 2. \end{cases} \end{aligned} \tag{8}$$

Following French and Jones (2011), we incorporate spousal income in reduced form rather than modeling the spouse's labor-supply decision. Net spousal income is a deterministic function of the household head's age and health:

$$ys_t = ys(t, H_t). \tag{9}$$

The function includes the spouse's employment and self-employment income and the government benefit for caregivers, but excludes pensions and means-tested trans-

fers, which enter separately. It applies throughout the life cycle, including after the household head reaches retirement age.

2.3 Assets, Pensions, and Government Transfers

Households save in a risk-free asset a_t earning a net return r per two-year period, subject to a no-borrowing constraint. The age-dependent function $y_t(\cdot)$ maps the household head's taxable income into after-tax income, incorporating features of the UK tax system (see O'Dea, 2018, Appendix D.7). Public and private pension payments begin at retirement, R_a, corresponding to ages 64–65.

Pension payments depend on average lifetime earnings at retirement, ae_{R_a}. Denoting the public and private pension benefit functions by $P^{pub}(\cdot)$ and $P^{priv}(\cdot)$, respectively, which we estimate in Appendix B.5, payments are

$$\begin{aligned} pbb_t &= \mathbf{1}\{t \geqslant R_a\} P^{pub}(ae_{R_a}), \\ privben_t &= \mathbf{1}\{t \geqslant R_a\} P^{priv}(ae_{R_a}). \end{aligned} \tag{10}$$

Average lifetime earnings evolve according to

$$ae_{t+1} = \begin{cases} \dfrac{W_t N_t + (n_0 + t - 1)(1+g)^2 ae_t}{n_0 + t}, & t < R_a, \\ ae_t, & t \geqslant R_a. \end{cases} \tag{11}$$

Past earnings are revalued at the annual rate g before current earnings are incorporated, and the resulting average is held fixed after retirement. We set $n_0 = 15$, the number of two-year periods between labor market entry at age 20 and age 50, and $g = 0.0184$, the revaluation rate in the data simulated by O'Dea (2018). The earnings accumulation and initial-earnings conventions are detailed in Appendix C.1.

Following French and Jones (2011) and De Nardi, Pashchenko and Porapakkarm (2025), the government guarantees a minimum level of resources through a consumption floor C_{min}, which is constant across ages. Transfers are

$$tr_t = \max\left\{0, C_{min} - \left[y_t(W_t N_t + r a_t + pbb_t + privben_t) + ys_t + a_t\right]\right\}. \tag{12}$$

Combining earnings, pensions, asset returns, spousal income, and transfers, the household's budget constraint is

$$C_t + a_{t+1} = y_t(r a_t + W_t N_t + pbb_t + privben_t) + ys(t, H_t) + tr_t + a_t, \tag{13}$$

subject to $a_{t+1} \geqslant 0$, nonnegative hours, and feasible consumption and leisure.

2.4 Recursive Household Problem

The household's state is

$$X_t = (H_t, a_t, ae_t, u_t), \tag{14}$$

consisting of health, assets, average lifetime earnings, and the persistent wage shock. The recursive problem is

$$V_t(X_t) = \max_{a_{t+1}, N_t} \Big\{ U(C_t, L_t) + \varpi + [1 - s_t(H_t)] b(a_{t+1}) + \beta s_t(H_t) \mathbb{E}_t \big[V_{t+1}(X_{t+1}) \mid H_{t+1} \neq D \big] \Big\}, \tag{15}$$

subject to the budget constraint (13), the time constraint (3), the earnings accumulation rule (11), the borrowing and feasibility constraints, and the exogenous laws of motion for health and productivity. The continuation expectation integrates over next-period health conditional on survival and the next productivity innovation. At the terminal age, $s_T(H_T) = 0$ and $V_{T+1} \equiv 0$. We solve the problem by backward induction. Computational details appear in Appendix D.

3 Data and Estimation

In the economic model, individuals make decisions knowing their underlying health, but the health measures they report in surveys need not accurately reflect that state. To assess the consequences for structural inference, we estimate the same economic model under two specifications. The first follows the conventional approach and treats binary self-reported health as the underlying state, while the second infers latent health from multiple noisy indicators. Accounting for measurement error requires correcting not only the estimated health process but also the initial distribution of states, economic outcomes conditional on health, and the moments used to estimate structural parameters. We first describe the data and health measures, then explain how we recover latent health and incorporate it into structural estimation, and finally present the parameter estimates and model fit.

3.1 Data and Health Measurement

We use waves 1–9 (2002–2019) of the English Longitudinal Study of Ageing (ELSA), a biennial survey of individuals aged 50 and older living in private households in England. ELSA follows age-eligible *core members* and collects information about their cohabiting partners.[7] We model the decisions of male core members. Labor supply,

[7] Core members originate from the Health Survey for England and remain eligible for follow-up unless they die or move outside Great Britain.

earnings, and health are measured at the individual level, whereas assets and pensions are measured at the household level. We incorporate spousal resources using observed partner income without explicitly modeling spouses' labor-supply decisions.

Our sample consists of men born between 1935 and 1950 who are married at their first interview, do not subsequently divorce or remarry during the observed panel, and are self-employed in no more than half of their employed waves. Men who subsequently become widowed or separated remain in the sample. These cohort restrictions align the sample with the UK pension environment modeled by O'Dea (2018). The resulting sample contains 2,223 men with complete initial-state information. Of these, 83% enter in wave 1 and the remainder in the wave-4 refreshment sample; entry ages range from 52 to 74 (see Appendix C). The health, wage, and spousal-income processes are estimated using broader samples of ELSA men, as detailed in Appendix B.1.

To measure underlying health while allowing for reporting error, we use three indicators: self-reported general health, the number of mobility difficulties, and the number of limitations in activities of daily living (ADLs) and instrumental activities of daily living (IADLs). The latter two use IFS-derived ELSA variables. Self-reported general health is collapsed into good (excellent, very good, or good) and bad (fair or poor). The counts of mobility difficulties and ADL/IADL limitations capture more detailed variation in physical functioning and are top-coded at eight and four, respectively. In the specification without measurement error, binary self-reported health alone defines the health state.[8]

We identify deaths using ELSA interview-status records and use mortality information for transitions beginning in waves 1–5, for which these records are considered reliable. After excluding proxy interviews and observations missing mobility or ADL/IADL measures, the health-estimation sample contains 37,209 person-wave observations on 8,680 men aged 50–91. Self-reported general health is available for 33,131 of these observations.

3.2 Health Dynamics and Identification

The two specifications differ in how they measure the health state H_t of Section 2. *Without measurement error*, H_t is the binary classification of self-reported health, and age-specific transitions between good health, bad health, and death are estimated from observed transitions. *With measurement error*, H_t is unobserved. Its transition

[8]The wording and availability of the general-health question differ across waves. For the latent-health specification, the preferred general-health question is treated as missing by design in wave 3. For the specification without measurement error, the alternative question is used in wave 3 and where needed in wave 1. The exact harmonization rules are documented in Appendix B.1.

probabilities and the distributions of the three indicators conditional on health are jointly estimated using a non-stationary hidden Markov model (see Appendix B.2). In both specifications, transitions between the living health states vary freely by age, and mortality is a logistic function of age with health-specific intercepts (Appendix B.2). This ensures that differences between specifications do not arise mechanically from imposing different age-profile restrictions on health dynamics. By contrast, Amengual, Bueren and Crego (2021) model transition probabilities as logit functions of age. Such restrictions may improve efficiency but, if misspecified, can confound the effects of measurement error with those of functional form.[9]

Let Y_{it}^{m} denote indicator $m \in \{1,2,3\}$ for individual i in period t, and define the emission probabilities

$$p_h^m(k) \equiv \Pr(Y_{it}^m = k \mid H_{it} = h), \qquad h \in \{G, B\}. \tag{16}$$

The emission matrix P^m collects these probabilities for indicator m, with one column for each health state. We assume that the emission distributions are constant across ages, whereas transition probabilities vary with age. The emission distributions describe how an individual in either latent health state may report each observed indicator. The baseline measurement model assumes that, conditional on current latent health, the three indicators are independent:

$$\Pr(Y_{it}^1, Y_{it}^2, Y_{it}^3 \mid H_{it} = h) = \prod_{m=1}^{3} \Pr(Y_{it}^m \mid H_{it} = h). \tag{17}$$

The resulting joint distribution of indicators is a two-component mixture, with mixture weights given by the shares of individuals in good and bad latent health. We orient the two components by requiring the good-health state to be associated with more favorable health indicators.

We adapt the identification methods of García-Vázquez (2021) for non-stationary hidden Markov models to accommodate mortality. The identification argument has three steps. First, conditional independence of the three indicators, together with appropriate rank and support conditions, identifies the emission distributions and the shares of individuals in each latent health state, up to relabeling. Second, observed deaths and the indicators recorded before death identify state-dependent survival probabilities. Third, joint observations of indicators at consecutive ages identify transitions between the two living states. Refreshment samples provide observations at additional ages,

[9] French (2005) and Bueren (2021) partially address measurement error in health, but impose parametric restrictions on the transition matrices for health and ignore measurement error when estimating the initial distribution of states.

allowing us to recover transitions over a longer portion of the life cycle than any single respondent is observed. Appendix A establishes the formal result and its assumptions.

Self-reported health anchors the comparison with the conventional specification but may also be especially sensitive to employment-related reporting behavior. Such *justification bias* (Bound, 1991; Kerkhofs and Lindeboom, 1995; Dwyer and Mitchell, 1999; Benítez-Silva et al., 2004; Lindeboom and Kerkhofs, 2009) threatens the separate assumption used below to recover economic outcomes conditional on latent health. We therefore replace self-reported general health with reported pain, retaining mobility and ADL/IADL limitations, and re-estimate the full model as a robustness exercise.

The estimated emission probabilities reveal substantial misclassification. Table 1 shows that 12% of individuals in latent good health report bad health, while 26% of those in latent bad health report good health. Binary self-reports therefore mix individuals from the two latent states. The mobility and ADL/IADL indicators are also informative about latent health. For example, an individual in latent good health reports no mobility difficulty with probability 0.69, compared with 0.01 for an individual in latent bad health (see Figure B.1 in Appendix B.2).

Table 1: Self-reported health conditional on latent health

	Latent health	
Self-reported health	Good	Bad
Good	0.88	0.26
Bad	0.12	0.74

3.3 Mapping the Model to the Data

Correcting health transitions alone is insufficient for consistent treatment of health measurement throughout structural estimation. Health also enters the initial distribution of states, wage and spousal-income processes, and the labor-supply and wealth statistics used in estimation. We therefore align these objects with the health-measurement convention used to estimate the health process. In the specification without measurement error, economic outcomes and initial health are grouped by binary self-reported health. In the latent-health specification, we use the estimated measurement system to recover or reproduce the corresponding statistics.

For outcomes summarized by health-specific conditional means, we use posterior probabilities of latent health. Let

$$q_{iw} \equiv \Pr(H_{iw} = G \mid Y^1_{iw}, Y^2_{iw}, Y^3_{iw}, t), \tag{18}$$

where the probability is computed from the estimated emission distributions and the estimated shares of each health state at age t; here w indexes survey waves and t the two-year age group of individual i at wave w. This posterior uses only the individual's indicators at wave w, not his full indicator history. If the indicators are conditionally independent of an economic outcome y_{iw} given latent health and the relevant age group t, the conditional mean satisfies

$$\mathbb{E}[y_{iw} \mid Y^1_{iw}, Y^2_{iw}, Y^3_{iw}, t] = \mu^{y,G}_t q_{iw} + \mu^{y,B}_t (1 - q_{iw}), \tag{19}$$

where $\mu^{y,h}_t \equiv \mathbb{E}[y_{iw} \mid H_{iw} = h, t]$. We recover these latent-health conditional means by regressing y_{iw} on q_{iw} and $1 - q_{iw}$ without a constant. This restriction differs from the conditional independence of the health indicators in equation (17): whereas the latter identifies the measurement system, the former allows us to recover economic outcomes conditional on latent health. It requires that, conditional on latent health and age, the indicators contain no additional information about economic outcomes. We discuss this assumption and its implications in Appendix B.3.

We apply this approach to the age–health profiles of hours worked and labor-force participation, which are central targets of structural estimation, and to the reduced-form spousal-income function $y_S(t, H_t)$ (see Appendices B.1 and B.3). In the simulated data, the corresponding hours and participation profiles are computed using each simulated individual's true health.

Wages require a distinct correction. Health enters the wage specification through the good-health premium, and we use posterior probabilities in place of an observed health dummy in the wage-profile regression. Because the wage regression uses pairs of waves, its posteriors condition on indicators at both dates and survival between them. We also account for the additional measurement component in wage residual moments when constructing the persistent wage process. Because wages are observed only for workers, the deterministic wage-profile coefficients are not treated as consistently estimated reduced-form parameters: they are disciplined jointly with the structural model by indirect inference (see Appendix B.4).

Initial conditions must also reflect the health-measurement convention. Each household enters the simulation at its first observed age. Observed state variables are initialized from the entry interview, while the unobserved components—including average prior earnings—are imputed. In the latent-health specification, initial health is drawn from its posterior distribution given the individual's whole history of indicators and survival; this is the econometrician's inference about health at entry, not information available to the household, which knows its health. In the specification without measurement error, initial health is the binary self-report. Appendix C describes the joint

initialization procedure. Because ELSA does not observe complete earnings histories before age 50, we use simulations from O'Dea (2018) to construct average earnings at entry and the pension benefit functions (see Appendices C.1 and B.5).

We handle wealth quantiles differently because they are not conditional means. The 25th, 50th, and 75th percentiles of wealth by age and health cannot be recovered by a linear regression on posterior probabilities. Instead, we classify individuals by their most likely latent-health state given the indicators observed at that wave. To apply the same classification in the data and model, we simulate the three health indicators conditional on each simulated individual's true health, recompute the posteriors, and assign simulated individuals to their modal health state. Thus, these targets are *wealth quantiles for groups classified using noisy health measurements*, not the quantiles of wealth conditional on the unobserved true health state. Household wealth is adjusted for household size and survey-wave effects as in De Nardi, Pashchenko and Porapakkarm (2025), and as described in Appendix E.4.

These distinctions matter for comparing the specifications with and without measurement error. Both share the economic environment, parameterization, and economic outcomes targeted in estimation. What changes are the health dynamics governing households' decisions and the mapping from observed health information to the moments used to estimate the model. Correcting health transitions alone is therefore insufficient: continuing to group economic outcomes by self-reported health would leave an important source of measurement error uncorrected.

3.4 Structural Estimation

We estimate the model in two steps, as is common in structural life-cycle analysis (see, e.g., Gourinchas and Parker, 2002; French, 2005; De Nardi, French and Jones, 2010). In the first step, we estimate the health and wage-shock processes, spousal income, pension functions, and initial conditions, while setting selected institutional and preference parameters using external information. These objects are constructed separately under the two treatments of health whenever they depend on health measurement; the pension functions and the externally set parameters do not, and they are reported in Appendices B.5 and B.6. In the second step, we estimate the remaining parameters by indirect inference (Gourieroux et al., 1993; Smith Jr, 1993).

The second-step parameter vector is

$$\theta = (\beta, \nu, \phi_P, \theta_B, \kappa_B, \phi_H, C_{\min}, a_0, a_1, a_2, a_H). \tag{20}$$

This vector contains preference and labor-supply parameters, the consumption floor, and the deterministic wage-profile coefficients. We estimate the wage-profile coefficients within the structural model to account for selection into employment, since wages are observed only for individuals who work. The coefficient of relative risk aversion is fixed at $\gamma = 3$.

Let ψ_n denote the empirical vector of target statistics and $\psi(\theta)$ its model counterpart. We solve

$$\widehat{\theta} = \arg\min_{\theta} [\psi(\theta) - \psi_n]' W_n [\psi(\theta) - \psi_n], \tag{21}$$

where W_n is diagonal, with entries given by the inverse bootstrap variances of the empirical targets. The targets comprise age–health profiles of wealth at the 25th, 50th, and 75th percentiles, mean hours worked, participation, and coefficients from the wage regression. Of 204 potential targets, 182 are used; cells with too few observations are excluded. For each specification, we apply the same health-measurement rule to simulated and observed targets. We simulate each individual 40 times and use a global optimization procedure (Arnoud, Guvenen and Kleineberg, 2019) because the criterion is non-smooth and has multiple local minima (see Appendix E).

Table 2 reports the parameter estimates (see Appendix E.3 for a discussion of parameter identification and approximate standard errors). Two differences stand out. Accounting for measurement error raises the estimated time cost of bad health from 1,230 to 1,735 hours per two-year period, an increase of 41%, and the estimated good-health wage premium from 0.15 to 0.20 log points. Both changes are consistent with misclassification attenuating the estimated effects of health on economic outcomes, as illustrated by the stylized model in Appendix F.

Table 2: Parameters estimated by indirect inference

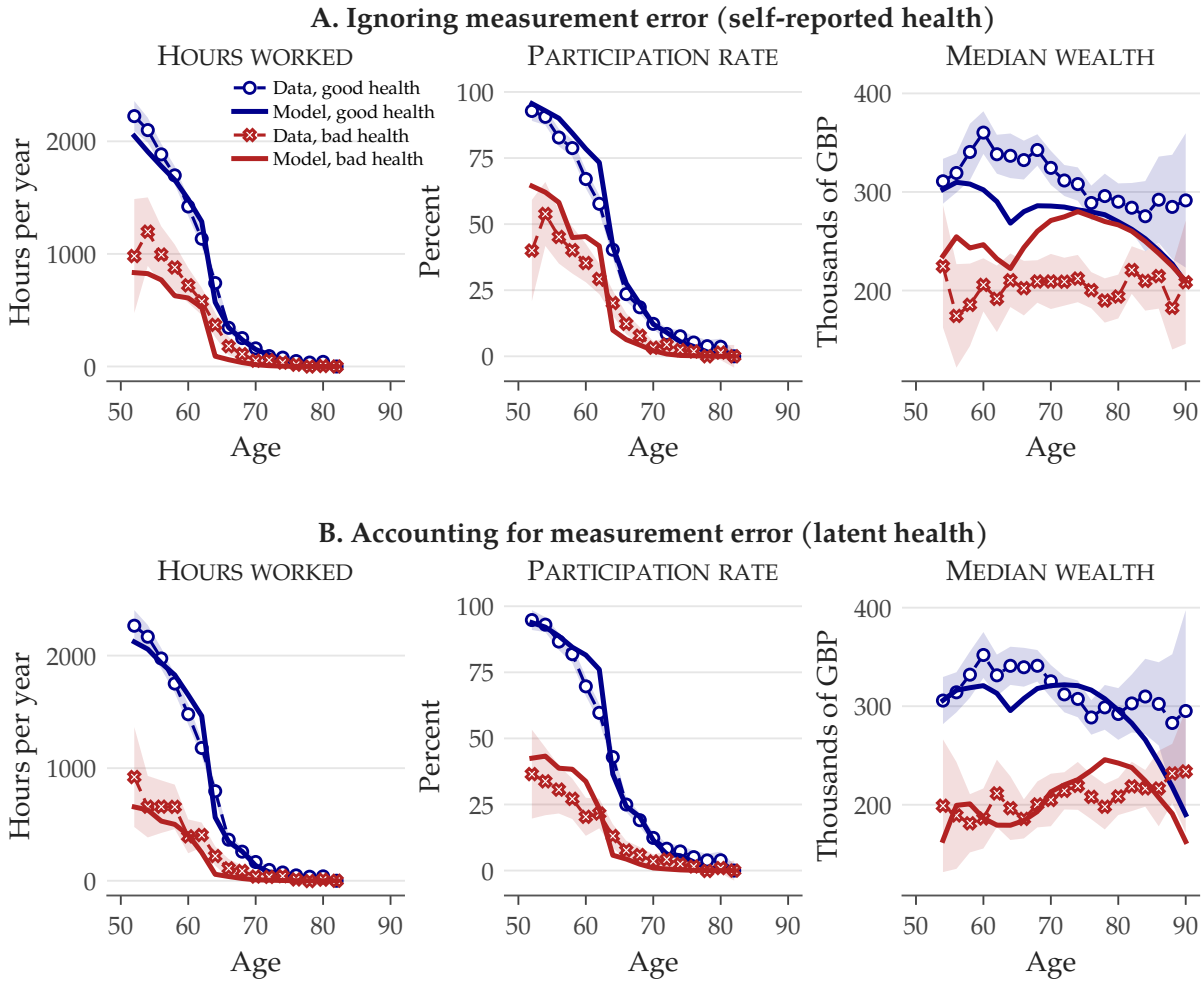
Parameter	**Ignoring measurement error**	**Accounting for measurement error**
β: biennial discount factor	1.035	1.040
ν: consumption weight	0.864	0.782
ϕ_P: participation time cost	3,305	2,131
θ_B: bequest weight	3.55	6.56
κ_B: bequest curvature parameter	186,336	865,535
ϕ_H: time cost of bad health	1,230	1,735
C_{min}: consumption floor	63,362	47,440
a_0: wage-profile constant	2.522	2.394
a_1: linear age term	−0.028	−0.005
a_2: quadratic age term	−0.005	−0.005
a_H: good-health wage premium	0.154	0.204

Notes. ϕ_P and ϕ_H are in hours per two-year period, C_{min} is in GBP per two-year period, and κ_B in GBP.

3.5 Model Fit

Figure 1 compares the simulated and observed profiles of hours worked, labor-force participation, and median wealth under both specifications. Both models reproduce the broad age patterns in hours worked and participation, although the common retirement-age rule produces a sharper decline in employment than observed in the data. The latent-health specification, however, fits the observed wealth differences between health groups considerably better. It also matches the observed health coefficient in the wage regression, whereas the specification without measurement error overpredicts it by approximately two standard errors (see Table E.1 in Appendix E.5).

Figure 1: Model Fit



Notes. Each row uses the targets of its specification. Shaded areas are 95% confidence intervals for the data targets. Wealth is adjusted for household size and wave, as explained in Section 3.3.

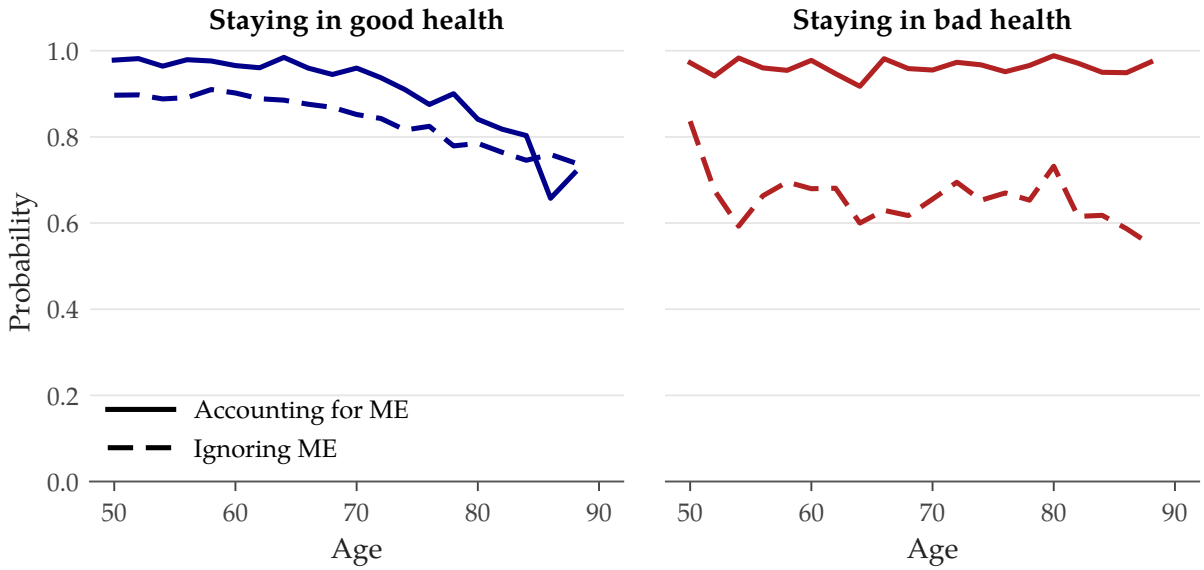
4 Health Dynamics and the Wealth–Health Gradient

We now examine how measurement error affects the estimated dynamics of health and the ability of the model to account for the wealth–health gradient. We first show that bad health is substantially more persistent than self-reported health suggests. We then show that a persistent latent health process observed with error can generate negative duration dependence in reported health, a pattern often attributed to permanent health heterogeneity. Finally, we show that the greater persistence of latent health allows the model to explain most of the observed wealth–health gradient without permanent health types or preference heterogeneity.

4.1 Health Persistence

Figure 2 compares the estimated two-year health transition probabilities with and without measurement error. Accounting for measurement error reveals substantially greater persistence, particularly in bad health: the probability of remaining in bad health exceeds 90% at every age, compared with 55–84% when measurement error is ignored. Good health is also more persistent at nearly every age.

Figure 2: Health Persistence With and Without Measurement Error



Notes. Two-year transition probabilities among survivors, by age at the start of the period. "Accounting for ME" refers to transitions in latent health, while "Ignoring ME" to transitions in self-reported health.

This difference reflects the fact that changes in reported health need not reflect changes in underlying health. For example, a healthy individual who reports bad health in one wave and good health in the next appears to have recovered, even though his underlying health has not changed. Such misclassification generates spurious

transitions. As Table 1 shows, 12% of individuals in latent good health report bad health, while 26% of those in latent bad health report good health. These reporting errors inflate the frequency of observed health transitions, making bad health appear less persistent than it is. This distinction matters for economic behavior: the longer bad health is expected to last, the greater its potential consequences for lifetime earnings, labor supply, and saving.

4.2 Duration Dependence in Reported Health

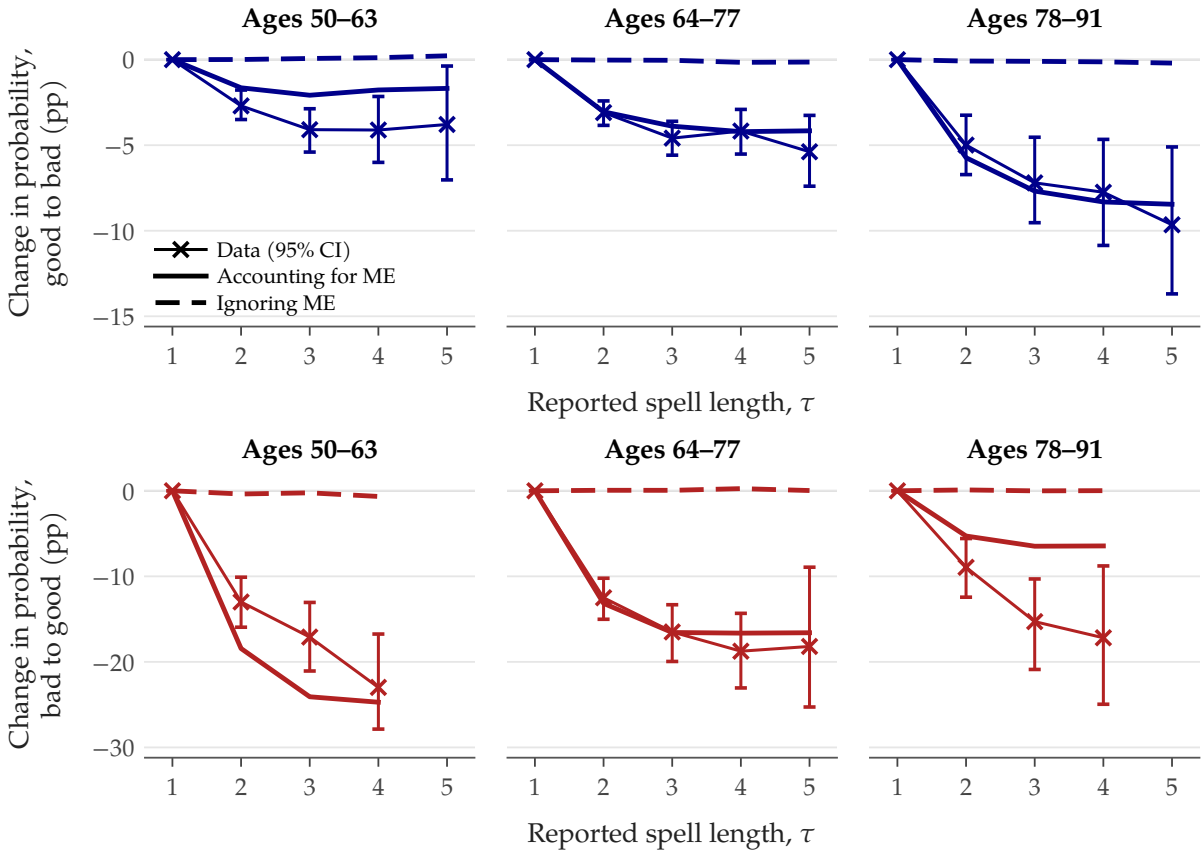
Measurement error also affects how we interpret observed health histories. De Nardi, Pashchenko and Porapakkarm (2025) document that the probability of leaving a self-reported health state declines with the duration of the reported spell, a pattern they reproduce by introducing permanent health types. We show that a first-order latent health process observed with error can generate negative duration dependence without such heterogeneity. To examine this, we calculate the probability of leaving a reported health state after at least τ consecutive two-year periods in that state, treating missing interviews and wave 3 as interruptions. We compare these probabilities in the data with those from simulations of both estimated health processes. For the latent-health specification, we simulate self-reports using the estimated emission probabilities.

Figure 3 plots changes in exit probabilities relative to the first period of a reported health spell, allowing us to compare how they decline with reported spell duration despite differences in levels. The data exhibit pronounced negative duration dependence for both good and bad health. Among individuals aged 50–63, the probability of reporting good health falls from 0.27 after at least one period of reported bad health to 0.04 after at least four, a decline of 23 percentage points. The specification without measurement error fails to reproduce this pattern, as expected from a first-order Markov process in which transition probabilities depend on age but not on spell duration. By contrast, for the same age group, the latent-health specification generates a decline of about 25 percentage points, from 0.54 to 0.30, closely matching the absolute decline in the data despite the first-order Markov structure of underlying health.

The intuition is simple. A single report of bad health provides imperfect information about underlying health: at ages 50–51, only 43% of those reporting bad health are actually in that state. Consecutive reports of bad health increase the posterior probability that an individual is truly unhealthy. Because latent bad health is highly persistent, longer reported spells are associated with lower subsequent probabilities of reporting recovery.

Although the latent-health specification captures much of the absolute decline in exit probabilities, it overstates their levels. At $\tau = 1$, among individuals aged 50–63,

Figure 3: Duration Dependence in Self-Reported Health



Notes. Change in the probability of leaving the reported health state relative to the initial period ($\tau = 1$), in percentage points, where τ is the minimum number of consecutive two-year periods in that state. Results are grouped by age at the start of the transition. "Accounting for ME" refers to self-reports simulated from the latent-health process and the emission probabilities. Data: 32,379 observations on 8,386 men; 95% confidence intervals from 300 bootstrap samples of individuals; points with fewer than 50 transitions are omitted. Models: 200 simulated panels using individuals' observed interview patterns.

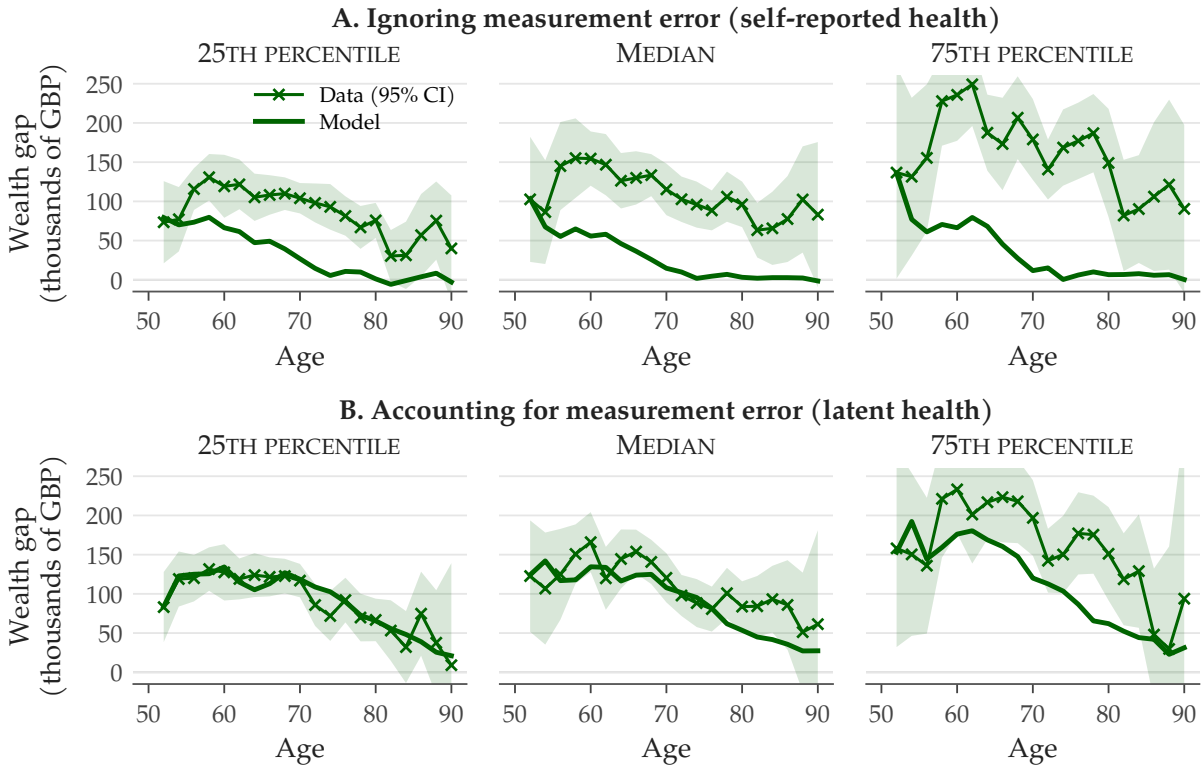
the probability of leaving reported bad health is 0.54 in the model, compared with 0.27 in the data. Consequently, the proportional decline by $\tau = 4$ is much smaller in the model (45%) than in the data (84%). This discrepancy arises despite similar unconditional frequencies of reported bad health (0.26 in the model and 0.24 in the data). With conditionally independent reporting errors, the estimated misclassification probabilities constrain how low exit probabilities can fall, whereas the data exhibit substantially lower probabilities after long reported spells. Changes in the survey question, mortality, and the composition of individuals observed in consecutive waves do not explain this discrepancy. Persistent reporting errors, as in Lange and McKee (2012), or a richer latent-health process may help account for it. Nevertheless, negative duration dependence in reported health does not, by itself, require permanent health heterogeneity.

4.3 The Wealth–Health Gradient

The greater persistence of latent health has important implications for wealth accumulation. Both specifications target wealth distributions by age and health status, yet they differ substantially in their ability to reproduce the observed wealth–health gradient. Figure 4 compares the observed and predicted wealth differences between individuals in good and bad health at the 25th, 50th, and 75th percentiles under each specification.

Both specifications reproduce the observed wealth–health gradient at the youngest ages, but their predictions diverge substantially with age. In the panel ignoring measurement error, the model increasingly understates the observed gradient, with predicted wealth differences nearly disappearing by the late seventies at all three percentiles. By contrast, the model accounting for measurement error preserves a large part of the observed gradient at all ages, although it understates the gap at the 75th percentile and, beyond age 78, at the median as well.

Figure 4: Wealth–Health Gradient



Notes. Difference in wealth between individuals in good and bad health at each percentile. Health classifications follow the measurement convention of each specification, so the data differ across rows. Shaded areas are 95% confidence intervals.

Averaging across the 20 age groups used in estimation, the observed wealth gap is approximately £87,000 at the 25th percentile, £109,000 at the median, and £160,000 at the 75th percentile. The specification without measurement error reproduces only 34%,

22%, and 19% of these gaps, respectively. Accounting for measurement error raises these shares to 101%, 82%, and 69%. Across all three percentiles, the model reproduces 81% of the observed wealth–health gradient when accounting for measurement error, compared with 23% when ignoring it.

Three mechanisms help explain why greater health persistence produces a larger wealth–health gradient. First, greater persistence preserves initial wealth differences. Individuals enter the model with wealth that differs by health, so both specifications start with a substantial gap: about £103,000 at age 52 when ignoring measurement error and £124,000 when accounting for it. Under the self-reported process, however, health at age 50 says little about health ten years later: among survivors, the probability of being in bad health at age 60 is 0.28 for those in bad health at age 50 and 0.22 for those in good health, compared with 0.83 and 0.10 under latent health. Under the self-reported process, current health therefore becomes increasingly uninformative about earlier health and accumulated wealth, causing the initial wealth differences between health groups to dissipate. Second, greater persistence allows earnings losses to accumulate. Individuals in bad health at age 64 have spent an average of about 10 of the preceding 14 years in bad health under latent health, compared with only 5 under self-reported health. Their lower wages and hours therefore translate into more prolonged earnings losses, contributing to larger wealth differences. Third, persistence affects saving incentives through expected lifetime income and survival. At age 50, remaining life expectancy is approximately six years lower in bad health than in good health under latent health, compared with one year under self-reported health. This larger survival gap shortens the expected retirement horizon of individuals in bad health, potentially reducing their incentive to accumulate retirement wealth.

These results offer an alternative explanation for the findings of De Nardi, Pashchenko and Porapakkarm (2025), who estimate a model with self-reported health and persistent health types. They find that reproducing the observed wealth–health gradient requires individuals with worse health types to be less patient. Without this correlation, their model generates a median wealth difference near retirement of approximately $35,000, compared with $140,000 in the data. This shortfall resembles that of our specification without measurement error. Our results point to a different mechanism: greater persistence in underlying health allows a model with a common discount factor and no permanent health types to reproduce most of the observed wealth–health gradient.

To assess the role of the health process, we re-estimate the specification without measurement error after replacing its health transitions and health-dependent mortality with those obtained when accounting for measurement error. The estimation targets, initial distribution, wage process, and spousal income process remain unchanged. The resulting model reproduces approximately 70% of the observed wealth–health gradient,

compared with 23% under the original self-reported-health process. This exercise captures the joint contribution of health transitions and mortality while allowing the structural parameters to adjust. Although it does not isolate persistence, the substantially larger differences in health transition probabilities than in health-specific mortality rates suggest that persistence plays an important role.

Together, these results show that a model's inability to reproduce the wealth–health gradient need not imply missing preference heterogeneity or permanent health types. In our application, much of the discrepancy stems from understating the persistence of bad health. Correcting health measurement therefore changes not only the estimated dynamics of health but also the model's ability to explain observed wealth inequality.

5 The (Mismeasured) Costs of Bad Health

We now examine how measurement error in health affects the estimated costs of bad health. We first quantify these costs under the two specifications, distinguishing between the population as a whole and individuals who enter the model in bad health. We then decompose the differences between specifications into the contributions of the estimated health process and the health-dependent moments used in structural estimation. Finally, we examine robustness to alternative health measures and briefly discuss the implications for retirement-age and pension-policy reforms.

5.1 Quantifying the Costs of Bad Health

We measure the costs of bad health by comparing the baseline economy, in which individuals face the estimated health risks, with a counterfactual in which they remain in good health throughout their lives and face the corresponding survival probabilities. Following the approach of De Nardi, Pashchenko and Porapakkarm (2025), we simulate both scenarios using the decision rules of the baseline economy and holding fixed the realizations of other exogenous shocks. Individuals therefore do not anticipate the elimination of health risk. We calculate differences in earnings, hours worked, years worked, consumption, assets, longevity, and welfare. Table 3 reports these differences for the population as a whole and for individuals who enter the model in bad health.

We distinguish between two measures of lost income. The *earnings cost* is the difference in average annual earnings of the household head through ages 64–65, with each average calculated over individuals alive in the corresponding scenario. The *monetary cost* also includes spousal income and pensions and accounts for income lost through premature mortality. Specifically, for each individual, we calculate the

difference between scenarios in the present value of income, discounted to the age at entry and assigning zero income after death, and divide it by the number of years lived in the always-healthy scenario. We then average this measure across individuals.

We also calculate the welfare cost of bad health, incorporating the value of additional years of life. For each individual, we measure this cost as the fraction of consumption in every period of the always-healthy scenario that equates realized lifetime utility across the two scenarios. We then average this consumption-equivalent measure across individuals. We calibrate the additive value of being alive, ϖ, so that the average value of a statistical life (VSL) among working-age individuals (ages 50–63) equals \$2 million in 2013 dollars, converted to pounds.[10] We calibrate ϖ separately for each specification and also report results for an alternative VSL of \$6 million (see the notes to Table 3). Because mortality is exogenous and ϖ enters utility additively, this calibration does not affect household decisions.

Table 3: Costs of Bad Health

Outcome	Ignoring ME	Accounting for ME	Difference
All individuals			
Annual earnings (£)	1,778	2,554	44%
Annual hours worked	151	220	45%
Years worked	0.69	0.99	44%
Annual consumption (£)	281	112	−60%
Assets (£)	2,432	2,550	5%
Annual monetary cost (£)	1,978	2,641	34%
Years of life gained	2.53	3.54	40%
Welfare cost (%)	6.67	10.85	63%
Initially unhealthy			
Annual earnings (£)	5,107	11,036	116%
Annual hours worked	483	980	103%
Years worked	1.61	3.50	117%
Annual consumption (£)	865	2,439	182%
Assets (£)	14,232	42,632	200%
Annual monetary cost (£)	3,720	7,345	97%
Years of life gained	3.51	7.50	114%
Welfare cost (%)	8.14	18.71	130%

Notes. Each entry is the difference between the scenario in which individuals are always in good health and the baseline. Earnings and hours worked are computed up to age 64–65, and the monetary cost over the remaining life. Consumption and assets are differences between means over the years in which individuals are alive in each scenario. The welfare cost uses a VSL of £1.39 million; with a VSL of £4.17 million, it is 11.82% and 18.51% for all individuals, and 14.53% and 31.56% for the initially unhealthy.

Accounting for measurement error substantially increases the estimated costs of bad health. For the population as a whole, the earnings cost rises from £1,778 to £2,554

[10] We calculate the average VSL over ages 50–63, whereas De Nardi, Pashchenko and Porapakkarm (2025) use ages 21–63 (see Appendix G.1 for further details).

per year, an increase of 44%. The estimated reduction in annual hours worked rises by 45%, from 151 to 220 hours, while years of work lost increase from 0.69 to 0.99. These results suggest that ignoring measurement error in health leads to understating not only the persistence of health but also its effects on labor supply and earnings.

The differences are particularly large for individuals who enter the model in bad health. Accounting for measurement error raises their estimated earnings cost from £5,107 to £11,036 per year, more than doubling it. Annual hours lost increase from 483 to 980, while years of work lost rise from 1.61 to 3.50. These larger costs are consistent with the greater persistence of latent bad health, which prolongs the exposure of initially unhealthy individuals to its adverse economic consequences.

The effects on consumption and assets are more heterogeneous. For individuals initially in bad health, accounting for measurement error nearly triples both the estimated consumption cost, from £865 to £2,439 per year, and the asset cost, from £14,232 to £42,632. For the population as a whole, however, the asset cost barely changes, while the consumption cost falls from £281 to £112 per year. Because these measures compare averages among survivors in each scenario, they partly reflect differences in age composition rather than solely changes in individual consumption and saving.

The monetary and longevity costs also increase substantially. For the population as a whole, the remaining-life monetary cost rises by 34%, from £1,978 to £2,641 per year. For individuals initially in bad health, it nearly doubles, from £3,720 to £7,345. Eliminating bad health increases expected longevity by 3.54 years in the latent-health model, compared with 2.53 years when measurement error is ignored. For individuals initially unhealthy, the corresponding gains are 7.50 and 3.51 years. Thus, accounting for measurement error raises the estimated longevity cost of bad health by approximately one year for the average individual and four years for those initially unhealthy.

These differences translate into larger welfare costs. At a VSL of £1.39 million, the estimated welfare cost rises from roughly 7% to 11% of consumption for the population as a whole, and from 8% to 19% for individuals initially in bad health. Taken together, these findings show that ignoring measurement error substantially understates the estimated costs of bad health across earnings, labor supply, longevity, and welfare, particularly for individuals initially in bad health.

5.2 The Sources of Measurement-Error Bias

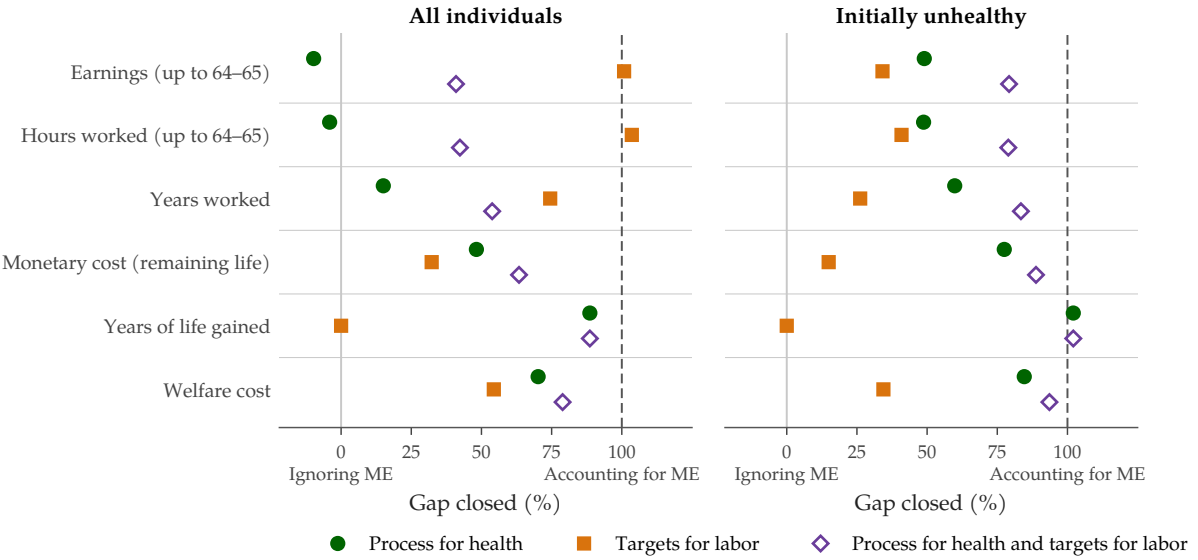
The two specifications differ in both the estimated health process and the economic moments conditional on health used to estimate structural parameters. Correcting measurement error reveals greater health persistence. It also uncovers larger differences in labor supply between healthy and unhealthy individuals, which imply a higher

estimated time cost of bad health. To assess the contribution of each source of difference, we re-estimate the specification without measurement error after replacing these inputs with those of the specification accounting for it.

Starting from the specification without measurement error, we replace three sets of inputs, separately and in combination: the health and mortality process, the labor-supply targets (hours worked and participation), and the wealth targets. Each replacement uses the corresponding inputs from the specification accounting for measurement error, while leaving the initial distribution, wage process, and spousal income unchanged. We then re-estimate the structural parameters and recompute the costs of bad health. For the wealth targets, the replacement follows the classification procedure described in Appendix G.2.

Figure 5 reports the percentage of the gap in estimated costs between the two specifications that each replacement closes. A value of 100% means that the replacement eliminates the gap, while a value of zero means that it leaves it unchanged. Because the inputs interact through the re-estimated model, their contributions need not add up. Even replacing all three sets of inputs need not eliminate the gap, since the initial distribution, wage process, and spousal income remain unchanged.

Figure 5: Decomposing Differences in the Estimated Costs of Bad Health



Notes. Each marker reports the result of re-estimating the specification without measurement error after replacing the indicated inputs with those from the specification accounting for measurement error. A value of 0% corresponds to the estimated cost when measurement error is ignored, while 100% (dashed line) corresponds to the cost when it is accounted for. Negative values indicate that the replacement widens, rather than closes, the gap with the specification accounting for measurement error. The welfare cost uses a VSL of £1.39 million.

The decomposition reveals two distinct sources of measurement-error bias, with markedly different implications for the overall population and those initially unhealthy.

For the population as a whole, correcting labor-supply targets is the main driver of the difference in the working-age costs of bad health. Replacing the targets for hours worked and participation accounts for essentially the entire difference in foregone earnings and hours worked through retirement age, and 74% of the difference in years worked. By contrast, replacing the health process alone does not increase the estimated earnings or hours costs. The larger working-age costs therefore arise primarily because correcting misclassification reveals larger differences in labor supply between health states, leading indirect inference to estimate a higher time cost of bad health.

For individuals initially in bad health, the estimated health process plays the dominant role. Replacing health transitions and mortality accounts for 77% of the difference in remaining-life monetary costs, 49% of the difference in earnings costs, and essentially the entire difference in longevity. These results reflect the importance of health dynamics for individuals already unhealthy at entry: greater persistence prolongs the economic consequences of bad health, while health-dependent mortality determines the years of life lost.

The longevity results also clarify the interpretation of the decomposition. Years of life gained depend on health transitions, mortality, and the initial distribution of health, but not on preferences or the economic targets. Replacing the health process therefore accounts for essentially the entire difference in longevity. Its contribution is not exactly 100%, however, because the initial distribution remains unchanged in these exercises.

Combining the health process with the corrected labor-supply targets accounts for most of the differences among individuals initially in bad health. Together, these inputs explain about 80% of the difference in earnings and hours worked, 90% of the difference in remaining-life monetary costs, and 95% of the difference in welfare costs. For the population as a whole, the interactions are more pronounced: correcting labor-supply targets alone closes more of the gap in working-age earnings and hours than correcting them jointly with the health process.

Correcting the wealth targets has little effect. Replacing them alone leaves the estimated costs almost unchanged and moves the estimated parameters by less than half a percent, and adding them to the other replacements barely changes the results. This is consistent with Section 4.3: the specification without measurement error fails to reproduce the wealth–health gradient even though wealth by health is among its targets, which points to its health process rather than the construction of the targets.

One exception concerns assets and consumption for the population as a whole. The differences in these costs between the two baseline specifications are small relative to the changes induced by individual replacements, producing shares below 0% or above 100%. These shares are therefore difficult to interpret as measures of the relative importance of the different inputs, Figure 5 therefore omits consumption and assets,

and Table G.1 in Appendix G.2 reports the remaining shares, omitting only the asset shares for the full population.

The decomposition establishes why correcting health transitions alone is insufficient. For individuals initially unhealthy, the health and mortality process accounts for most of the difference in remaining-life monetary costs and longevity. For the population as a whole, however, the difference in working-age earnings costs arises primarily from correcting the labor-supply moments used in estimation. Ignoring measurement error therefore creates two distinct distortions in our application: it understates health persistence and attenuates the observed behavioral differences between health states. Correcting one without the other can leave substantial errors in the estimated costs of bad health.

Robustness to Alternative Health Measures. Our baseline latent-health specification uses self-reported general health, mobility difficulties, and ADL/IADL limitations. An important concern is that self-reported health may be subject to justification bias: individuals may report worse health to justify lower labor supply, even conditional on their underlying health. To assess the sensitivity of our results, we replace self-reported health with reported pain, retaining the other two indicators, and re-estimate the full model. The results, reported in Appendix G.4, are qualitatively and quantitatively similar. For the population as a whole, the estimated earnings cost is £2,450 per year when using pain, compared with £2,554 in the baseline latent-health specification and £1,778 when ignoring measurement error. For individuals initially in bad health, the corresponding estimates are £9,535, £11,036, and £5,107. The estimated longevity gains from eliminating bad health are likewise substantially larger than when measurement error is ignored. Although the somewhat smaller costs obtained using pain could reflect justification bias, they could also indicate that the alternative indicators capture different dimensions of health. The larger estimated costs of bad health are therefore robust to replacing self-reported general health with a symptom-based measure.

Implications for Pension Policy. The two specifications also produce relatively similar predictions for pension-policy experiments that do not directly change health risk. Appendix G.3 considers increases in the pension age of one and two years and reductions in the state pension of 5% to 20%. Accounting for measurement error implies somewhat larger labor-supply and saving responses to these reforms but similar welfare costs. These differences are much smaller than those found for the costs of bad health: the importance of correcting health measurement depends on the counterfactual under consideration.

6 Conclusion

Economists do not observe health directly. Instead, they typically rely on self-reported health and treat it as if it perfectly revealed the underlying state. This paper shows that this seemingly innocuous shortcut can substantially distort what structural models tell us about health, economic behavior, and wealth inequality. Using a life-cycle model of labor supply and saving, we compare a specification that treats self-reported health as the true health state with one that infers latent health from multiple noisy indicators. Crucially, we account for measurement error throughout structural estimation, rather than only when estimating health transitions.

Ignoring measurement error has two important consequences. First, it makes health appear substantially less persistent because some observed transitions between good and bad health reflect reporting errors rather than changes in underlying health. Second, misclassification understates differences in labor supply and earnings between healthy and unhealthy individuals, leading the model to underestimate the economic burden of bad health. In particular, accounting for measurement error raises the estimated time cost of bad health by 41%, to approximately 17 hours per week. Correcting health measurement also changes what the model can explain: without preference heterogeneity or permanent health types, the latent-health model reproduces about 80% of the observed wealth–health gradient, compared with only about 20% when measurement error is ignored. It also generates negative duration dependence in reported health, a pattern often attributed to permanent health heterogeneity.

These differences have substantial economic implications. Accounting for measurement error raises the estimated earnings cost of bad health by 44% for the population as a whole and more than doubles it for the initially unhealthy. Our decomposition shows that correcting labor-supply moments accounts for most of the difference in earnings costs for the population as a whole, whereas correcting the health and mortality process accounts for most of the difference in remaining-life monetary costs and longevity for those initially unhealthy. Correcting the health process alone is therefore insufficient to recover the full economic costs of bad health.

The broader lesson of our analysis extends beyond health. When a persistent economic state is observed through a noisy proxy, treating that proxy as the true state can understate both its persistence and its effects on economic behavior. It can also create the appearance that additional persistent heterogeneity is needed to explain the data. Measurement error therefore affects not only the quantitative predictions of structural models, but also the economic mechanisms they appear to require.

References

Altonji, J. G. and Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures, *Journal of Business & Economic Statistics* **14**(3): 353–366.

Amengual, D., Bueren, J. and Crego, J. A. (2021). Endogenous health groups and heterogeneous dynamics of the elderly, *Journal of Applied Econometrics* **36**(7): 878–897.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, *Econometrica* **68**(2): 399–405.

Andrews, I., Gentzkow, M. and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments, *The Quarterly Journal of Economics* **132**(4): 1553–1592.

Arnoud, A., Guvenen, F. and Kleineberg, T. (2019). Benchmarking global optimizers, *NBER Working Paper 26340*, National Bureau of Economic Research.

Benítez-Silva, H., Buchinsky, M., Chan, H. M., Cheidvasser, S. and Rust, J. (2004). How large is the bias in self-reported disability?, *Journal of Applied Econometrics* **19**(6): 649–670.

Blundell, R., Britton, J., Costa Dias, M. and French, E. (2023). The impact of health on labor supply near retirement, *Journal of Human Resources* **58**(1): 282–334.

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

Borella, M., De Nardi, M., Yang, F. and Torres Chain, J. P. (2025). Why do households save and work?, *NBER Working Paper 33874*, National Bureau of Economic Research.

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.

Dal Bianco, C. and Moro, A. (2026). The economic effects of nonlinear health dynamics: Estimates from a dynamic life-cycle model. Working paper, University of Padova and Vanderbilt University.

David, M., Little, R. J. A., Samuhel, M. E. and Triest, R. K. (1986). Alternative methods for CPS income imputation, *Journal of the American Statistical Association* **81**(393): 29–41.

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. (2025). The lifetime costs of bad health, *The Review of Economic Studies* **92**(3): 1987–2026.

Dwyer, D. S. and Mitchell, O. S. (1999). Health problems as determinants of retirement: Are self-rated measures endogenous?, *Journal of Health Economics* **18**(2): 173–193.

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.

Hosseini, R., Kopecky, K. A. and Zhao, K. (2026). How important is health inequality for lifetime earnings inequality?, *The Review of Economic Studies* **93**(1): 556–599.

Kerkhofs, M. and Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors, *Health Economics* **4**(3): 221–235.

Lange, F. and McKee, D. (2012). The evolution of latent health over the life course. Mimeo, Yale University.

Lindeboom, M. and Kerkhofs, M. (2009). Health and work of the elderly: Subjective health measures, reporting errors and endogeneity in the relationship between health and work, *Journal of Applied Econometrics* **24**(6): 1024–1046.

Little, R. J. A. (1988). Missing-data adjustments in large surveys, *Journal of Business & Economic Statistics* **6**(3): 287–296.

Low, H. and Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off, *American Economic Review* **105**(10): 2986–3029.

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.

Russo, N. (2026). Health-dependent preferences, consumption, and insurance. Working paper, Goethe University Frankfurt.

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, J. P. (2004). Unraveling the ses: health connection, *Population and development review* **30**: 108–132.

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

White, M. N. (2023). Self-reported health status and latent health dynamics, *Journal of Health Economics* **88**: 102727.

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

Appendix

Bad Health Is Worse Than It Looks: Measurement Error and Structural Inference

Martín García-Vázquez — Washington University in St. Louis
Yuta Maeda — University of Minnesota
Luis Pérez — Southern Methodist University

November 6, 2026

A Identification of the Measurement-Error Model **33**

B Data and Auxiliary Estimation **36**
B.1 Samples . . . 36
B.2 Emission Matrices, Health Transitions, and Mortality . . . 37
B.3 Spousal Income . . . 43
B.4 Wages . . . 46
B.5 Pension Parameters . . . 53
B.6 Assigned Parameters . . . 53

C Initial Distribution of States **54**
C.1 Average Earnings . . . 55
C.2 Health . . . 58
C.3 Wages . . . 60
C.4 Assets and Participation . . . 62
C.5 Predictive Mean Matching . . . 62
C.6 Sampling . . . 63

D Numerical Solution **63**

E Structural Estimation **65**
E.1 Weighting Matrix and Objective . . . 65
E.2 Minimization of the Indirect-Inference Objective . . . 66
E.3 Parameter Identification . . . 69
E.4 Construction of the Estimation Targets . . . 72
E.5 Additional Fit Results . . . 74

F Health Misclassification and Attenuation Bias **75**

G Additional Results and Robustness **80**
G.1 Welfare Calibration . . . 80
G.2 Additional Decomposition Results . . . 81
G.3 Pension Policy Counterfactuals . . . 83
G.4 Robustness to Alternative Health Indicators . . . 84

A Identification of the Measurement-Error Model

This appendix establishes identification of the latent-health process and its measurement system in the presence of mortality and staggered sample entry. We extend the identification approach for non-stationary hidden Markov models with refreshment samples in García-Vázquez (2021) to accommodate an absorbing, perfectly observed death state. Using the finite-mixture results of Bonhomme et al. (2016), we establish identification of the measurement system, the distribution of latent health among survivors, and age-specific health transitions, including mortality.

Environment. We distinguish age a from survey wave t because individuals enter the survey at different ages and observations from different cohorts are pooled to estimate age-specific health dynamics. Let A denote the number of age groups and T the number of survey waves. We allow $T < A$, so identification must rely on overlapping observations from different cohorts rather than complete individual health histories. Elsewhere in the paper, t indexes age groups, w indexes waves, and a denotes assets; the matrix K_a defined below collects the rows of the transition matrix Π_t of Section 2 for the two living states.

Let the underlying health state be

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

where states 1 and 2 denote good and bad health, respectively, and state 3 denotes death. Death is absorbing and perfectly observed. Health follows a first-order Markov process with transition probabilities that vary with age but not with cohort or calendar time. We suppress individual and survey-wave subscripts when describing age-specific distributions.

For each living individual, we observe three noisy health indicators, indexed by $m \in \{1,2,3\}$. Indicator m takes values in $\{1,\dots,J_m\}$ while the individual is alive. We augment its support with a death category, $J_m + 1$, which is observed without error.

Let P^m denote the $J_m \times 2$ emission matrix for indicator m among living individuals:

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

where p_c^m contains the probabilities of each indicator category conditional on health state $c \in \{1,2\}$. The corresponding emission matrix including death is

$$Q^m = \begin{pmatrix} P^m & 0_{J_m \times 1} \\ 0_{1\times 2} & 1 \end{pmatrix}.$$

We assume that the three indicators are conditionally independent given current latent health, and that observations of indicators at different ages are independent conditional on the corresponding latent health states. The emission distributions are invariant across ages and survey waves.

Define the distribution of latent health conditional on survival to age a by

$$\widetilde{\pi}_{ac} = \mathbb{P}(H_a = c \mid H_a \neq 3), \qquad c \in \{1,2\},$$

and let $\widetilde{\pi}_a = (\widetilde{\pi}_{a1}, \widetilde{\pi}_{a2})'$. For individuals alive at age a, define the 2×3 transition matrix K_a, with entries

$$K_a(c,d) = \mathbb{P}(H_{a+1} = d \mid H_a = c),$$

for $c \in \{1,2\}$ and $d \in \{1,2,3\}$. The third row of the full transition matrix is known because death is absorbing.

Proposition 1 (Identification with mortality and staggered sample entry). *Suppose that, for every age a under consideration, the following conditions hold:*

(i) Y_a^1, Y_a^2, Y_a^3 *are mutually independent conditional on* H_a, *and* Y_a^m *and* $Y_{a+1}^{m'}$ *are independent conditional on* (H_a, H_{a+1}) *for all* m, m';

(ii) $\operatorname{rank}(P^m) = 2$ *for* $m = 1,2,3$, *and* P^m *does not depend on* a *or* t;

(iii) $\widetilde{\pi}_{ac} > 0$ *for* $c = 1,2$;

(iv) $\mathbb{P}(Y_a^1, Y_a^2, Y_a^3 \mid H_a \neq 3)$ *and* $\mathbb{P}(Y_a^1, Y_{a+1}^1 \mid H_a \neq 3)$ *are identified from the pooled cohorts, with* $Y_{a+1}^1 = J_1 + 1$ *indicating death between* a *and* $a+1$;

(v) K_a *and* $\{P^m\}_{m=1}^3$ *are common to the cohorts and waves pooled in (iv), and sample entry and nonmortality attrition are independent of* $(H_a, Y_a^1, Y_a^2, Y_a^3)$ *conditional on* $H_a \neq 3$.

Then $\{P^m\}_{m=1}^3$, $\{\widetilde{\pi}_a\}_a$, *and* $\{K_a\}_a$ *are identified up to a common permutation of the two living health states* $\{1,2\}$. *In particular, the state-dependent mortality probabilities* $\mathbb{P}(H_{a+1} = 3 \mid H_a = c)$ *are identified for* $c = 1,2$ *and every* a *at which (iv) holds.*

Proof. We establish identification in two steps.

Step 1: Identification of the measurement system and of latent health among survivors. For each age a, the joint distribution of the indicators among survivors is identified from the pooled sample under conditions (iv) and (v), and condition (i) implies

$$\begin{aligned}\mathbb{P}(Y_a^1 = j_1, Y_a^2 = j_2, Y_a^3 = j_3 \mid H_a \neq 3) \\ = \sum_{c=1}^{2} \widetilde{\pi}_{ac} \prod_{m=1}^{3} p_c^m(j_m).\end{aligned}$$

This is a two-component finite mixture with three conditionally independent measurements. Since $\operatorname{rank}(P^m) = 2$ for $m = 1,2,3$ and $\widetilde{\pi}_{ac} > 0$, Theorems 2 and 3 of Bonhomme et al. (2016) identify $\{P^m\}_{m=1}^3$ and $\widetilde{\pi}_a$ up to a permutation of the two states. Because the emission distributions are invariant across ages, the same labeling applies throughout the age range. We resolve the remaining permutation by labeling the state associated with more favorable indicator distributions as good health. Q^m is then identified from P^m by adding the known entries.

Step 2: Identification of transitions and mortality. Fix an age a at which the consecutive-age distribution in (iv) is identified, and let M_a be the $J_1 \times (J_1 + 1)$ matrix with entries

$$M_a(j,k) = \mathbb{P}(Y_a^1 = j, Y_{a+1}^1 = k \mid H_a \neq 3).$$

By the Markov property and condition (i),

$$M_a = P^1 \operatorname{diag}(\widetilde{\pi}_a) K_a (Q^1)'.$$

All objects on the right-hand side except K_a have been identified in Step 1. Since P^1 has full column rank,

$$L_P \equiv \left[(P^1)'P^1\right]^{-1}(P^1)'$$

satisfies $L_P P^1 = I_2$. Similarly, Q^1 has full column rank, so

$$R_Q \equiv Q^1\left[(Q^1)'Q^1\right]^{-1}$$

satisfies $(Q^1)'R_Q = I_3$. Hence,

$$L_P M_a R_Q = \operatorname{diag}(\widetilde{\pi}_a) K_a,$$

and, since $\widetilde{\pi}_{ac} > 0$,

$$K_a = \operatorname{diag}(\widetilde{\pi}_a)^{-1}\left[(P^1)'P^1\right]^{-1}(P^1)'M_a Q^1\left[(Q^1)'Q^1\right]^{-1}.$$

Thus, K_a is identified. Its first two columns give transitions between the living health states, and its third column gives state-dependent mortality probabilities. Since death is absorbing, the remaining row of the full transition matrix is $(0,0,1)$. Repeating the argument at every age at which (iv) holds completes the proof. □

Remarks. The identification result has two implications for our application. First, mortality and transitions between living health states are jointly identified, although

they need not be estimated jointly. In Appendix B.2, we estimate mortality probabilities first and then recover transitions between the living states.

Second, identification does not require observing any individual throughout the entire age range: overlapping observations from different cohorts and refreshment samples suffice under conditions (iv) and (v). Identification is therefore age-specific and requires that the relevant distributions can be recovered at each age considered.

B Data and Auxiliary Estimation

This appendix describes the data and auxiliary estimation procedures used to construct inputs to the structural model. We document the estimation samples, the health measurement system and transitions, health-dependent spousal income and wages, pension benefits, and parameters fixed outside structural estimation. Throughout, we distinguish between specifications that account for measurement error in health and those that treat self-reported health as the true state.

B.1 Samples

To estimate wage profiles and wage-shock parameters, we use male core members born between 1940 and 1970. We exclude observations with missing or nonpositive wages, fewer than 300 annual hours worked, or earnings below the Jobseeker's Allowance threshold. The resulting sample contains 7,607 person-wave observations on 2,711 individuals. Hourly wages are measured as annual earnings divided by annual hours.

For spousal income, we use male core members from all birth cohorts who satisfy the marital and self-employment restrictions of the structural model. We exclude proxy interviews and assign zero spousal income to men without a partner. Spousal income includes the partner's net employment income, net self-employment income, and Carer's Allowance, as detailed in Appendix B.3. The resulting sample contains 21,510 person-wave observations with the required income and health information.

We construct health-specific profiles of labor-force participation and hours worked using men born between 1935 and 1950. For wealth, we extend this sample to include men born before 1935 to cover the oldest ages of the model. The estimation targets use 11,860 person-wave observations for labor supply and 17,782 observations for wealth.

ELSA observes individuals from age 50 onward, and waves 1–9 span at most seventeen years. Because individual histories do not span the entire life cycle, we pool observations across waves and birth cohorts into two-year age groups. The 1935–1950

cohorts cover much of the transition from employment into retirement, while the earlier cohorts extend the wealth profiles through ages 90–91.

The self-reported-health question differs across survey waves. In most waves, respondents classify their health as excellent, very good, good, fair, or poor. Wave 3 instead uses the categories very good, good, fair, bad, and very bad. In wave 1, half of respondents answer the alternative question at the beginning of the health module and the standard question at its end. We classify excellent, very good, and good responses to the standard question as good health, and fair and poor responses as bad health. In the specification that accounts for measurement error, we use the standard question whenever available, including the end-of-module question in wave 1, and treat self-reported health as missing by design in wave 3. In the specification that ignores measurement error, we use the alternative question in wave 3 and for the relevant wave-1 respondents, classifying bad and very bad responses as bad health.

B.2 Emission Matrices, Health Transitions, and Mortality

We estimate the non-stationary hidden Markov model described in the main text using an adaptation of the constrained Baum–Welch procedure developed by García-Vázquez (2021). The procedure accommodates age-dependent health transitions and staggered sample entry. We extend it to allow for mortality, which requires estimating state-dependent survival probabilities and adjusting the restrictions linking latent-health distributions across ages.

Estimation proceeds in three steps. We first estimate the measurement system and latent-health distributions, then state-dependent mortality probabilities, and finally transitions between living health states.

Let K_h denote the number of living health states, with death indexed by $K_h + 1$. Define

$$\begin{aligned}
\widetilde{\Pi}_t(i,j) &= \mathbb{P}(H_{t+1} = j \mid H_t = i, H_{t+1} \neq K_h + 1),\\
P^m(i,j) &= \mathbb{P}(Y_t^m = i \mid H_t = j),\\
\widetilde{\pi}_t(i) &= \mathbb{P}(H_t = i \mid H_t \neq K_h + 1),\\
p_t(i) &= \mathbb{P}(H_{t+1} = K_h + 1 \mid H_t = i),
\end{aligned}$$

for living states $i, j \in \{1, \ldots, K_h\}$. Thus, $\widetilde{\Pi}_t$ governs transitions between living states conditional on survival, $\widetilde{\pi}_t$ gives the distribution of latent health among survivors, and $p_t(i)$ is the probability of dying conditional on current health. The transition

probabilities entering the structural model are recovered from these objects:

$$\mathbb{P}(H_{t+1} = j \mid H_t = i) = [1 - p_t(i)]\,\widetilde{\Pi}_t(i,j), \qquad j \leqslant K_h,$$

with probability $p_t(i)$ assigned to death.

As established in Appendix A, the measurement system and age-specific transition probabilities are identified from cross-sectional and consecutive-age distributions of the health indicators, including observed mortality. The estimation procedure exploits this structure while imposing restrictions that ensure consistency between the age-specific distributions and transition matrices.

Because death is absorbing, the distribution of latent health among survivors evolves according to

$$\widetilde{\pi}_{t+1}(c) = \frac{\sum_{j=1}^{K_h} \widetilde{\pi}_t(j)[1 - p_t(j)]\,\widetilde{\Pi}_t(j,c)}{\sum_{j=1}^{K_h} \widetilde{\pi}_t(j)[1 - p_t(j)]}. \tag{B.1}$$

When mortality differs across health states, the distribution of health among individuals surviving to the next period differs from the distribution at the beginning of the current period. We therefore define the survival-adjusted distribution

$$\widetilde{\pi}^S_t(j) = \frac{\widetilde{\pi}_t(j)[1 - p_t(j)]}{\sum_{k=1}^{K_h} \widetilde{\pi}_t(k)[1 - p_t(k)]}. \tag{B.2}$$

The consistency restriction (B.1) can then be written as

$$\widetilde{\pi}_{t+1}(c) = \sum_{j=1}^{K_h} \widetilde{\pi}^S_t(j)\,\widetilde{\Pi}_t(j,c). \tag{B.3}$$

Thus, transitions conditional on survival map the survival-adjusted health distribution at t, rather than the initial distribution $\widetilde{\pi}_t$, into the distribution at $t+1$. This distinction is central to Step 3.

Data and sample construction. Throughout this subsection, $t = 1, \ldots, T$ indexes the model's two-year age groups. Individuals contribute an observation to an age group for each survey wave in which they are interviewed at that age. The resulting cross-sections pool observations across waves and birth cohorts, including refreshment samples.

We retain observations with valid mobility and ADL/IADL measures and exclude proxy interviews. Self-reported general health is unavailable in wave 3 and is treated as missing by design. We assume that item and unit nonresponse are independent of latent health conditional on age. For the longitudinal estimator, we further assume that

selection into the consecutive-wave sample is independent of the latent-health path, conditional on age and survival to the second observation.

Transition pairs consist of observations from consecutive survey waves in which the individual's age group advances by exactly one model period. We exclude the small number of pairs for which interview timing produces a change of zero or two age groups.

Steps 1 and 3 use waves 1–9. We exclude waves 10 and 11 because the COVID-19 pandemic disrupted the regular spacing of interviews and introduced changes in interview modes that could affect the stability of the measurement system.

Mortality estimation uses a more restricted observation window. We construct the death indicator D_{t+1} from harmonized vital-status records rather than from interview response. It equals one if a death is recorded between the interview at t and the following wave, and zero if the individual is known to be alive at the following wave, regardless of whether he is interviewed. Observations with unknown subsequent vital status, or deaths that cannot be assigned to a single two-year interval, are treated as censored and excluded.

We use mortality observations from origin waves 1–5, for which vital-status records are complete. Steps 1 and 3 use the broader panel, while Step 3 incorporates the mortality estimates through the survival-adjusted distribution. We therefore assume that age-specific mortality probabilities are stable across the survey periods used in these steps. The early-wave sample covers every model age group, so this assumption does not require extrapolating mortality across ages.

Let $\mathcal{A}_t$ denote the set of eligible observations on individuals alive at age t, and let $\mathcal{A}_t^D \subseteq \mathcal{A}_t$ denote the subset for which subsequent vital status is observed. For each observation, let $\mathcal{O}_{it} \subseteq \{1,2,3\}$ denote the set of observed health indicators.

Step 1: Measurement system and latent-health distributions. We jointly estimate the emission matrices $\{P^m\}_{m=1}^3$ and the age-specific latent-health distributions $\{\widetilde{\pi}_t\}_{t=1}^T$ by maximum likelihood, imposing common emission distributions across ages:

$$\max_{\{P^m\},\{\widetilde{\pi}_t\}} \sum_{t=1}^{T} \sum_{i \in \mathcal{A}_t} \log \left[\sum_{h=1}^{K_h} \widetilde{\pi}_t(h) \prod_{m \in \mathcal{O}_{it}} P^m(Y_{it}^m, h) \right].$$

The product includes only the indicators observed for each individual. In particular, wave-3 observations contribute through mobility and ADL/IADL limitations, while the missing self-reported-health indicator is integrated out. In implementation, an unobserved indicator contributes a factor of one to the likelihood.

We maximize the likelihood using the EM algorithm. In the E-step, we compute posterior probabilities of latent health using the observed indicators, the age-specific distribution, and the common emission matrices. In the M-step, we update each age-specific distribution by averaging the corresponding posterior probabilities and update the emission matrices using posterior-weighted counts pooled across ages. Pooling observations across ages improves the estimation of the common emission matrices and provides a consistent labeling of latent states across ages.

Because latent states are identified only up to permutation, we order them by the probability of reporting zero ADL/IADL limitations:

$$\widetilde{P}^3(1,1) > \widetilde{P}^3(1,2) > \cdots > \widetilde{P}^3(1,K_h).$$

Under this normalization, state 1 is the healthiest state. In our two-state model, states 1 and 2 denote good and bad health, respectively.

Step 2: State-dependent mortality. Although the mortality probabilities are identified nonparametrically under the conditions of Appendix A, the number of observed deaths at individual ages is limited. We therefore impose a logistic age profile with health-specific intercepts:

$$p_t(h) = \mathbb{P}(H_{t+1} = K_h + 1 \mid H_t = h) = \frac{\exp(b_h + d_1 t + d_2 t^2)}{1 + \exp(b_h + d_1 t + d_2 t^2)}.$$

We estimate $\{b_h\}_{h=1}^{K_h}$, d_1, and d_2 using observations in $\mathcal{A}_t^D$, treating latent health as unobserved and holding the measurement-system estimates from Step 1 fixed.

We estimate the mortality parameters using an EM algorithm. Given the current parameter estimates, the E-step computes the posterior probability of each latent-health state conditional on the observed health indicators and subsequent vital status:

$$\begin{aligned}\rho_{it}(h) &= \mathbb{P}(H_{it} = h \mid Y_{it}^{obs}, D_{i,t+1}) \\ &= \frac{\widetilde{\pi}_t(h)\,\mathbb{P}(D_{i,t+1} \mid H_{it} = h)\prod_{m\in\mathcal{O}_{it}} P^m(Y_{it}^m, h)}{\sum_{s=1}^{K_h} \widetilde{\pi}_t(s)\,\mathbb{P}(D_{i,t+1} \mid H_{it} = s)\prod_{m\in\mathcal{O}_{it}} P^m(Y_{it}^m, s)}.\end{aligned}$$

The M-step updates the mortality parameters by maximizing the posterior-weighted Bernoulli likelihood:

$$\sum_{t=1}^{T}\sum_{i\in\mathcal{A}_t^D}\sum_{h=1}^{K_h} \rho_{it}(h)\Big[D_{i,t+1}\log p_t(h) + (1 - D_{i,t+1})\log\big(1 - p_t(h)\big)\Big].$$

We iterate between these steps until convergence. The objective pools marginal likelihood contributions across person-period observations, using the estimated posterior probabilities to account for latent health.

Step 3: Transitions between living health states. For each $t = 1, \ldots, T-1$, we use observations from individuals interviewed alive in consecutive age groups to estimate the transition matrix conditional on survival, $\widetilde{\Pi}_t$. Under the maintained sampling assumptions, the origin-state distribution among observed pairs equals $\widetilde{\pi}_t^S$, the population health distribution conditional on survival to the next period. We use this distribution in the pairwise likelihood and estimate $\widetilde{\Pi}_t$ through a constrained EM algorithm.

In the E-step, given a candidate transition matrix $\widetilde{\Pi}_t^{(\iota)}$, we compute the posterior probability of each pair of latent states:

$$\widehat{\nu}_{i,k,j} = \frac{\ell_{it}(k)\,\widetilde{\pi}_t^S(k)\,\widetilde{\Pi}_t^{(\iota)}(k,j)\,\ell_{i,t+1}(j)}{\sum_{k'=1}^{K_h}\sum_{j'=1}^{K_h}\ell_{it}(k')\,\widetilde{\pi}_t^S(k')\,\widetilde{\Pi}_t^{(\iota)}(k',j')\,\ell_{i,t+1}(j')},$$

where

$$\ell_{i\tau}(k) = \prod_{m \in \mathcal{O}_{i\tau}} P^m(Y_{i\tau}^m, k).$$

The prior $\widetilde{\pi}_t^S$ accounts for differential mortality across health states. Although identification of transitions requires only one indicator at each date, we use all available indicators to improve precision.

In the M-step, we update the transition matrix by solving

$$\begin{aligned}
\widetilde{\Pi}_t^{(\iota+1)} = \arg\max_{\Pi} \quad & \sum_i \sum_{k=1}^{K_h}\sum_{j=1}^{K_h} \widehat{\nu}_{i,k,j} \log \Pi(k,j) \\
\text{s.t.} \quad & \sum_{j=1}^{K_h} \Pi(k,j) = 1 \quad \text{for all } k, \\
& \sum_{j=1}^{K_h} \Pi(j,c)\,\widetilde{\pi}_t^S(j) = \widetilde{\pi}_{t+1}(c) \quad \text{for all } c.
\end{aligned}$$

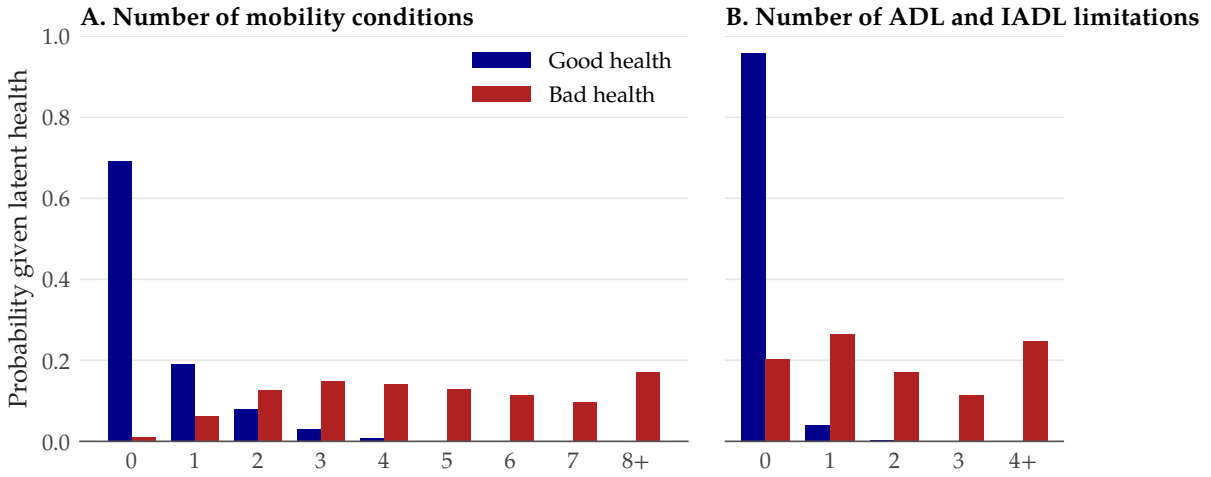
The second restriction imposes the consistency restriction (B.3) between the estimated transition matrix and the age-specific latent-health distributions.

The survival adjustment is required in the E-step regardless of whether the aggregation restriction is imposed. Under correct specification and the maintained sampling assumptions, the unconstrained estimator also identifies the transition probabilities. The constraint ensures that the estimated transitions and mortality process reproduce

the cross-sectional latent-health distributions. A full-information maximum-likelihood approach could impose this restriction by parameterizing later latent-health distributions as functions of all preceding transitions. Such an approach would make each transition matrix enter the likelihood of subsequent cohorts. The constrained Baum–Welch procedure instead imposes the restrictions separately for each age pair, substantially reducing the computational burden.

We estimate each step from 1,000 random starting values. Across convergent runs, the maximized objective values agree to within 10^{-9}, including in the age-specific transition problems of Step 3. This provides strong numerical evidence that the estimates are insensitive to initialization.

Figure B.1: Emission probabilities for mobility conditions and ADL/IADL limitations.



Notes. Each bar reports the estimated probability of observing a given number of conditions or limitations conditional on latent health.

Estimated emission distributions. Figure B.1 reports the estimated emission probabilities for mobility conditions and ADL/IADL limitations. The two states have sharply different distributions of physical limitations. Among individuals in good latent health, 69 percent report no mobility conditions and 96 percent report no ADL/IADL limitations. Only 4 percent report more than two mobility conditions, and almost none report more than one ADL/IADL limitation. Among individuals in bad latent health, the corresponding probabilities of reporting no conditions or limitations fall to 1 and 20 percent. Eight or more mobility conditions is the most likely category, while 25 percent report at least four ADL/IADL limitations. These differences help distinguish latent health states despite the misclassification in self-reported general health documented in Table 1.

B.3 Spousal Income

We estimate mean spousal income conditional on latent health, $\mathbb{E}[ys_t \mid H_t = c]$, for $c = 1, \ldots, K_h$. Because health is unobserved, identification requires that the indicators contain no additional information about spousal income conditional on latent health. Formally, we impose the following restriction.

Assumption 1 (Exclusion Restriction). $\mathbb{E}[ys_t \mid Y_t^1, Y_t^2, Y_t^3, H_t = c] = \mathbb{E}[ys_t \mid H_t = c]$.

The restriction requires that reporting errors do not predict spousal income once latent health is held fixed. One concern is justification bias, whereby economic circumstances affect how individuals report their health. Mobility and ADL/IADL limitations measure physical functioning rather than work capacity directly and may therefore be less susceptible to such bias than self-reported general health. Nevertheless, they may contain information about economic outcomes beyond that captured by binary latent health. As a robustness exercise, we replace self-reported general health with reported pain and re-estimate the model. This tests sensitivity to the subjective indicator but does not establish the validity of the broader restriction. The same assumption, with wages or labor supply in place of spousal income, underlies the corresponding auxiliary estimates and structural estimation targets.

Let $q_t(c \mid y)$ denote the posterior probability of health state c, given the vector of observed indicators $y = (y^1, y^2, y^3)$:

$$q_t(c \mid y) = \frac{\widetilde{\pi}_t(c) \prod_{m=1}^{3} P^m(y^m, c)}{\sum_{s=1}^{K_h} \widetilde{\pi}_t(s) \prod_{m=1}^{3} P^m(y^m, s)}.$$

Under Assumption 1,

$$\mathbb{E}[ys_t \mid Y_t = y] = \sum_{c=1}^{K_h} \mathbb{E}[ys_t \mid H_t = c]\, q_t(c \mid y). \tag{B.4}$$

The observed conditional mean is therefore a weighted average of the means by latent health, with weights given by the posterior probabilities. Identification follows even from a single indicator. In matrix notation,

$$\mathbb{E}[ys_t \mid Y_t^1] = \operatorname{diag}\left(\mathbb{P}(Y_t^1)\right)^{-1} \widetilde{P}^1 \operatorname{diag}(\widetilde{\pi}_t)\, \mathbb{E}[ys_t \mid H_t],$$

where the vectors on either side collect the corresponding conditional means. Because $\widetilde{P}^1$ has full column rank and the latent-state and retained indicator probabilities are positive, the matrix multiplying the latent-health means has full column rank. The health-specific means are therefore identified.

We estimate these means by minimum distance, imposing the restrictions implied by (B.4). Equivalently, within each age group, we regress spousal income on the vector of posterior probabilities without a constant:

$$ys_t = \sum_{c=1}^{K_h} \mathbb{E}[ys_t \mid H_t = c]\, q_t(c \mid Y_t) + \epsilon_t.$$

Since the posterior probabilities sum to one, including a constant would create perfect collinearity. Equation (B.4) implies $\mathbb{E}[\epsilon_t \mid Y_t] = 0$ and therefore $\mathbb{E}[\epsilon_t \mid q_t(\cdot \mid Y_t)] = 0$. Consequently, the population least-squares coefficients recover the latent-health conditional means provided that $\mathbb{E}[q_t(\cdot \mid Y_t)\, q_t(\cdot \mid Y_t)']$ is nonsingular. To see this, suppose that $\nu' q_t(\cdot \mid Y_t) = 0$ almost surely for some vector ν. Taking expectations conditional on Y_t^1 gives $\nu' \mathbb{P}(H_t = \cdot \mid Y_t^1) = 0$. Because the matrix of these posterior probabilities has full column rank, $\nu = 0$. Thus, the least-squares coefficients identify mean spousal income by latent health.

B.3.1 Construction of Spousal Income and Estimation Sample

We construct spousal income using the ELSA IFS and financial derived-variable files for waves 1–9. The measure refers to income received by the respondent's co-resident partner. We consider four definitions. The first, $ys^{(1)}$, includes net employment income; $ys^{(2)}$ adds Carer's Allowance; $ys^{(3)}$ adds net self-employment income instead; and $ys^{(4)}$ includes all three components. Net employment income comprises take-home pay and income from occasional jobs, while net self-employment income includes profits and drawings. We retain reported self-employment losses, which account for approximately 0.2 percent of person-wave observations. The structural model uses $ys^{(4)}$.

We exclude means-tested benefits, which are generated endogenously by the model's consumption floor; public and private pensions, which enter through the pension functions; annuity income, which is treated as asset income; and disability benefits, which are not explicitly modeled. Spousal income therefore captures partner-related resources not generated elsewhere in the structural model.

We include Carer's Allowance because it compensates caregiving rather than low consumption and cannot be reproduced by the model's consumption floor. Eligibility depends on care provision and earnings, with cash benefits paid only below the recipient's State Pension age. Comparing definitions with and without the allowance separates this transfer from the partner's labor income. Its annual value is approximately £2,900–£3,150 in 2013 prices over the sample period.

Income components are reported at weekly frequencies in the source data. We annualize them and deflate nominal values to 2013 prices using the ONS all-items

Consumer Price Index. Spousal income is measured net of taxes, while the respondent's own earnings are gross. Accordingly, the model taxes the respondent's income, while spousal income enters the budget constraint directly. Men without a co-resident partner receive zero spousal income, so the estimated function incorporates both partner earnings and partner absence.

The estimation sample consists of male core members satisfying the marital and self-employment restrictions of the main text, observed in waves with the required income and health information. We exclude proxy interviews because the health measurement system is estimated from respondents' own reports and may differ when information is provided by a proxy. Such interviews account for 3,734 of the 90,074 person-wave observations in waves 1–9 (4.1 percent). Most are excluded automatically because proxy respondents do not answer the relevant health questions.

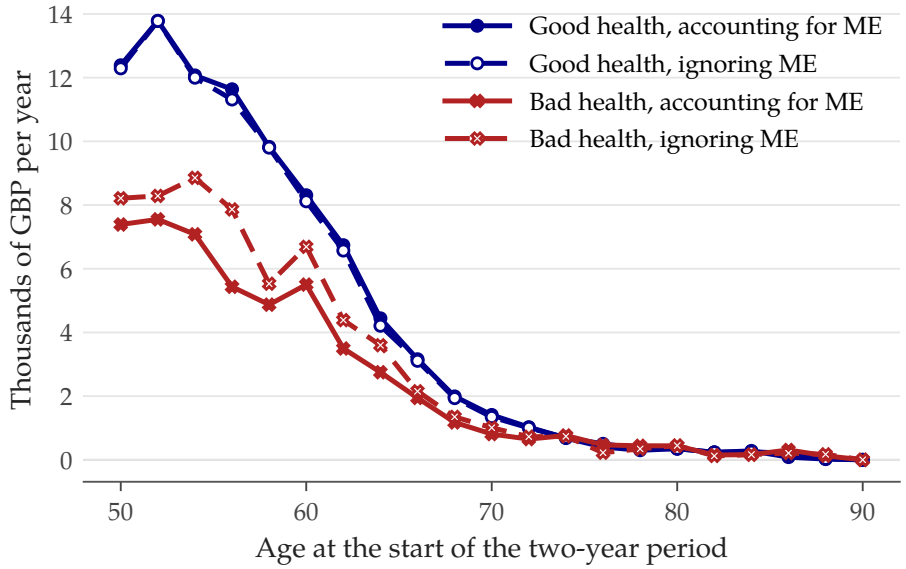
The resulting sample contains 21,981 person-wave observations. Of these, 120 fall outside the model's age range and 23 lack one of the required health measures. Spousal income is missing for approximately 1.6 percent of partnered observations, primarily when the partner is not interviewed; these observations are excluded under the maintained missing-at-random assumption. Including Carer's Allowance removes a further 0.5 percent because of missing benefit information. The baseline sample contains 21,510 person-wave observations.

B.3.2 Results

Figure B.2 compares the estimated spousal-income profiles across the specifications that account for and ignore measurement error. The profiles are similar for individuals in good health. For those in bad health, accounting for measurement error reduces estimated spousal income at working ages, by as much as £2,400 annually at ages 56–57. The resulting health gradient is steeper, consistent with self-reported health misclassifying some healthier individuals, whose partners earn more, as unhealthy.

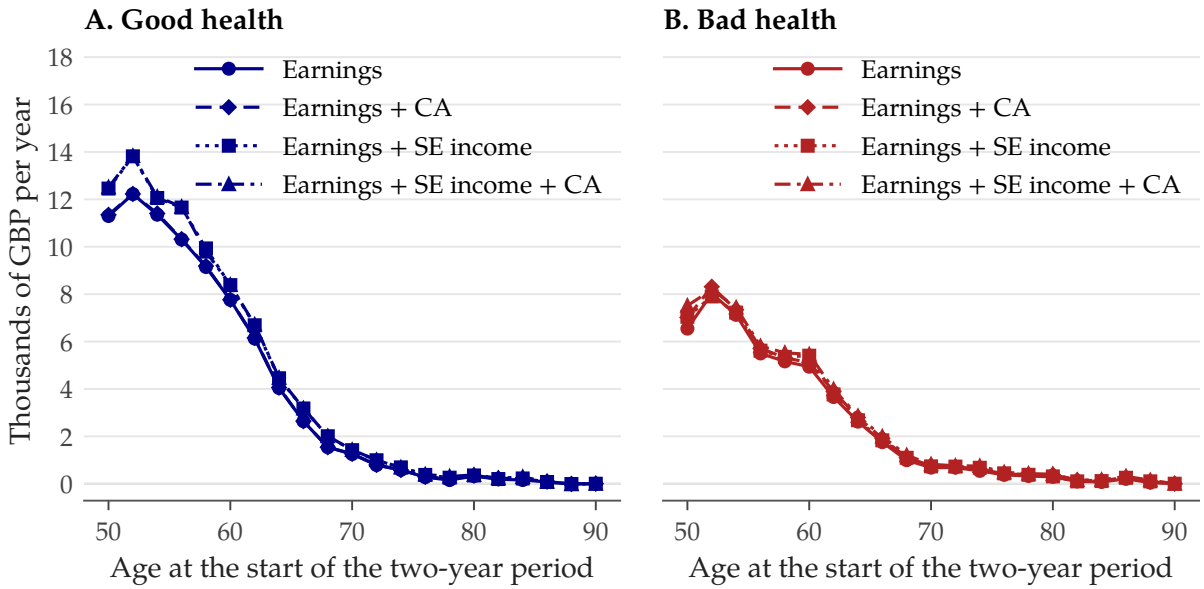
Figure B.3 compares the four definitions of spousal income. At ages 50–54, partners of men in bad health receive roughly half the income of partners of men in good health. Carer's Allowance contributes primarily to households with an unhealthy respondent, while self-employment income is more important among households in good health.

Figure B.2: Estimated spousal-income profiles by age and health.



Notes. Annual spousal income in pounds sterling, including the partner's net employment and self-employment income and Carer's Allowance.

Figure B.3: Spousal income by age, health, and income definition.



Notes. Annual spousal income in pounds sterling. CA denotes Carer's Allowance, and SE denotes net self-employment income.

B.4 Wages

In the model, log wages depend on a quadratic age profile, a health premium, and persistent idiosyncratic productivity:

$$\log\left(w_{it}^{\text{model}}\right) = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}_{\{\text{Health}_{it}=\text{Good}\}} + u_{it}. \tag{B.5}$$

The idiosyncratic component follows an AR(1) process,

$$u_{it} = \rho u_{i,t-1} + \xi_{it}, \qquad \rho \in (0,1),$$

with innovations

$$\xi_{i1} \sim \mathcal{N}(0, \sigma^2_{\xi,1}), \qquad \xi_{it} \sim \mathcal{N}(0, \sigma^2_{\xi}), \quad t > 1,$$

where the initial innovation variance may differ from the variance in subsequent periods.

In the data, hourly wages are observed with measurement error and health is observed through noisy indicators. For a given health regressor χ_{it}, we write observed log wages among employed individuals as

$$\log\left(\tilde{w}^{\text{data}}_{it}\right) = \tilde{a}_0 + \tilde{a}_1 t + \tilde{a}_2 t^2 + \tilde{a}_H \chi_{it} + \zeta_{it}, \tag{B.6}$$

where

$$\zeta_{it} = u_{it} + m_{it} + \tilde{a}_H e_{it}, \qquad e_{it} = \mathbf{1}_{\{\text{Health}_{it}=\text{Good}\}} - \chi_{it}.$$

The term m_{it} is serially uncorrelated measurement error in wages, with variance σ^2_m. The term e_{it} captures misclassification of the underlying health state.

We estimate two specifications that differ in how health enters the wage equation. When measurement error is ignored, χ_{it} is the indicator for good self-reported health. When measurement error is accounted for, it is the posterior probability of good latent health, following French (2005). Both specifications use the same sample, pairwise regression design, weights, and wage-shock moment conditions. The latent-health specification additionally corrects the residual moments for health measurement error.

The estimation has three steps. We first estimate the health process and measurement system as described in Appendix B.2. We then estimate the wage-profile coefficients from within-individual wage changes (Appendix B.4.1) and the wage-shock parameters, $\vartheta = (\rho, \sigma^2_{\xi,1}, \sigma^2_{\xi}, \sigma^2_m)$, by minimum distance using second moments of wage residuals (Appendices B.4.2–B.4.5). The wage level $\tilde{a}_0$ is obtained from the mean residual among employed individuals and used as an auxiliary statistic in structural estimation.

Observed wages are selected because employment depends on health and productivity. We account for this selection in the indirect-inference comparison by applying the same wage-profile regression to employed individuals in the data and simulated model. The auxiliary coefficients therefore incorporate selection in both samples. The wage-shock parameters, by contrast, are estimated outside the structural model and

are not selection-corrected, as in French (2005) and O'Dea (2018). We discuss the implications for the measurement-error correction below. Standard errors are obtained from a person-level bootstrap that re-estimates the auxiliary parameters in each replication.

B.4.1 Estimation of the Wage Profile

Consider two periods $s < t$ in which individual i is observed working. Differencing (B.6) gives

$$\Delta_{st} \log \tilde{w}_i = \tilde{a}_1 (t - s) + \tilde{a}_2 (t^2 - s^2) + \tilde{a}_H (\chi_{it} - \chi_{is}) + (u_{it} - u_{is}) + (m_{it} - m_{is}) + \tilde{a}_H (e_{it} - e_{is}). \tag{B.7}$$

Differencing eliminates the wage-level intercept and any time-invariant individual component. We estimate $(\tilde{a}_1, \tilde{a}_2, \tilde{a}_H)$ by least squares using all pairs of periods in which an individual works. Each pair is weighted by $1/T_i$, where T_i is the number of working observations for individual i. The regression contains no intercept.

For observation-specific regressors, the pairwise estimator is algebraically equivalent to the within estimator. To see this, note that

$$\sum_{s<t} (z_t - z_s)(v_t - v_s)' = T_i \sum_{t} (z_t - \bar{z}_i)(v_t - \bar{v}_i)'$$

for any two variables z and v. Thus, the $1/T_i$-weighted all-pairs regression coincides with the within estimator in the specification that uses self-reported health.

In the latent-health specification, the health regressor is instead pair-specific. For each pair (s, t), let $\mathcal{I}_{i,st}$ contain the indicators observed at both dates and the corresponding survival information, and define

$$\chi_{i\tau}^{st} = \mathbb{P}(H_{i\tau} = G \mid \mathcal{I}_{i,st}), \qquad \tau \in \{s, t\}.$$

We replace the change in the self-reported-health dummy with $\chi_{it}^{st} - \chi_{is}^{st}$. Since the posterior probability at a given date can depend on the other date in the pair, this estimator is not generally equivalent to a conventional within regression. It retains the same all-pairs design while using the information available for each pair to correct health misclassification.

Let

$$e_{i\tau}^{st} = \mathbf{1}\{H_{i\tau} = G\} - \chi_{i\tau}^{st}.$$

Under the correctly specified measurement system,

$$\mathbb{E}[e^{st}_{i\tau} \mid \mathcal{I}_{i,st}] = 0, \qquad \tau \in \{s, t\}.$$

It follows that

$$\mathbb{E}\big[(\chi^{st}_{it} - \chi^{st}_{is})(e^{st}_{it} - e^{st}_{is})\big] = 0.$$

The posterior therefore removes the mechanical attenuation from health misclassification. Consistency of the wage-profile coefficients additionally requires that the observed indicators do not predict changes in the wage shock, conditional on the information used to construct the posterior. Using all available pairs exploits wage variation over multiple horizons rather than restricting estimation to consecutive observations.

As a specification check, we estimate the health wage premium separately for pairs grouped by their distance $t-s$. Under the maintained assumptions, the pairwise posteriors are correctly calibrated at each horizon, so the estimated health premium should not systematically vary with pair length. The estimates, reported in the replication package, do not differ significantly across pair lengths, although the bootstrap confidence bands widen quickly with the distance between observations.

B.4.2 Identification of Wage-Shock Parameters

Let the residual from the wage equation be

$$\varepsilon_{it} = u_{it} + m_{it} + \tilde{\alpha}_H e_{it}. \tag{B.8}$$

We first establish identification from the moments of $u_{it} + m_{it}$, abstracting from health measurement error. We then describe how the corresponding measurement-error contribution is removed from the empirical moments.

Iterating the AR(1) process gives

$$u_{it} = \sum_{\tau=1}^{t} \rho^{t-\tau} \xi_{i\tau},$$

so that

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

These expressions imply

$$\begin{aligned}\operatorname{Cov}(\varepsilon_{i4}-\varepsilon_{i2},\varepsilon_{i1}) &= \rho(\rho^2-1)\sigma^2_{\xi,1},\\ \operatorname{Cov}(\varepsilon_{i3}-\varepsilon_{i2},\varepsilon_{i1}) &= \rho(\rho-1)\sigma^2_{\xi,1}.\end{aligned} \tag{B.9}$$

Their ratio identifies the persistence parameter:

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

Given ρ, either covariance identifies $\sigma^2_{\xi,1}$. The covariance

$$\operatorname{Cov}(\varepsilon_{i2}-\varepsilon_{i1},\varepsilon_{i1}) = (\rho-1)\sigma^2_{\xi,1}-\sigma^2_m \tag{B.10}$$

then identifies σ^2_m. Finally,

$$\operatorname{Var}(\varepsilon_{i2}-\varepsilon_{i1}) = 2\sigma^2_m+(\rho-1)^2\sigma^2_{\xi,1}+\sigma^2_\xi \tag{B.11}$$

identifies σ^2_ξ.

Together, these restrictions identify the four wage-shock parameters from second moments involving four periods, provided the relevant covariances are nonzero.

B.4.3 Overidentifying Restrictions

To improve precision, we supplement the identifying restrictions with additional variances and covariances implied by the wage-shock process. Let $V_t \equiv \operatorname{Var}(u_{it}) = \rho^{2(t-1)}\sigma^2_{\xi,1}+G_t\sigma^2_\xi$, where $G_t = \sum_{\tau=1}^{t-1}\rho^{2(\tau-1)}$. The additional moment restrictions are

$$\begin{aligned}
\operatorname{Var}(\varepsilon_{it}) &= V_t+\sigma^2_m, && t=1,\dots,14,\\
\operatorname{Cov}(\varepsilon_{it},\varepsilon_{is}) &= \rho^{t-s}V_s, && s\in\{1,2,3\},\ 1\leqslant t-s\leqslant 9,\\
\operatorname{Cov}(\varepsilon_{i,t+1},\varepsilon_{it}) &= \rho V_t, && t=4,\dots,13,\\
\operatorname{Cov}(\varepsilon_{i,t+1}-\varepsilon_{it},\varepsilon_{it}) &= (\rho-1)V_t-\sigma^2_m, && t=1,\dots,13,\\
\operatorname{Cov}(\varepsilon_{i,t+2}-\varepsilon_{it},\varepsilon_{it}) &= (\rho^2-1)V_t-\sigma^2_m, && t=1,\dots,12,\\
\operatorname{Cov}(\varepsilon_{i3}-\varepsilon_{i2},\varepsilon_{i3}) &= \rho^3(\rho-1)\sigma^2_{\xi,1}+\left[1+\rho(\rho-1)\right]\sigma^2_\xi+\sigma^2_m,\\
\operatorname{Cov}(\varepsilon_{i2}-\varepsilon_{i1},\varepsilon_{i2}) &= \rho(\rho-1)\sigma^2_{\xi,1}+\sigma^2_\xi+\sigma^2_m.
\end{aligned} \tag{B.12}$$

Together with the identifying moments in (B.9)–(B.11), these restrictions yield 81 statistics. We retain a statistic only if at least 30 individuals contribute to its sample counterpart and apply the same rule in each bootstrap replication.

B.4.4 Correcting the Residual Moments for Health Misclassification

When health is measured with error, wage residuals contain the additional component $\tilde{a}_H e_{it}$. Under the maintained orthogonality assumptions, their second moments equal those of the wage shock and wage measurement error plus the contribution of health misclassification. Because this contribution is determined by the estimated health process and measurement system, we subtract it before estimating the wage-shock parameters.

The correction uses posteriors constructed from the information relevant to each moment. For variances, we condition on indicators observed at that date; for two-date moments, on the indicators and survival information at both dates; and for three-date moments, on the corresponding three-date information. The posterior, and hence the residual, is therefore specific to the statistic being computed.

Let q_{it} denote the posterior probability of good health using information from a single date. For a pair (a, b), let $p_{ab}^{(a)}$ and $p_{ab}^{(b)}$ denote the posterior probabilities of good health at each date, and let J_{ab} denote the joint posterior probability of good health at both dates. The required moments include

$$\begin{aligned} \mathrm{Var}(e_{it}) = \mathbb{E}[q_{it}(1-q_{it})], \qquad & \mathrm{Cov}(e_{ia}, e_{ib}) = \mathbb{E}\big[J_{ab} - p_{ab}^{(a)} p_{ab}^{(b)}\big], \\ & \mathrm{Var}(e_{ia})\big|_{(a,b)} = \mathbb{E}\big[p_{ab}^{(a)}\big(1-p_{ab}^{(a)}\big)\big]. \end{aligned} \tag{B.13}$$

The last expression applies to a pair-specific residual. Moments involving differences or multiple residuals are constructed from these expressions. For each statistic, we subtract the relevant health measurement-error moment multiplied by $\hat{\tilde{a}}_H^2$, then estimate the wage-shock parameters using the restrictions above. No correction is applied when self-reported health is treated as the true state.

The correction is exact under the assumed orthogonality of wage shocks and health measurement errors. Among workers, however, employment selection can induce dependence between wage shocks and latent health, while the posterior health probabilities are constructed using population rather than employment-conditional health distributions. The cross term between wage shocks and classification errors therefore need not vanish in the selected sample.

Indirect inference applies the same wage-profile regression to observed and simulated workers, incorporating employment selection into the matched auxiliary coefficients. The wage-shock parameters, however, are estimated outside the structural model and are not selection-corrected. At the estimated health wage premium, the measurement-error correction is small, and the resulting wage-shock parameter estimates are similar across specifications.

B.4.5 Estimation of Wage Parameters by Minimum Distance

We estimate the wage-shock parameters by minimum distance:

$$\hat{\vartheta} = \arg\min_{\vartheta \in \Theta} Q(\vartheta),$$

where

$$Q(\vartheta) = \frac{1}{n}\big[\widehat{\pi} - \boldsymbol{h}(\vartheta)\big]' \big[\widehat{\pi} - \boldsymbol{h}(\vartheta)\big]. \tag{B.14}$$

Here, n is the number of retained moments, $\widehat{\pi}$ contains their empirical counterparts after correcting for health measurement error where appropriate, and $\boldsymbol{h}(\vartheta)$ contains the corresponding theoretical moments. The baseline uses identity weighting, following Altonji and Segal (1996), who show that estimated optimal weighting matrices can introduce substantial finite-sample bias in covariance-structure estimation. We also compute inverse-variance-weighted estimates as a sensitivity check. The estimates reported in Table B.1 use identity weighting.

Conditional on ρ, the theoretical moments are linear in $(\sigma^2_{\xi,1}, \sigma^2_{\xi}, \sigma^2_{m})$. We exploit this structure by concentrating the objective: for each candidate ρ, we estimate the variance parameters by least squares, then minimize the concentrated objective over ρ, refining the solution through a one-dimensional search. We use the unconstrained estimator for inference and compute the solution imposing nonnegative variances as a diagnostic. This avoids imposing boundary constraints that can generate nonstandard bootstrap behavior (Andrews, 2000). Standard errors and confidence intervals are computed from 250 person-level bootstrap replications. In each replication, we re-estimate the health process, measurement system, wage-profile coefficients, and wage-shock parameters, incorporating sampling uncertainty from all three steps. Table B.1 reports the estimates.

Table B.1: Minimum-distance estimates of wage-shock parameters.

Parameter	Ignoring ME	Accounting for ME
ρ	0.9431	0.9446
	(0.041)	(0.041)
$\sigma^2_{\xi,1}$	0.2397	0.2374
	(0.047)	(0.047)
σ^2_{ξ}	0.0378	0.0371
	(0.018)	(0.018)
σ^2_{m}	0.1103	0.1109
	(0.021)	(0.021)

Notes. Estimates use identity weighting. Standard errors from 250 person-level bootstrap replications are reported in parentheses.

B.5 Pension Parameters

Following O'Dea (2018), we approximate public and private pension benefits as quadratic functions of average lifetime earnings. We estimate

$$\mathrm{pbb}_{i,65} = \mathrm{ss}_1 \mathrm{ae}_{i,64} + \mathrm{ss}_2 \mathrm{ae}_{i,64}^2 + e_i^{ss}, \qquad \mathrm{ae}_{i,64} \leqslant \widehat{\mathrm{ae}}_{ss}, \tag{B.15}$$

and

$$\mathrm{privben}_{i,65} = \mathrm{pp}_0 + \mathrm{pp}_1 \mathrm{ae}_{i,64} + \mathrm{pp}_2 \mathrm{ae}_{i,64}^2 + e_i^{pp}. \tag{B.16}$$

Here, $\mathrm{pbb}_{i,65}$ and $\mathrm{privben}_{i,65}$ denote household public and private pension benefits at age 65, and $\mathrm{ae}_{i,64}$ denotes average lifetime earnings at age 64. The public pension function is quadratic up to the earnings threshold $\widehat{\mathrm{ae}}_{ss} = 77{,}163$ and constant above it. Private pension benefits follow the estimated quadratic specification. Table B.2 reports the parameter estimates.

Table B.2: Estimated pension parameters

Parameter	Estimate	Standard error
ss_1	0.6518	0.0006
ss_2	-3.56×10^{-6}	1.31×10^{-8}
pp_0	5,980.80	591.35
pp_1	0.3426	0.0266
pp_2	5.23×10^{-7}	2.45×10^{-7}

Average lifetime earnings are revalued according to the growth of national average earnings, consistent with the UK pension system and the accumulation rule in the main text. At age 50, fifteen two-year periods have elapsed since labor-market entry at age 20, so we set $n_0 = 15$. Appendix C.1 describes the imputation of initial average earnings.

B.6 Assigned Parameters

We fix three parameters outside structural estimation. We set the coefficient of relative risk aversion to $\gamma = 3$, the time endowment to 8,760 hours per two-year period, and the annual real interest rate to 4 percent, which implies a return of $(1.04)^2 - 1 \approx 8.2$ percent per two-year model period. Table B.3 lists these parameters.

Table B.3: Assigned parameters

Parameter	Value	Justification
γ: coefficient of relative risk aversion	3	Standard in the literature
$\bar{L}$: hours per two-year period	8,760	Twelve hours per day
r: annual real interest rate	0.04	Standard in the literature

We also fix the income tax schedule. Following O'Dea (2018), we approximate the UK tax system using

$$\mathrm{Tax}(ti, t) = \begin{cases} 0, & ti \leqslant \tau_1^t, \\ 0.2(ti - \tau_1^t), & \tau_1^t < ti \leqslant \tau_2^t, \\ 0.2(\tau_2^t - \tau_1^t) + 0.4(ti - \tau_2^t), & ti > \tau_2^t. \end{cases}$$

Taxable income ti includes labor earnings, asset income, public pension payments, and private pension payments. The latter two enter only after retirement. Table B.4 reports the age-specific tax thresholds. We multiply the annual thresholds in O'Dea (2018) by two to match the model's biennial frequency and assign them to the corresponding two-year age groups.

Table B.4: Income tax thresholds for the two-year model period

	Age		
Parameter	< 64	64–73	$\geqslant 74$
τ_1	16,210	21,000	21,200
τ_2	84,940	89,740	89,940

C Initial Distribution of States

The model requires a joint distribution of initial states comprising wealth a, average earnings ae, health H, the persistent wage shock u, and labor-force participation. Wealth and participation are observed in ELSA, but average earnings are not. Health is observed through noisy indicators, while the wage shock can be inferred from observed wages only for individuals who work. Following French and Jones (2011), we initialize each individual at his first interview, using observed states where available and imputing the remaining components. The empirical joint distribution of these state vectors provides the initial conditions for the model, from which individuals are sampled with replacement. This appendix describes their construction and the validation of the imputation procedures.

Sample and timing. We restrict attention to male core members born between 1935 and 1950. This is the cohort simulated by O'Dea (2018), whose data we use to estimate pension benefits and impute average earnings. All men in this cohort faced a State Pension age of 65 under the pre-2016 rules, so the pension functions estimated in Appendix B.5 apply throughout the sample.[11]

We use individuals who are married at their first interview, are not subsequently observed to divorce or remarry, and are self-employed in no more than half of their employed waves, drawing on waves 1–9 of ELSA. Each individual enters the model at the two-year age group t_0 of his first interview, which ranges from ages 52–53 to 74–75. Of these individuals, 83 percent enter at wave 1 and the remainder at the wave-4 refreshment.

Throughout, t indexes the model's two-year age groups. Let x_i denote the covariates used in the imputations: a quadratic in entry age t_0, the posterior probability of good health defined in Appendix C.2, signed log wealth $\text{sgn}(a)\log(1+|a|)$, an indicator for leaving full-time education after age 15, and an indicator for access to a defined-benefit pension. The variables refer to characteristics at entry, although their construction may use information observed in subsequent waves, as described below.

C.1 Average Earnings

Because ELSA interviews individuals only from age 50 onward, their complete earnings histories are unavailable. Although the ELSA Life History Interview provides retrospective employment and earnings information, imputations based on these histories yield a weak relationship between average earnings and pension benefits, unlike the patterns documented using UK administrative records. We therefore supplement ELSA with simulated earnings histories from O'Dea (2018), whose model reproduces earnings profiles observed in UK administrative data.[12]

Our imputation uses ELSA information to rank individuals within the distribution of average earnings in the simulated data. The resulting assignment preserves dependence between average earnings and other initial states to the extent that the ranking variable is informative. We use the same simulated data to estimate the pension functions in Appendix B.5, ensuring that the earnings measure is defined consistently across the model and its initial conditions.

[11] Men born on or after 6 April 1951 are covered by the new State Pension introduced in April 2016, while men born after 5 December 1953 face a State Pension age above 65. These changes do not affect our cohort. Waves 10 and 11 are excluded for the reasons described in Appendix B.2.

[12] The ELSA Life History Interview, conducted alongside wave 3, collects retrospective information on employment and earnings. Reconstructing earnings histories requires assumptions about wage growth within employment spells, historical currency conversions, and the tax treatment of reported earnings. These limitations make the implied average-earnings measures noisy.

C.1.1 The Earnings Measure

Under the UK State Earnings-Related Pension Scheme and State Second Pension, past earnings are revalued with the growth of national average earnings before being averaged. In constant prices, this revaluation is governed by the real growth rate g. The model's law of motion for average earnings is

$$ae_{t+1} = \frac{W_t N_t + (15 + t - 1)(1+g)^2 \, ae_t}{15 + t}, \qquad t = 1, \ldots, 7. \tag{C.1}$$

Average earnings remain constant from age 64 onward. The factor $(1+g)^2$ accounts for the model's biennial frequency. The simulated data of O'Dea (2018) use the annual counterpart of this law of motion,

$$ae_{\tau+1} = \frac{(n-1)(1+g)\, ae_\tau + e_\tau}{n},$$

where n is the number of years since labor-market entry and e_τ denotes annual earnings. Reconstructing individual earnings histories from age 20 yields $g = 0.0184$ and reproduces the simulated average earnings to within £0.01 at every age. The imputed initial state therefore corresponds to the same revalued earnings measure used in the model and in the pension functions of Appendix B.5.

C.1.2 Imputation by Rank

We impute average earnings using a rank-based procedure. Within household types, we rank ELSA individuals by private pension income and assign average earnings from similarly ranked households in O'Dea's simulated data. The key assumption is that private pension income provides a comparable ranking of average earnings across the two datasets. The procedure has four steps.

Household type. O'Dea classifies households into four types based on education (leaving full-time education at age 15 or earlier versus later) and access to a defined-benefit pension. We construct the same classification. Education is measured by the age at which the respondent completed full-time schooling. Defined-benefit access is identified from pension histories, pension-wealth files, and current scheme contributions. We classify an individual as having access if he has ever reported an employer defined-benefit pension, holds defined-benefit pension wealth in waves 2–5, or contributes to such a scheme in any wave. Pension records reflecting only refunds of contributions are excluded. Among individuals born between 1941 and 1950, 48 percent have defined-benefit access, compared with 50 percent in O'Dea's data.

Ranking variable. We rank individuals by their own private pension income, annualized and expressed in 2013 prices. We use the first observation at or after State Pension age, provided it is no more than six years later. Individuals without private pension income form a tied group at the bottom of the distribution. For the two defined-benefit types, however, we treat zero private pension income as missing because it may reflect deferral, commutation, or a pension held in the spouse's name. We break ties using the individual's own State Pension income and, if necessary, randomly. Private pension income is more informative about individuals' economic circumstances than State Pension income within household type. In ELSA, its rank correlations with initial wealth and wages are 0.26 and 0.32, compared with -0.04 and 0.04 for the individual's own State Pension, and -0.02 and 0.04 for the couple's State Pension. State Pension benefits are relatively compressed for this cohort, with variation partly reflecting contribution histories, contracting out, deferral, and reporting rather than lifetime earnings.

Ranks. Within cells defined by household type and the wave group in which pension income is observed (waves 1–4, 5–7, or 8–11; unlike the health histories, pension readings are taken from all eleven waves), we assign percentile rank

$$q_i = \frac{r_i - 1}{N_c - 1},$$

where r_i is the individual's rank and N_c is the number of observations in the cell. Ranking within wave groups accounts for changes in the level of pensions in payment over the sample period.

Donors. For each household type j, the 1,000 simulated households of O'Dea are ranked by the head's private pension income at age 65. An ELSA individual with percentile rank q_i is matched to a nearby simulated household using donor position

$$\pi_i = \min\left\{\max\{1 + \text{round}(999\, q_i) + v_i,\ 1\},\ 1000\right\}, \qquad v_i \sim \text{Uniform}\{-25, \ldots, 25\}. \quad \text{(C.2)}$$

Thus, each individual draws from a window of 51 positions centered on his rank, with donor positions capped at the boundaries. This introduces variation in average earnings among individuals with similar pensions rather than assigning them all the same donor, while preserving the information contained in their ranking. Each individual receives the donor's average earnings at his own entry age. For individuals first observed after age 65, we use the donor's average earnings at 65, when the measure is frozen. The imputed value ae_{i,t_0} uses the revalued earnings definition in (C.1) and is converted to biennial units by multiplying by two.

C.1.3 Validation

The imputation requires private pension income to be informative about average earnings within household type. In O'Dea's simulated data, the rank correlations between private pension income and average earnings at age 64 are 0.86, 0.78, 0.35, and 0.97 across the four types, with similar values at ages 50–65. The ranking is informative for three types but much weaker for the high-education type without defined-benefit access, whose private pension income is more exposed to returns on defined-contribution wealth. In ELSA, the corresponding within-type correlations of imputed average earnings with initial wealth and wages are (0.12, 0.07), (0.23, 0.36), (−0.01, −0.04), and (0.36, 0.45).

As an additional check, we construct an alternative imputation that ranks working individuals by earnings at entry and matches them to simulated households ranked by earnings at the same age. The two imputations have within-type rank correlations ranging from 0.12 to 0.27. The evidence indicates that the imputation is most informative for the two types with defined-benefit access, less informative for the low-education type without it, and provides little individual-specific information for the high-education type without defined-benefit access. The simulated data determine the distribution and level of imputed average earnings within each type, while ELSA supplies individuals' relative positions to the extent supported by the observed pension information.

Individuals with no pension reading—those not observed at State Pension age within the available waves, or individuals in defined-benefit types with a zero reading—receive average earnings using the predictive mean matching procedure in Appendix C.5. The procedure is estimated on individuals with a rank-based imputation.

C.2 Health

We initialize health by drawing from its posterior distribution at entry, conditional on the individual's observed health history. The posterior combines all available health indicators from waves 1–9 and, when observed, vital status following the last interview. It uses the emission matrices, health transitions, mortality probabilities, and age-specific latent-health distributions estimated in Appendix B.2.

Let $t_0 = b_1 < b_2 < \cdots < b_n$ denote the age groups in which individual i is interviewed. Because the individual is alive at each interview, $H_{b_k} \neq K_h + 1$ for $k = 1, \ldots, n$. Let $Y_k = (Y^1_{b_k}, Y^2_{b_k}, Y^3_{b_k})$ denote the health indicators at the k-th interview, and let D_{b_n+1} indicate death in the period following the last interview. The death indicator is available for origin waves 1–5 and otherwise treated as censored. Define the emission likelihood

at interview k and the transition probabilities between living health states by

$$\ell_k(h) = \prod_{m:\, Y^m_{b_k} \text{ observed}} P^m\big(Y^m_{b_k}, h\big),$$

$$\Pi_t(h,h') = \big[1 - p_t(h)\big]\,\widetilde{\Pi}_t(h,h') = \mathbb{P}\big(H_{t+1} = h' \mid H_t = h\big), \qquad h' \leqslant K_h.$$

For interviews separated by $\Delta = b_{k+1} - b_k$ age groups, the transition matrix over the interval is

$$\Pi_{b_k \to b_{k+1}} = \Pi_{b_k}\Pi_{b_k+1}\cdots\Pi_{b_{k+1}-1}.$$

Its (h,h') entry gives the probability of surviving to the next interview in health state h', conditional on health state h at the previous interview. When interviews are separated by more than one age group, the matrix product integrates over the intervening latent-health states.

By the Markov property and conditional independence of the indicators, the joint probability of the observed history and a latent-health path $(h_1,\ldots,h_n)$ is

$$\widetilde{\pi}_{b_1}(h_1)\prod_{k=1}^{n}\ell_k(h_k)\prod_{k=1}^{n-1}\Pi_{b_k\to b_{k+1}}(h_k,h_{k+1})\;B_n(h_n),$$

where

$$B_n(h) = \begin{cases} p_{b_n}(h), & D_{b_n+1} = 1, \\ 1 - p_{b_n}(h), & D_{b_n+1} = 0, \\ 1, & D_{b_n+1} \text{ is censored.} \end{cases}$$

The prior $\widetilde{\pi}_{b_1}$ is the estimated distribution of latent health among survivors at entry. We integrate over subsequent latent-health states using the backward recursion

$$B_k(h) = \sum_{h'=1}^{K_h}\Pi_{b_k\to b_{k+1}}(h,h')\,\ell_{k+1}(h')\,B_{k+1}(h'), \qquad k = n-1,\ldots,1.$$

Here, $B_k(h)$ is the likelihood of all information observed after interview k, conditional on latent health h at that interview. Combining this likelihood with the entry-age prior and the initial emission likelihood gives

$$\mathbb{P}\big(H_{t_0} = h \mid Y_1,\ldots,Y_n, D_{b_n+1}\big) = \frac{\widetilde{\pi}_{t_0}(h)\,\ell_1(h)\,B_1(h)}{\sum_{h'=1}^{K_h}\widetilde{\pi}_{t_0}(h')\,\ell_1(h')\,B_1(h')}. \tag{C.3}$$

We draw initial health H_{i,t_0} from (C.3) and use the posterior probability of good health as a covariate in subsequent imputations. Because health evolves exogenously in the model, subsequent health observations can inform its initial state without requiring a model of economic decisions. The resulting posterior is highly concentrated: the

median of the largest posterior state probability is 1.000, and this probability exceeds 0.90 for 93 percent of individuals.

The posterior does not condition on wealth or average earnings, which may contain additional information about health before entry. Its high concentration suggests that incorporating this information would change the modal health classification for relatively few individuals, although it does not establish that the joint distribution of health and economic states is unaffected. The posterior uses the same measurement system and health dynamics as the wage estimation in Appendix B.4, but conditions on the full observed health history rather than the information relevant to a particular wage moment.

C.3 Wages

We initialize the persistent wage shock u_{t_0} using observed wages when available and imputations based on individual characteristics otherwise. For workers with a usable wage at entry, the residual from the wage equation is

$$z_i = \log \tilde{w}_{i,t_0} - \left(\tilde{a}_0 + \tilde{a}_1 t_0 + \tilde{a}_2 t_0^2 + \tilde{a}_H \mathbf{1}\{H_{i,t_0} = G\}\right) = u_{i,t_0} + m_{i,t_0}.$$

The health indicator is evaluated at the drawn initial state. The residual contains the persistent wage shock and wage measurement error. For nonworkers and individuals without a usable wage, this residual is unobserved. We use a common prior for the wage shock conditional on observable characteristics and update it using the wage residual when available.[13]

Prior. Using workers with observed wages, we estimate

$$z_i = x_i'\lambda + \zeta_i,$$

with x_i augmented by imputed average earnings ae_{i,t_0}. Under classical wage measurement error, the fitted value estimates the conditional mean of the persistent wage shock among workers. The residual variance combines persistent wage variation and measurement error. We therefore define

$$p_i = x_i'\hat{\lambda}, \qquad \sigma^2_{u|x} = \max\left\{\hat{\sigma}^2_\zeta - \hat{\sigma}^2_m, 0\right\},$$

[13] Appendix B.4 uses posterior probabilities of good health to estimate wage profiles and correct wage-residual moments for health misclassification. Here, health has already been drawn as an initial state, so the corresponding health indicator enters the wage equation directly.

and specify the conditional prior

$$u_{i,t_0} \mid x_i \sim \mathcal{N}(p_i,\ \sigma^2_{u|x}). \tag{C.4}$$

The measurement-error variance $\hat{\sigma}^2_m$ is estimated in Appendix B.4. The prior thus combines a conditional mean estimated among workers with residual dispersion net of wage measurement error.

Workers. Given the prior and the observed residual $z_i = u_{i,t_0} + m_{i,t_0}$, with independent Gaussian measurement error, the posterior is

$$u_{i,t_0} \mid z_i, x_i \sim \mathcal{N}\Big(p_i + \omega\,(z_i - p_i),\ (1-\omega)\,\sigma^2_{u|x}\Big), \tag{C.5}$$

where

$$\omega = \frac{\sigma^2_{u|x}}{\sigma^2_{u|x} + \hat{\sigma}^2_m}.$$

We draw the initial wage shock from this posterior. The observed wage residual is shrunk toward its predicted mean, with the weight determined by the relative variances of the persistent shock and measurement error. Importantly, drawing from the posterior rather than assigning its mean preserves the dispersion of the initial state. Under the Gaussian measurement model, the posterior mean has conditional-on-x_i variance $\omega\,\sigma^2_{u|x}$ across wage observations, whereas posterior draws have variance $\sigma^2_{u|x}$.

Nonworkers. For nonworkers, we follow French and Jones (2011) and impute a wage residual using workers with similar characteristics. We sort individuals into deciles of predicted shocks p_i, randomly select a working donor j from the same decile, and set

$$z_i = p_i + \hat{\zeta}_j.$$

We then treat this imputed residual as a signal and draw the initial wage shock from (C.5). The donor residual contains both persistent wage variation and measurement error. Under the maintained variance assumptions, combining shrinkage with posterior uncertainty preserves the prior variance:

$$\omega^2\left(\sigma^2_{u|x} + \sigma^2_m\right) + (1-\omega)\,\sigma^2_{u|x} = \sigma^2_{u|x}.$$

Thus, the imputation retains the estimated dispersion of wage shocks while using the empirical residual distribution to generate variation among nonworkers. Unlike French and Jones (2011), we cannot use wages from the individual's last job because this information is unavailable. A limitation is that the prediction equation is estimated on workers, whose wage shocks may differ systematically from those of nonworkers

with the same characteristics. We do not correct for this selection, which cannot be identified from observed wages alone.

Comparison with the wage process. Following French and Jones (2011), we estimate the dispersion of initial wage shocks from observed wages rather than imposing the distribution implied by the estimated wage process. The latter implies

$$\mathrm{Var}(u_t) = \rho^{2(t-1)}\sigma^2_{\xi,1} + \sigma^2_{\xi}\,\frac{1-\rho^{2(t-1)}}{1-\rho^2}.$$

The data-based initialization implies somewhat less dispersion. For individuals entering at ages 52–65, the standard deviation of initialized shocks ranges from 0.45 to 0.50, compared with 0.50–0.54 implied by the estimated wage process. For those entering at age 66 or older, the corresponding ranges are 0.34–0.43 and 0.55–0.56. The larger discrepancy at older ages reflects the scarcity of observed wages: only 35 of 429 individuals entering at age 66 or older report a wage. Most initial shocks at these ages are therefore imputed from workers with similar characteristics. We retain the data-based initialization rather than imposing the dispersion from the wage process, which is estimated on men born between 1940 and 1970. Among workers in our initialization cohort, the wage residual net of measurement error has a standard deviation of 0.46 at the first interview.

C.4 Assets and Participation

Total wealth, including housing and non-housing assets, is observed at entry. We set negative wealth to zero, consistent with the model's borrowing constraint, and winsorize values above the 99th percentile. Following French and Jones (2011), individuals participate in the labor market if annual hours exceed 300. Individuals who report being employed but whose weekly hours are missing or below five are assigned 20 hours per week and classified as working.

C.5 Predictive Mean Matching

For variables observed only for a subset of individuals, we use the predictive mean matching procedure of French and Jones (2011), following David et al. (1986) and Little (1988). The method combines predicted values with residuals drawn from individuals with similar characteristics.

Let y be observed for individuals in $\mathcal{O}$. We estimate

$$y_i = x_i'\lambda + e_i$$

by least squares on $\mathcal{O}$, obtaining fitted values $\hat{y}_i = x_i'\hat{\lambda}$ for every individual and residuals $\hat{e}_j = y_j - \hat{y}_j$ for observed individuals. We sort all individuals into deciles of fitted values. For each individual $i \notin \mathcal{O}$, we randomly select a donor $j \in \mathcal{O}$ from the same decile and impute

$$y_i^{\text{imp}} = \hat{y}_i + \hat{e}_j. \tag{C.6}$$

Unlike conditional-mean imputation, this procedure preserves residual variation from observed individuals with similar predicted values. We apply it to average earnings for individuals without a pension reading, using x_i, and to wage residuals for nonworkers, augmenting x_i with imputed average earnings at entry.

C.6 Sampling

The initial distribution is the empirical joint distribution of the state vectors

$$(t_{0,i},\ a_{i,t_0},\ ae_{i,t_0},\ H_{i,t_0},\ u_{i,t_0},\ \mathbf{1}\{N_{i,t_0} > 0\})$$

among individuals with complete initialized states. We sample complete state vectors with replacement and simulate each individual from his entry-age group. Drawing health and wage shocks from their posterior distributions preserves uncertainty about latent states, while stochastic imputations of average earnings and wage residuals retain variation that conditional-mean imputations would eliminate. Sampling the complete vectors jointly preserves the dependence across initial states captured by the data and the maintained imputation procedures.

The estimation targets use the same cohort across its observed survey waves. In simulated data, health-conditioned moments are constructed to match the corresponding empirical estimator, as described in Section 3.3. Hours and participation are conditioned on true latent health in the model, while their empirical counterparts are recovered through regressions on posterior probabilities. Wealth percentiles are instead classified by the modal posterior health state in both the data and simulated samples, with the latter obtained from simulated health indicators.

D Numerical Solution

We solve the household problem by backward induction. At each age and state, we maximize the Bellman equation over savings and hours worked, with consumption

determined by the budget constraint:

$$V_t(H, a, ae, u) = \max_{a', N} \Big\{ U\big(c,\ \bar{L} - \phi_P \mathbf{1}_{\{N>0\}} - N - \phi_H \mathbf{1}_{\{H=B\}}\big) + \varpi \\ + [1 - s_t(H)]\, b(a') + \beta\, s_t(H)\, \mathbb{E}\big[V_{t+1}(\cdot, a', ae', \cdot)\big] \Big\} \tag{D.1}$$

subject to

$$c + a' = y_t\big(ra + W_t N + \mathbf{1}_{\{t \geqslant R_a\}}(pbb_t + privben_t)\big) + ys(t, H) + tr_t + a, \qquad a' \geqslant 0.$$

The expectation integrates over next-period health and wage shocks, conditional on survival. Average earnings evolve according to their law of motion, while mortality depends on age and health.

For the numerical solution, we use the normalized flow-utility and bequest functions

$$\frac{(C^\nu L^{1-\nu})^{1-\gamma} - 1}{1-\gamma}, \qquad \theta_B \frac{(\kappa_B + a)^{(1-\gamma)\nu} - 1}{1-\gamma},$$

which have well-defined limits as $\gamma \to 1$. These normalizations leave optimal decisions unchanged because mortality is exogenous and the additive constants do not depend on household choices. They matter, however, for welfare, since an additive constant in flow utility changes the value of being alive relative to death. Welfare calculations therefore use the corresponding unnormalized utility functions and the value-of-life parameter ϖ, as described in Section 5.1.

Discretization and continuation values. We discretize assets and average earnings using power-spaced grids, with more points near zero to capture the effects of borrowing constraints and nonlinearities in household decisions. The persistent wage shock is discretized using an age-varying Rouwenhorst approximation that accommodates changes in its variance over the life cycle. Health is discrete.

At each age and grid point, we compute expected continuation values by integrating over health and wage-shock transitions. Before retirement, we use bilinear interpolation over assets and average earnings to evaluate continuation values at the states implied by household choices. After retirement, average earnings are fixed, so interpolation is linear in assets.

Optimization over household choices. The Bellman objective need not be globally concave. In particular, the fixed time cost of labor-force participation creates a discontinuity at zero hours, while the tax schedule introduces kinks in disposable income. We therefore combine a structured search over feasible choices with local optimization.

At each state, we construct candidate choices directly in levels of consumption and hours worked. The hours grid includes zero, the hours corresponding to taxable-income thresholds, and fractions of the maximum feasible hours. The consumption grid includes half the consumption floor, the floor plus fractions of resources available above it, and consumption of all available resources. We evaluate the Bellman objective at each feasible combination, using interpolated continuation values.

We then refine the best candidates using three complementary procedures. First, we optimize consumption holding hours at zero, explicitly solving the nonparticipation problem. Second, we apply the Nelder–Mead algorithm to consumption and hours, starting from the best interior candidates. Third, we optimize hours subject to a binding borrowing constraint, $a' = 0$.

For each state, we retain the feasible choice delivering the highest value across the initial candidate search and the three refinements. This procedure considers non-participation, interior solutions, and borrowing-constrained choices separately, while accounting for nonlinearities induced by taxes and the fixed participation cost. By searching from multiple candidates and explicitly evaluating economically relevant boundaries, it reduces reliance on concavity or convergence from a single starting point.

E Structural Estimation

This appendix provides additional details on the indirect-inference procedure used to estimate the structural parameters. We describe the estimation targets and weighting matrix, the numerical minimization of the objective, the sources of parameter identification, and the construction of the empirical moments. We also report additional results on the fit of the wage profiles.

E.1 Weighting Matrix and Objective

Let ψ_{jn}, for $j = 1, \ldots, J$, denote the empirical targets, and let $\psi_j(\theta)$ denote their simulated counterparts at structural parameter vector θ. We use $J = 204$ targets: for each of the 20 two-year age groups from ages 52–53 to 90–91, the 25th, 50th, and 75th percentiles of wealth, mean hours worked, and the labor-force participation rate, each by health status. These yield ten targets per age group, supplemented by the four coefficients of the wage profile.

We weight each target by the inverse of its estimated sampling variance. The weighting matrix is diagonal:

$$W_n = \operatorname{diag}\left(\frac{1}{\widehat{\operatorname{Var}}(\psi_{1n})}, \ldots, \frac{1}{\widehat{\operatorname{Var}}(\psi_{Jn})}\right).$$

The variances of the age- and health-specific targets are estimated by resampling individuals, using 500 bootstrap draws per cell. The variances of the wage-profile coefficients are obtained from the bootstrap procedure described in Appendix B.4. We exclude targets with insufficient observations or zero estimated bootstrap variance, leaving 182 of the 204 targets active.

The indirect-inference objective is

$$Q_n(\theta) = \frac{1}{J_a} \sum_{j \in \mathcal{A}} \frac{\left[\psi_j(\theta) - \psi_{jn}\right]^2}{\widehat{\operatorname{Var}}(\psi_{jn})}, \tag{E.1}$$

where $\mathcal{A}$ denotes the set of active targets and $J_a = 182$. This objective equals that of (21) divided by J_a, which leaves the minimizer unchanged. The objective measures the average squared discrepancy between simulated and empirical moments, expressed in units of their estimated sampling standard errors. Its square root is therefore the root-mean-square standardized deviation across targets.

We simulate 40 replicas of each individual in the initial-state distribution, holding the realizations of all exogenous shocks fixed across parameter vectors. This common-random-numbers approach eliminates variation in the objective arising solely from redrawing shocks during optimization. Nevertheless, the finite simulated sample and discrete household decisions generate local irregularities in the objective, which also exhibits multiple local minima. These features motivate the optimization procedure described next.

E.2 Minimization of the Indirect-Inference Objective

We minimize (E.1) using the TikTak global-search procedure of Arnoud et al. (2019), followed by local refinement and comparisons across estimation runs. The procedure combines a broad initial exploration of the parameter space with searches concentrated around promising candidates.

The parameter space is a bounded box:

$$\begin{aligned}
\beta &\in [0.8,\ 1.2],\\
\nu &\in [0.4,\ 0.98],\\
\phi_P &\in [10,\ 5000],\\
\phi_H &\in [1,\ 3500],\\
\theta_B &\in [0,\ 100],\\
\kappa_B &\in [5000,\ 1500000],\\
C_{\min} &\in [100,\ 80000].
\end{aligned}$$

The time costs ϕ_P and ϕ_H are measured in hours per two-year period, while $C_{\min}$ is expressed in pounds sterling per two-year period and κ_B in pounds sterling. Each of the four wage-profile coefficients is restricted to lie within ten estimated standard errors of its empirical counterpart, with the additional restriction $a_H \geqslant 0$. No estimated parameter lies at a boundary. The algorithm has four stages:

1. *Global search.* We evaluate the objective at 32,768 points generated by a Sobol sequence over the parameter space, discarding infeasible points at which the household problem cannot be solved, such as parameter combinations for which the time costs exceed the available time endowment. The evaluations are parallelized, and we retain the 364 points with the lowest objective values as starting candidates.

2. *Local search.* We process the 364 candidates in ascending order of their objective values. For candidate i, the starting point is a convex combination of the candidate $\theta^{(i)}$ and the current best solution $\hat{\theta}$:

   $$\theta_0^{(i)} = (1 - w_i)\,\theta^{(i)} + w_i\,\hat{\theta},$$

   where

   $$w_i = \min\left\{0.995,\ \max\left\{\sqrt{\frac{i}{364}},\ 0.05\right\}\right\}.$$

   As the search progresses, starting points place increasing weight on the incumbent, concentrating subsequent searches around the most promising region. Each candidate is refined using Nelder–Mead, with a budget of objective evaluations that grows from 100 for the first candidate to 200 for the last. We update the incumbent whenever a search improves the objective and retain the endpoint of every local search.

3. *Polishing.* Because the local searches may stop before converging, we refine the 52 best distinct candidates obtained from the incumbent, the local-search endpoints, and estimates from earlier runs. The latter include estimates from the alternative treatment of health measurement error, the alternative set of health indicators, and the decomposition exercises in Section 5.2. We restrict these candidates to the admissible parameter space and evaluate them using the objective of the current specification. From each candidate, we run both Nelder–Mead and BOBYQA (Powell, 2009), allowing 1,500 objective evaluations for each search. The candidate with the lowest objective after these refinements is retained as the estimate.

4. *Cross-run verification.* Once all specifications have been estimated, we evaluate their final parameter vectors under the objective of each specification. If another run provides a lower value, we repeat the polishing stage using the combined set of candidates.

The cross-run refinement matters because closely related specifications can converge to different local minima; in the final estimation it improved the estimate in one of the eight runs. The complete procedure takes approximately four hours per run using 104 workers. Although this multistart procedure does not guarantee attainment of the global minimum, it substantially reduces dependence on a single initialization and exploits information from closely related specifications.

Treatment of employment selection. The four wage-profile coefficients are estimated by matching the auxiliary regression applied to observed and simulated workers. This approach follows the logic of French (2005), who proposes iteratively adjusting structural wage coefficients until the profiles estimated among simulated workers coincide with those estimated in the data.

That iterative procedure need not converge, a difficulty we also encountered in implementation. Instead, we include the four wage-profile coefficients directly among the indirect-inference targets. The same auxiliary estimator is applied to the observed and simulated samples, so the comparison incorporates employment selection generated by the structural model. Unlike exact iterative matching, the indirect-inference objective permits discrepancies in individual wage coefficients to be traded off against the remaining targets according to their estimated sampling variances.

This distinction is important because observed wages are available only for workers, whose health and wage shocks influence participation decisions. Consequently, the wage-profile coefficients recovered from employed individuals need not coincide with the structural coefficients governing potential wages. The indirect-inference procedure accounts for this selection when estimating the structural wage-profile parameters.

E.3 Parameter Identification

The structural parameters are identified jointly through their effects on the age- and health-specific profiles of wealth, labor supply, and observed wages. Each parameter affects several outcomes, so the mechanisms described below are the main sources of identifying variation rather than one-to-one moment restrictions.

The biennial discount factor β primarily governs saving behavior and is disciplined by asset accumulation at younger ages. The consumption weight ν determines the trade-off between consumption and leisure and is identified mainly by hours worked. The fixed participation cost ϕ_P governs the extensive margin of labor supply and is primarily disciplined by participation rates. The time cost of bad health ϕ_H affects labor supply conditional on health, particularly the differences in hours worked between individuals in good and bad health. The bequest parameters θ_B and κ_B determine the strength and curvature of the bequest motive and are primarily disciplined by wealth holdings toward the end of life. The consumption floor C_{min} affects saving incentives among households with few resources and is identified mainly by the lower tail of the wealth distribution. The four structural wage coefficients (a_0, a_1, a_2, a_H) are disciplined by the wage profiles estimated among workers in the observed and simulated samples. Because participation depends on health and productivity, the coefficients recovered from observed wages generally differ from those governing potential wages. Applying the same auxiliary estimator to observed and simulated workers allows the model to account for this selection when matching wage profiles.

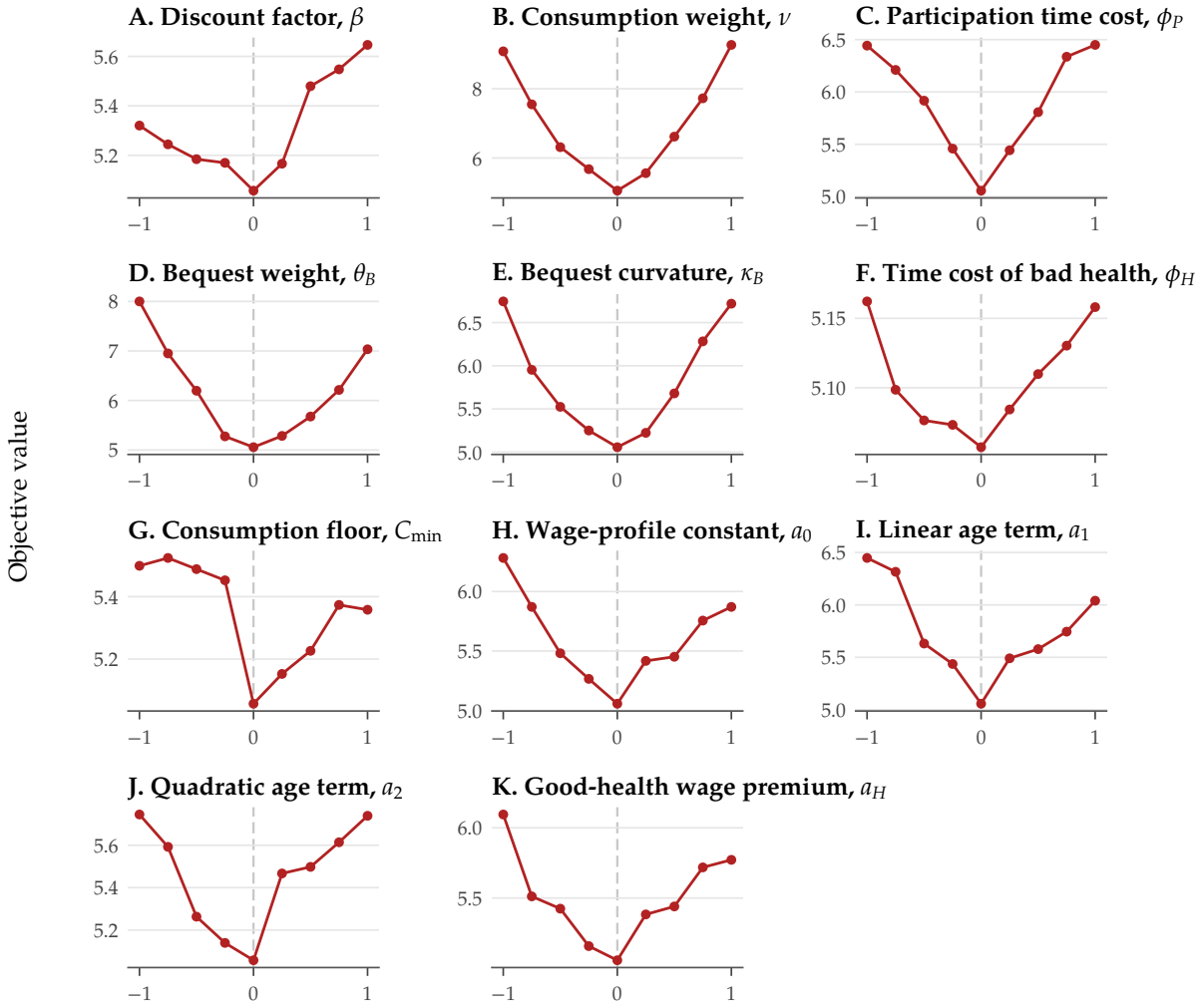
Identification is particularly intertwined for ϕ_H and a_H. A higher time cost of bad health reduces labor supply among unhealthy individuals and changes the composition of workers whose wages are observed. Consequently, ϕ_H affects not only labor-supply moments but also the health wage premium. In the estimated model, its effect on that premium runs opposite to the effect of a_H, generating a negative relationship between the two estimates and weakening their separate determination by the wage moments. Similarly, the linear and quadratic age coefficients of the wage profile can offset each other over the observed age range, making them difficult to distinguish individually.

We examine the local sensitivity of the estimates to the targets following Andrews et al. (2017), evaluated at estimates from an earlier round of optimization over a narrower parameter space, which differ by less than 5 percent from the final estimates in Table 2. The resulting approximate standard errors range from 0.6 to 2.9 percent of the estimates for β, ν, C_{min}, and a_0, and are approximately 10 percent for ϕ_P and ϕ_H. The bequest parameters are less precisely determined, with relative standard errors of 33 percent for θ_B and 13 percent for κ_B, and the linear and quadratic wage-age coefficients are difficult to identify separately. These approximations ignore simulation

uncertainty and the sampling uncertainty of the first-stage estimates; we use them to characterize the local behavior of the estimator and to set the parameter ranges in the figures below, not as confidence intervals.

Local sensitivity of the objective. Figure E.1 plots the indirect-inference objective along each of the eleven structural parameters in the specification accounting for health measurement error. For each parameter, we evaluate the objective at nine points spanning one approximate standard error on either side of its estimate, holding the remaining parameters and simulated shocks fixed. The horizontal range therefore depends on the parameter's approximate standard error.

Figure E.1: Local sensitivity of the objective, accounting for measurement error.

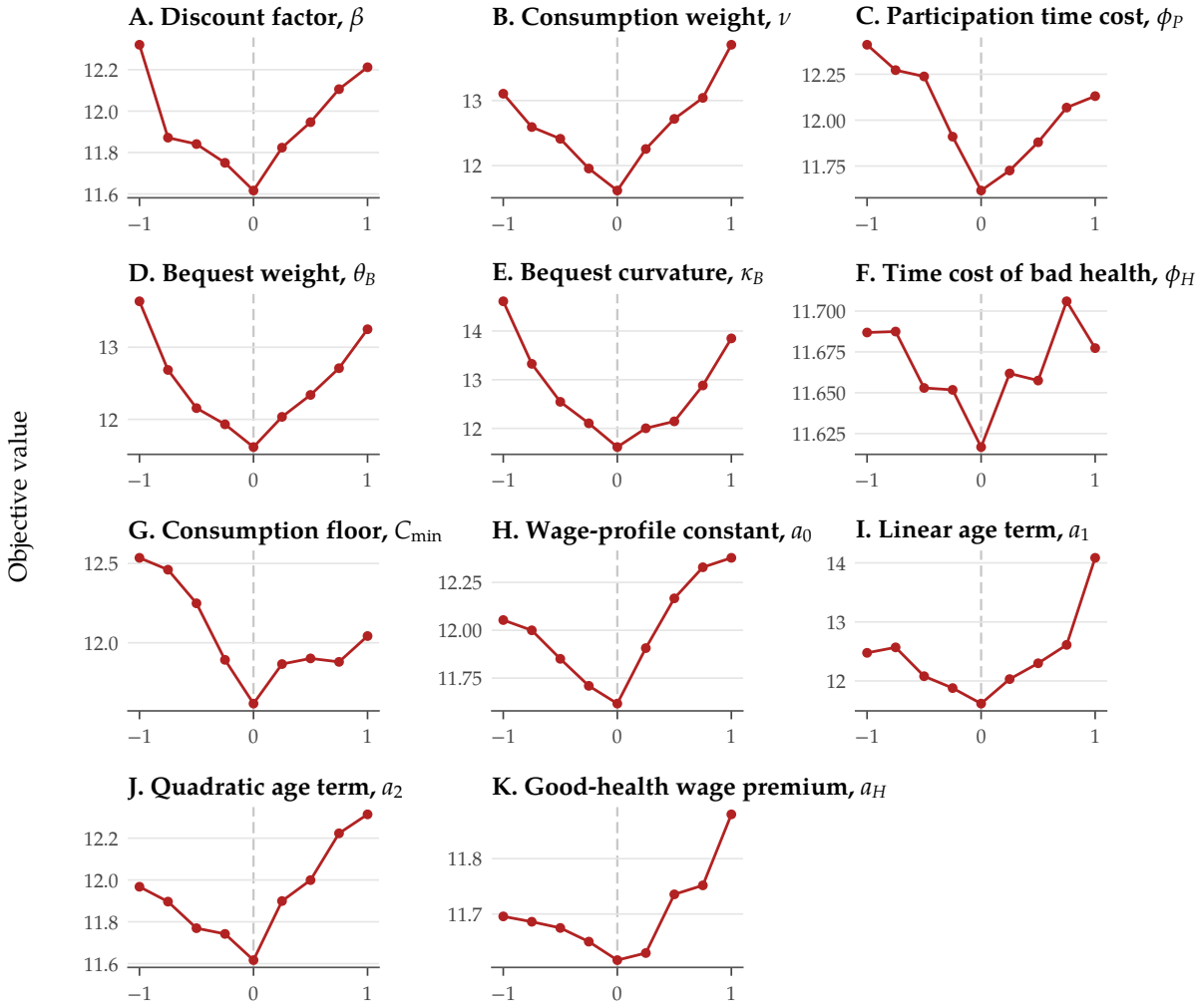


Notes. The objective is evaluated at nine points within one approximate standard error on either side of each parameter estimate, holding the other parameters and simulated shocks fixed.

The objective has a local minimum at the estimate along each of the plotted directions, but its sensitivity varies substantially: over the displayed range it rises by approximately four units for ν, two to three units for θ_B and κ_B, one unit for ϕ_P, a_0, a_1, and a_H, and only 0.1 for ϕ_H.

These profiles do not establish joint identification: a sharply curved one-dimensional profile can conceal directions along which combinations of parameters move with little effect on the objective, as for ϕ_H and a_H and for the linear and quadratic wage-age coefficients. The small irregularities reflect the finite simulated sample. Figure E.2 repeats the exercise for the specification that ignores health measurement error, using that specification's approximate standard errors.

Figure E.2: Local sensitivity of the objective, ignoring measurement error.



Notes. The construction is identical to Figure E.1, using the parameter estimates and approximate standard errors from the specification that treats self-reported health as the true state.

E.4 Construction of the Estimation Targets

Wealth profiles. Following De Nardi, Pashchenko and Porapakkarm (2025), we adjust wealth for differences in household size and survey-wave effects that are not explicitly modeled. We estimate

$$w_{i,w} = \mu_{t,h} + \sum_{k=1}^{3} d_k^h n_{i,w}^k + \delta_w + e_{i,w}, \tag{E.2}$$

where $w_{i,w}$ denotes total wealth of individual i in survey wave w, $\mu_{t,h}$ are age-group-by-health effects, $n_{i,w}$ denotes household size, and δ_w are wave effects. The coefficients d_k^h allow the relationship between household size and wealth to differ by health status.

We construct adjusted wealth as

$$\widetilde{w}_{i,w} = \widehat{\mu}_{t,h} + \sum_{k=1}^{3} \widehat{d}_k^h \bar{n}_t^k + \bar{\delta} + \widehat{e}_{i,w}, \tag{E.3}$$

where $\bar{n}_t$ is average household size at age t, and $\bar{\delta}$ is the average wave effect. The adjustment evaluates the fitted household-size component at the average household size for each age and replaces the wave effect with its average, retaining the individual's regression residual. It therefore removes the variation attributed to household size and survey timing while preserving residual differences across individuals.

The wealth targets are the 25th, 50th, and 75th percentiles of adjusted wealth by age and health. When measurement error is ignored, health groups are defined using self-reported health. When measurement error is accounted for, individuals are classified according to their most likely latent-health state given the observed indicators, as described in Section 3.3. Initial wealth is adjusted consistently with the empirical wealth targets, so that the simulated distribution starts from the same measurement basis.

Labor-supply profiles. We target mean hours worked and labor-force participation rates by age and health, using the samples and procedures described in Section 3.3.

The construction of comparable empirical and simulated moments depends on how health is treated. When accounting for measurement error, the empirical hours and participation profiles are recovered through regressions on posterior probabilities of latent health. These regressions identify health-specific conditional means under the exclusion restrictions described in Appendix B.3. Their simulated counterparts are computed conditional on the model's true latent-health state.

The wealth targets use a different classification procedure: observations are assigned to health groups according to their modal posterior state. In simulated data, we generate health indicators from the estimated measurement system and apply the corresponding posterior classification. Thus, wealth and labor-supply targets use different mappings from observed indicators to latent health, consistent with their respective empirical estimators. In the specification that ignores measurement error, health-conditioned empirical targets are constructed using self-reported health, and the corresponding simulated targets follow the same treatment of health.

Wage profiles. We estimate the wage-profile targets using male core members born between 1940 and 1970 and observed at ages 77 or younger. The broader cohort range increases precision because many individuals in the structural estimation cohort are already retired when first interviewed. The age restriction excludes later age groups with relatively few usable wage observations.

The structural wage equation is

$$\log W_{it} = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}\{H_{it} = G\} + u_{it}. \tag{E.4}$$

Observed wages are available only for employed individuals and contain measurement error. We therefore construct four auxiliary wage-profile coefficients from observed workers, using the procedure detailed in Appendix B.4. Because employment depends on health and wage shocks, these auxiliary coefficients need not equal the corresponding structural coefficients.

When health measurement error is ignored, the health regressor is the indicator for good self-reported health. We estimate the age and health coefficients using all pairs of working observations for each individual, weighted by the inverse of the number of working periods. As shown in Appendix B.4.1, this estimator is algebraically equivalent to the conventional within estimator when regressors are defined at the individual-period level.

When measurement error is accounted for, we instead use pair-specific posterior probabilities of good latent health. For each observation pair, the posterior incorporates the observed indicators and relevant survival information at both dates. Because the posterior at one date can depend on the information observed at the other, this estimator is not generally equivalent to the conventional within estimator. Its construction and identifying assumptions are described in Appendix B.4.1.

The auxiliary wage-level coefficient is obtained from the mean wage residual among employed individuals. Together with the coefficients on age, age squared, and good health, it gives the four wage-profile targets.

We compute the corresponding auxiliary coefficients in simulated data using the same regression design and the treatment of health appropriate to each specification. The comparison is therefore between coefficients estimated among employed individuals in both samples, rather than between the empirical auxiliary coefficients and the structural wage parameters themselves.

This distinction is central to the indirect-inference strategy. The structural model determines which individuals work and therefore whose wages enter the simulated auxiliary regression. Matching its coefficients to those obtained from the data allows the estimation to account for employment selection without requiring potential wages to be observed for nonworkers.

E.5 Additional Fit Results

Table E.1 compares the auxiliary wage-profile coefficients estimated in ELSA with those obtained from simulated data. Results are reported for both specifications, ignoring and accounting for health measurement error.

Table E.1: Fit of estimated wage profiles

Parameter	Ignoring ME			Accounting for ME		
	Data	**Model**	t	**Data**	**Model**	t
a_0	2.597	2.710	2.79	2.553	2.654	1.58
a_1	−0.000	−0.070	−5.65	−0.002	−0.046	−3.65
a_2	−0.001	0.001	1.87	−0.000	0.001	1.75
a_H	0.035	0.080	2.23	0.083	0.050	−0.60

Notes. The reported coefficients are estimates from the auxiliary wage-profile regressions, rather than structural wage coefficients. The statistic t is the difference between the simulated and empirical coefficient, divided by the estimated standard error of the empirical target, obtained by resampling individuals.

Accounting for health measurement error improves the fit of the auxiliary health wage premium. The model generates a coefficient of 0.050, compared with 0.083 in the data, a discrepancy of −0.60 standard errors. When measurement error is ignored, the corresponding coefficients are 0.080 in the model and 0.035 in the data, a discrepancy of 2.23 standard errors. Thus, the specification accounting for measurement error more closely reproduces the health wage premium estimated among workers.

Both specifications fit the age profile less well. In particular, the simulated linear age coefficient is more negative than its empirical counterpart, with discrepancies of −5.65 and −3.65 standard errors in the specifications ignoring and accounting for measurement error, respectively. The quadratic coefficient also differs from its empirical counterpart, although by less than two standard errors in both specifications.

Because the linear and quadratic age coefficients jointly determine the wage-age profile, their individual discrepancies do not translate directly into discrepancies in predicted wages at particular ages. Nevertheless, the results indicate that neither specification matches all aspects of the empirical wage-age relationship. This remains a limitation of the estimated model.

The model accounting for measurement error fits the auxiliary health wage premium more closely despite these discrepancies in the age coefficients. This distinction matters for the quantitative analysis because the health wage premium, together with selection into employment and the persistence of latent health, determines the effects of bad health on earnings.

F Health Misclassification and Attenuation Bias

This appendix illustrates how measurement error in health can attenuate the effects of health recovered from structural estimation. We develop a static labor-supply model in which individuals differ in their labor-supply elasticities and health affects the disutility of work. We show that treating noisy health classifications as true states attenuates the estimated labor-supply cost of bad health, even when multiple indicators improve classification accuracy. We then show how incorporating posterior health probabilities into estimation can eliminate this source of attenuation.

Labor supply and the true effect of health. Consider a worker i whose health state is $m_i \in \{G, B\}$, denoting good and bad health. Health affects the disutility of work through $\phi_m > 0$, with $\phi_B > \phi_G$. Individuals also differ in their Frisch elasticities, $\epsilon_i > 0$. Given a common wage $w > 0$, worker i solves

$$\begin{aligned} \max_{c,h \geqslant 0} \quad & c - \phi_m \frac{h^{1+1/\epsilon_i}}{1+1/\epsilon_i} \\ \text{s.t.} \quad & c \leqslant wh. \end{aligned}$$

Consumption equals labor income at the optimum, $c = wh$. The first-order condition for hours is $w = \phi_m h^{1/\epsilon_i}$, which gives

$$h_i(m) = \left(\frac{w}{\phi_m}\right)^{\epsilon_i},$$

or in logs,

$$\log h_i(m) = \epsilon_i(\log w - \log \phi_m).$$

Suppose the distribution of Frisch elasticities ϵ_i is independent of health, with mean $\bar{\epsilon} = \mathbb{E}[\epsilon_i]$. The mean of log hours conditional on health is then

$$\mu_m \equiv \mathbb{E}[\log h_i \mid m_i = m] = \bar{\epsilon}(\log w - \log \phi_m).$$

Thus, given the wage and the distribution of labor-supply elasticities, the two conditional means identify ϕ_G and ϕ_B. The average effect of moving from bad to good health on log hours is

$$\Delta \equiv \mu_G - \mu_B = \bar{\epsilon} \log\left(\frac{\phi_B}{\phi_G}\right) > 0. \tag{F.1}$$

Since $\phi_B > \phi_G$, individuals work fewer hours in bad health. Moreover, because labor-supply elasticities are independent of health, Δ also equals the average counterfactual increase in log hours from improving an individual's health from bad to good, holding wages and individual elasticities fixed.

Attenuation from a noisy health measure. Suppose the econometrician does not observe true health m_i, but instead observes a noisy indicator m_i^*. The indicator reports the true state with probability $1-p$ and a random state with probability p:

$$m_i^* = \begin{cases} m_i, & \text{with probability } 1-p, \\ e_i, & \text{with probability } p, \end{cases}$$

where e_i equals G or B with probability one-half each. Reporting errors are independent of labor-supply elasticities conditional on true health. The resulting misclassification probabilities are symmetric:

$$\mathbb{P}(m_i^* = B \mid m_i = G) = \mathbb{P}(m_i^* = G \mid m_i = B) = \frac{p}{2}.$$

To obtain a simple closed-form attenuation factor, suppose initially that the two health states are equally prevalent, $\mathbb{P}(m_i = G) = \mathbb{P}(m_i = B) = 1/2$. Bayes' rule then implies

$$\mathbb{P}(m_i = G \mid m_i^* = G) = 1 - \frac{p}{2}, \qquad \mathbb{P}(m_i = G \mid m_i^* = B) = \frac{p}{2}.$$

Consequently, mean log hours conditional on reported health are

$$\begin{aligned} \mathbb{E}[\log h_i \mid m_i^* = G] &= \left(1 - \frac{p}{2}\right)\mu_G + \frac{p}{2}\mu_B, \\ \mathbb{E}[\log h_i \mid m_i^* = B] &= \frac{p}{2}\mu_G + \left(1 - \frac{p}{2}\right)\mu_B. \end{aligned}$$

Subtracting yields

$$\underbrace{\mathbb{E}[\log h_i \mid m_i^* = G] - \mathbb{E}[\log h_i \mid m_i^* = B]}_{\Delta^*} = (1-p)\underbrace{(\mu_G - \mu_B)}_{\Delta}. \tag{F.2}$$

Thus, the observed difference in mean log hours is attenuated by a factor $1-p$.

Suppose the econometrician ignores measurement error and estimates ϕ_G and ϕ_B by matching the conditional means of log hours using reported health as if it were true health. The estimated difference in health-related work disutility satisfies

$$\log \hat{\phi}_B - \log \hat{\phi}_G = (1-p)(\log \phi_B - \log \phi_G).$$

The model therefore interprets the smaller observed hours gap as evidence of a smaller difference in the disutility of work across health states. Consequently, the counterfactual effect implied by the estimated parameters is

$$\hat{\Delta} = (1-p)\Delta < \Delta. \tag{F.3}$$

The intuition is straightforward. As health reports become noisier, the groups classified as healthy and unhealthy contain increasingly similar mixtures of true health states. Their observed labor supply therefore converges, and a structural model that treats reported health as accurate interprets this convergence as evidence that health has a smaller effect on labor supply.

The exact attenuation factor depends on the prevalence of the two health states. More generally, let $\pi = \mathbb{P}(m_i = G)$ and $r = \mathbb{P}(m_i^* = G) = p/2 + (1-p)\pi$. The observed hours gap is

$$\Delta^* = \frac{(1-p)\pi(1-\pi)}{r(1-r)}\Delta. \tag{F.4}$$

For $0 < \pi < 1$ and $0 < p < 1$, the multiplier lies strictly between zero and one. Thus, measurement error also attenuates the estimated health effect when the prevalence of good and bad health differs, although the attenuation factor generally depends on both reporting accuracy and health prevalence. Equation (F.2) obtains when $\pi = 1/2$.

Combining multiple health indicators. Suppose now that the econometrician observes three noisy indicators, $m_i^{*1}, m_i^{*2}, m_i^{*3}$, rather than one. Each indicator follows the same reporting process as above, with independent reporting errors conditional on true health:

$$m_i^{*j} = \begin{cases} m_i, & \text{with probability } 1-p, \\ e_i^j, & \text{with probability } p, \end{cases} \qquad j = 1,2,3,$$

where each e_i^j independently equals G or B with probability one-half.

We interpret these indicators as responses to three survey questions measuring the same underlying health state. Their conditional independence allows the joint responses to provide more information about latent health than any single indicator.

An intuitive way to use this information is to classify individuals according to their most likely health state. Let

$$\hat{m}_i = \arg\max_{m \in \{G,B\}} \mathbb{P}(m_i = m \mid m_i^{*1}, m_i^{*2}, m_i^{*3})$$

denote the modal posterior classification.

Under the balanced health-state prior, $\mathbb{P}(m_i = G) = 1/2$, and the identical symmetric reporting processes assumed above, this classification follows a majority rule: individuals are classified according to the state reported by at least two indicators. For example, an individual reporting (G, G, B) is classified as being in good health.

Combining the indicators improves classification accuracy. Let $q = p/2$ denote the misclassification probability of each indicator. The majority classification is incorrect only when at least two indicators misclassify health, which occurs with probability

$$\mathbb{P}(\hat{m}_i \neq m_i \mid m_i) = 3q^2(1-q) + q^3 = 3q^2 - 2q^3.$$

The majority classification can be represented by the same symmetric reporting process as an individual indicator, but with an effective noise parameter $\hat{p}$ satisfying

$$\frac{\hat{p}}{2} = 3\left(\frac{p}{2}\right)^2 - 2\left(\frac{p}{2}\right)^3.$$

Hence,

$$\hat{p} = \frac{p^2(3-p)}{2} < p, \qquad 0 < p < 1. \tag{F.5}$$

Combining three independent indicators therefore reduces misclassification relative to relying on any one of them.

Nevertheless, the modal classification remains imperfect. If the econometrician treats $\hat{m}_i$ as true health and estimates the structural model by matching hours conditional on this classification, the resulting difference in mean log hours is

$$\begin{aligned} \hat{\Delta}_{\text{modal}} &\equiv \mathbb{E}[\log h_i \mid \hat{m}_i = G] - \mathbb{E}[\log h_i \mid \hat{m}_i = B] \\ &= (1-\hat{p})\Delta. \end{aligned} \tag{F.6}$$

Because $0 < \hat{p} < p$,

$$(1-p)\Delta < (1-\hat{p})\Delta < \Delta. \tag{F.7}$$

Combining indicators thus reduces attenuation but, under modal classification, does not eliminate it.

Accounting for the measurement process. The preceding attenuation arises from treating noisy health classifications as true states, rather than from latent health itself. To see this, suppose that the measurement process is known or consistently estimated. Let $Y_i = (m_i^{*1}, m_i^{*2}, m_i^{*3})$ denote the vector of observed health indicators, and define the posterior probability of good health as $q(Y_i) = \mathbb{P}(m_i = G \mid Y_i)$. Under the maintained assumption that reporting errors contain no additional information about labor supply conditional on true health, $\mathbb{E}[\log h_i \mid Y_i, m_i] = \mathbb{E}[\log h_i \mid m_i]$. The conditional expectation of log hours given the observed indicators is therefore

$$\mathbb{E}[\log h_i \mid Y_i] = q(Y_i)\mu_G + [1 - q(Y_i)]\mu_B = \mu_B + \Delta\, q(Y_i). \tag{F.8}$$

Equation (F.8) shows that accounting for the measurement process eliminates attenuation under the maintained assumptions. Regressing log hours on the posterior probability of good health, with an intercept, identifies the true difference in health-specific means, $\Delta = \mu_G - \mu_B$, provided the posterior probabilities are correctly specified and exhibit sufficient variation. Unlike modal classification, this approach preserves information about uncertainty in latent health rather than assigning each individual to a single state. This result motivates our treatment of health in structural estimation. Multiple noisy indicators allow us to identify latent-health probabilities, while the exclusion restriction ensures that these probabilities can be used to recover health-specific economic moments. The key distinction is not simply that more indicators improve health classification, but that explicitly modeling measurement error prevents residual classification mistakes from being interpreted as smaller economic effects of health.

Implications. The stylized model illustrates why ignoring health measurement error can understate the economic consequences of bad health. Combining multiple indicators reduces this bias, but modal classification generally leaves residual attenuation. Incorporating the estimated measurement process instead permits recovery of the underlying health-specific moments under the maintained identification assumptions. In the quantitative life-cycle model, the mapping from health to economic outcomes is more complex, and the direction of bias in a particular counterfactual need not follow the simple attenuation formula derived here. Nevertheless, the same mechanism applies: treating imperfect health classifications as true states can distort structural inference about the economic costs of bad health.

G Additional Results and Robustness

This appendix provides additional details on the welfare calculations and decomposition exercises in Section 5. We also examine whether accounting for health measurement error changes the predicted effects of pension policies and report the results of re-estimating the model using alternative health indicators.

G.1 Welfare Calibration

As described in Section 5.1, the welfare cost of bad health incorporates the value of additional years of life. We calibrate the additive value of being alive, ϖ, to match an average value of a statistical life (VSL) of \$2 million in 2013 dollars among individuals aged 50–63. Converted to 2013 pounds with the World Bank purchasing-power-parity factor for the United Kingdom (0.695248), the target is approximately £1.39 million; we also consider a \$6 million target, approximately £4.17 million.

With $\gamma = 3$, the unnormalized flow-utility function is negative, so the model does not necessarily assign a positive value to additional survival. Following Section 5.1, we introduce ϖ to capture the value of being alive beyond consumption and leisure. We calibrate ϖ separately for each specification to match the VSL target. Since mortality is exogenous and ϖ enters utility additively, this calibration does not affect household decisions.

At each age and state, the VSL is the marginal value of an increase in the probability of surviving to the next period, expressed in monetary units using the marginal value of assets:

$$\mathrm{VSL}_t = \frac{\partial V_t / \partial s_t}{\partial V_t / \partial a_t}.$$

We exclude from the calibration person-periods in which the consumption floor binds: approximately 5 percent of working-age observations in the latent-health specification and 8 percent in the specification ignoring measurement error. For households receiving transfers, additional resources can be offset one for one by reductions in transfers, making the value function locally insensitive to wealth when the consumption floor and borrowing constraint bind. In this region, the marginal value of assets can be zero, rendering the VSL ratio undefined.

To calculate consumption-equivalent welfare costs, we find the proportional adjustment to consumption in every period of the always-healthy counterfactual that equates realized lifetime utility across the two scenarios, holding bequests fixed. Welfare calculations use the unnormalized utility functions and the calibrated ϖ, as in Section 5.1.

G.2 Additional Decomposition Results

Table G.1 reports the full decomposition underlying Figure 5. Starting from the specification that ignores measurement error, we replace the estimated health and mortality process, the labor-supply targets, and the wealth targets with their counterparts from the specification accounting for measurement error, separately and in combination. In each exercise, we re-estimate the structural parameters while leaving the initial distribution, wage process, and spousal-income process unchanged.

For an outcome y, the reported share is

$$S_y^{(k)} = \frac{C_y^{(k)} - C_y^{\text{NoME}}}{C_y^{\text{ME}} - C_y^{\text{NoME}}},$$

where $C_y^{(k)}$ is the estimated cost under replacement k, and C_y^{NoME} and C_y^{ME} are the costs in the two baseline specifications; the table reports $100\,S_y^{(k)}$. A value of 100 indicates that the replacement closes the entire difference, while a negative value indicates that it widens the gap. Because the inputs interact through structural re-estimation, the entries need not add up to 100 and should not be interpreted as independent contributions.

Table G.1: Percentage of the Difference in the Costs of Bad Health Closed by Each Input

Outcome	Process for health (1)	Targets for labor (2)	Targets for assets (3)	Health + labor (4)	Labor + assets (5)	All three (6)
All individuals						
Earnings, up to 64–65	−10	101	0	41	100	40
Hours worked, up to 64–65	−4	104	0	42	95	41
Years worked	15	74	0	54	70	54
Consumption	149	−33	0	141	−67	141
Monetary cost, remaining life	48	32	0	63	32	63
Monetary cost, up to 64–65	6	57	0	35	58	35
Years of life gained	89	0	0	89	0	89
Welfare cost, VSL £1.39m	70	54	0	79	52	79
Welfare cost, VSL £4.17m	79	36	0	83	36	83
Initially unhealthy						
Earnings, up to 64–65	49	34	0	79	33	79
Hours worked, up to 64–65	49	41	0	79	37	78
Years worked	60	26	0	83	24	84
Consumption	52	33	0	74	36	73
Assets	57	25	0	69	23	70
Monetary cost, remaining life	77	15	0	89	15	89
Monetary cost, up to 64–65	46	32	0	70	32	70
Years of life gained	102	0	0	102	0	102
Welfare cost, VSL £1.39m	85	34	0	94	31	94
Welfare cost, VSL £4.17m	92	21	0	96	20	96

Notes. Columns re-estimate the model ignoring measurement error with the indicated inputs replaced by those of the model accounting for it: (1) health and mortality process; (2) hours and participation targets; (3) asset targets, with simulated person-periods classified by the most likely health state given three drawn measures; (4)–(6) combinations. Entries are in percent and need not add up; assets for all individuals are not reported because the baseline gap is only £119.

The results reinforce the distinction between correcting health dynamics and correcting the behavioral moments used in structural estimation. For the population as a whole, replacing the labor-supply targets closes essentially the entire gap in working-age earnings and hours costs, while replacing the health process alone does not. For individuals initially in bad health, the health process is more important: replacing it closes 77 percent of the gap in remaining-life monetary costs and approximately the entire gap in longevity. Combining the health process and labor-supply targets closes 89 percent of the monetary-cost gap and 94 percent of the welfare-cost gap at the lower VSL target for this group.

Replacing the wealth targets has negligible effects on the estimated health costs, both alone and in combination with other inputs. The one caveat concerns asset costs in the full population, which Table G.1 does not report: the difference in asset costs between the two baseline specifications is only £119, so modest changes induced by individual replacements translate into large shares that are not informative about the relative importance of the inputs.

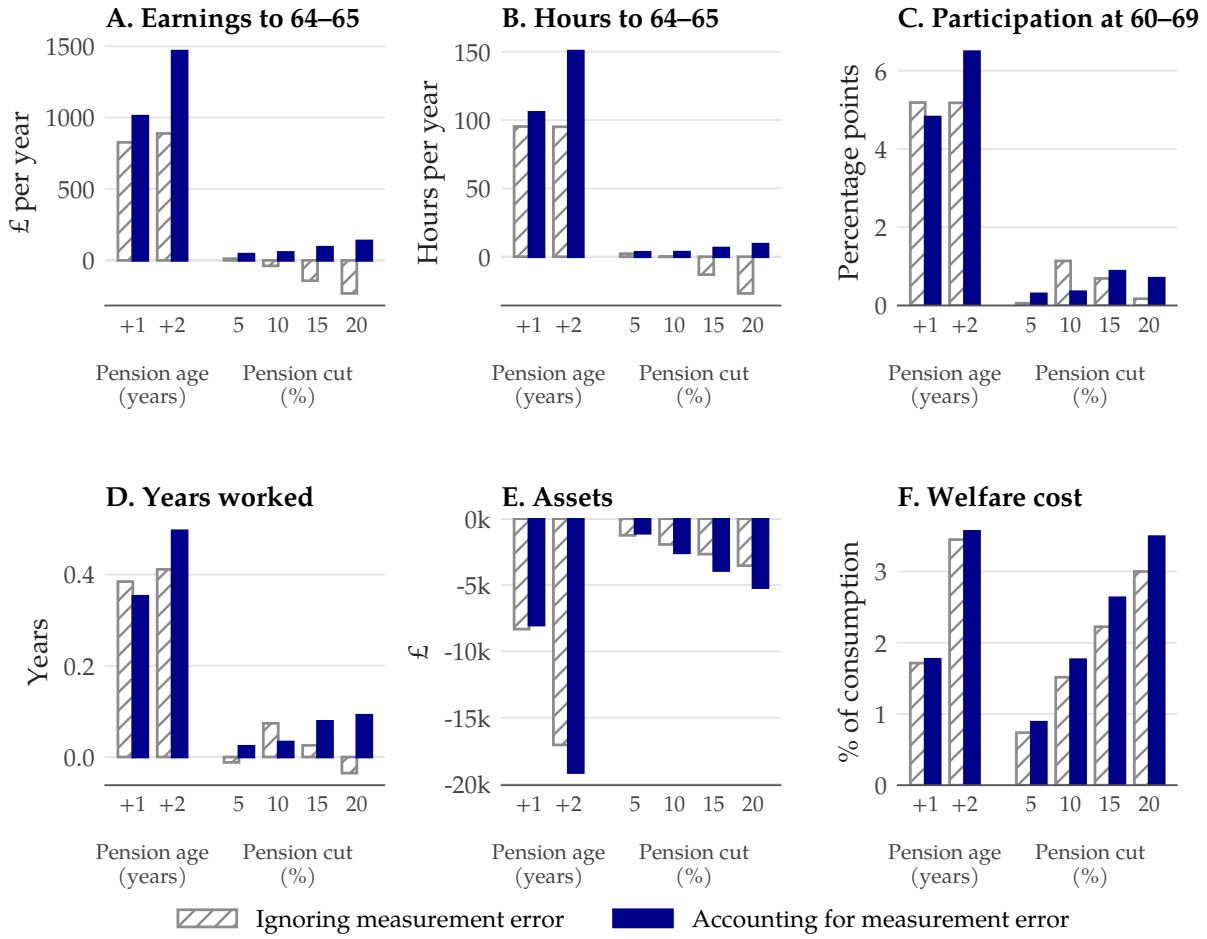
Construction of the wealth-target replacement. The wealth-target replacement requires a different procedure from the labor-supply replacement because wealth targets are percentiles rather than conditional means. While regressions on posterior health probabilities recover health-specific mean hours and participation rates under the maintained exclusion restriction, they do not directly recover conditional wealth percentiles. As described in Section 3.3 and Appendix E.4, the latent-health specification instead constructs wealth targets by classifying individuals according to their modal posterior health state. To implement the corresponding replacement, we generate three health indicators for each simulated observation in the specification ignoring measurement error, drawing from the estimated emission distributions conditional on its simulated health state. We then compute posterior health probabilities, classify observations by their modal state, and calculate the wealth percentiles of the resulting groups.

This procedure reproduces the classification rule of the latent-health specification, although the simulated health state generating the indicators corresponds to self-reported rather than latent health. The modal classification agrees with the generating state in approximately 97 percent of simulated person-periods. The initial distribution and labor-supply targets remain unchanged. The negligible effects of the replacement indicate that changing the construction of wealth targets, holding the other inputs fixed and allowing structural parameters to adjust, contributes little to the differences in estimated health costs.

G.3 Pension Policy Counterfactuals

Our main counterfactual eliminates exposure to bad health. We now examine whether accounting for health measurement error also changes the predicted effects of policies that alter retirement incentives while leaving health risks unchanged. We consider two reforms: increasing the pension eligibility age by one or two years, and reducing state pension benefits by 5, 10, 15, or 20 percent. Each reform is announced at model entry, and households reoptimize under the new policy. We simulate the same individuals using identical realizations of exogenous shocks and measure policy effects relative to each specification's own baseline. Figure G.1 reports the results.

Figure G.1: Pension policy counterfactuals: Changes relative to baseline



Notes. Each bar reports the change in an outcome relative to the corresponding specification's baseline, averaged over all simulated individuals. The reforms increase pension eligibility age by one or two years or reduce state pension benefits by 5–20 percent. Policies are announced at model entry, households reoptimize, and simulations use the same individuals and shock realizations as the baseline.

Raising the pension eligibility age increases labor supply in both specifications, although the magnitude of the response differs. Under a two-year increase, annual earnings through ages 64–65 rise by £1,465 in the latent-health model, compared with

£889 when measurement error is ignored. The corresponding increases in hours worked are 151 and 95 per year. Despite the larger labor-supply response, the predicted welfare costs are similar across specifications: 3.56 percent of consumption when accounting for measurement error and 3.45 percent when ignoring it. Reductions in state pension benefits generate relatively small labor-supply responses in both specifications but larger effects on assets and welfare. For a 20 percent reduction, assets fall by approximately £5,200 in the latent-health model and £3,500 when measurement error is ignored, while the welfare costs are 3.49 and 3.00 percent of consumption.

Overall, accounting for health measurement error changes some predicted labor-supply and saving responses to pension reforms, but has relatively modest effects on their welfare costs. This contrasts with the much larger differences in the estimated costs of bad health in Section 5.1. The implications of health measurement error for counterfactual analysis therefore depend on the policy considered and are most pronounced when the counterfactual directly changes exposure to bad health.

G.4 Robustness to Alternative Health Indicators

The baseline latent-health specification uses self-reported general health, mobility conditions, and ADL/IADL limitations. A potential concern is that self-reported health may depend on labor-market circumstances, violating the assumption that the indicators contain no additional information about economic outcomes conditional on latent health. For example, individuals' employment status may influence how they report their general health, potentially biasing the health-specific economic moments used in structural estimation. To assess the sensitivity of our results, we replace self-reported general health with reported pain, retaining the other two indicators, and re-estimate the entire model. Table G.2 reports the estimated costs of bad health under the two measurement systems, alongside those of the model ignoring measurement error, which is unchanged because it continues to treat self-reported health as the observed state.

The estimated costs of bad health remain substantially larger than those obtained when measurement error is ignored. For the population as a whole, the earnings cost is £2,450 per year under the alternative indicators, compared with £2,554 under the baseline indicators and £1,778 when measurement error is ignored. For individuals initially unhealthy, the corresponding costs are £9,535, £11,036, and £5,107. The result extends beyond earnings. Among individuals initially unhealthy, the estimated gain in longevity from remaining healthy is 6.39 years under the alternative indicators, compared with 3.51 years when measurement error is ignored. The corresponding welfare costs, calibrated to a VSL of £1.39 million, are 16.08 and 8.14 percent of consumption.

The alternative indicators generally produce somewhat smaller costs than the baseline latent-health specification, particularly for individuals initially unhealthy. These differences may reflect employment-related reporting behavior in self-reported health, but they may also arise because pain and general health capture different aspects of underlying health. The exercise cannot distinguish these explanations or rule out reporting biases in the alternative indicators. Nevertheless, the central quantitative conclusion is robust to replacing self-reported health: accounting for measurement error continues to imply substantially larger economic costs of bad health.

Table G.2: Costs of Bad Health under Alternative Health Indicators

Outcome	**Ignoring ME**	**Accounting for ME**	
		Self-reported health, mobility, ADL/IADL	Pain, mobility, ADL/IADL
All individuals			
Annual earnings (£)	1,778	2,554	2,450
Annual hours worked	151	220	190
Years worked	0.69	0.99	0.88
Annual consumption (£)	281	112	159
Assets (£)	2,432	2,550	3,147
Annual monetary cost (£)	1,978	2,641	2,577
Annual monetary cost, up to 64–65 (£)	1,863	2,886	2,839
Years of life gained	2.53	3.54	3.33
Welfare cost, VSL £1.39m (%)	6.67	10.85	9.90
Welfare cost, VSL £4.17m (%)	11.82	18.51	17.16
Initially unhealthy			
Annual earnings (£)	5,107	11,036	9,535
Annual hours worked	483	980	773
Years worked	1.61	3.50	2.77
Annual consumption (£)	865	2,439	1,984
Assets (£)	14,232	42,632	35,362
Annual monetary cost (£)	3,720	7,345	6,516
Annual monetary cost, up to 64–65 (£)	6,105	11,609	10,330
Years of life gained	3.51	7.50	6.39
Welfare cost, VSL £1.39m (%)	8.14	18.71	16.08
Welfare cost, VSL £4.17m (%)	14.53	31.56	27.46

Notes. The last two columns report the model accounting for measurement error, estimated using either self-reported general health, mobility difficulties, and ADL/IADL limitations (baseline), or pain, mobility difficulties, and ADL/IADL limitations (alternative). For each measurement system, we re-estimate the health and mortality process, emission distributions, wage process, spousal income, initial distribution, and structural parameters. Each entry reports the difference between the always-healthy counterfactual and the corresponding baseline, as defined in Table 3. The specification ignoring measurement error is unchanged across measurement systems.