The Welfare Cost of Late-life Depression

Ray Miller* Sayorn Chin† Ashish Kumar Sedai†

July 5, 2022

Abstract

We quantify the welfare cost of depression among older Americans by estimating a panel VAR model of mental and physical health, labor supply, and consumption using data from the Health and Retirement Study. We use the estimated model and age sixty joint distribution of outcomes to simulate life-cycle paths with and without prevalence of depressive symptoms after age sixty. We estimate that the prevalence of late-life depression costs an average of between 0.85 and 2.1 years in quality-adjusted life expectancy per person. Moreover, depression may result in an average loss of labor supply of up to 1.1 months and lifetime consumption of up to $16,000. Combining into a single compensating variation welfare metric, we estimate a bound on the average welfare cost of depression of 8-15% of annual consumption after age sixty. On aggregate, this amounts to roughly $180-360 billion annually. We also project that while the average welfare cost of late-life depression is declining slightly over birth cohorts, the welfare burden is becoming significantly more unequal.

JEL classifications: I14, J14, J11, J26

Keywords: depression, mental health, aging, retirement, consumption, cost-utility analysis

*Assistant Professor, Colorado State University, Department of Economics, Fort Collins, CO, 80523. Email: ray.miller@colostate.edu

†PhD Candidate, Colorado State University, Department of Economics, Fort Collins, CO, 80523.

1 Introduction

Depression is one of the leading causes of emotional distress and lower quality of life among older adults (Blazer, 2003; Sivertsen et al., 2015). Depression and depressive symptoms are also highly correlated with other physical and psychiatric conditions in older populations (Moussavi et al., 2007; Vaughan et al., 2015; Soysal et al., 2017; Chu et al., 2019). Increasing depressive symptoms with age have also been shown to be predictive of an increased risk of mortality (Bruce et al., 1994; Chui et al., 2015). Nonetheless, under-treatment of depression remains prevalent in older populations despite the wide availability of effective treatments (Barry et al., 2012; Kok and Reynolds, 2017). Improving our understanding of the comprehensive costs of late-life depression may be a fruitful avenue for expanding uptake of effective antidepressants and treatment therapies.

While preventing and treating late-life depression is of major social importance in its own right, significant spillover benefits are also possible. Empirical studies have found depression to be related to increased risk of frailty, reduced mobility, functional limitations, and progression of chronic diseases (Stuck et al., 1999; Penninx et al., 1999; Ciechanowski et al., 2000; De Groot et al., 2001; Geerlings et al., 2001; Rubio-Guerra et al., 2013; Vaughan et al., 2015; Chiang et al., 2015; Soysal et al., 2017; Penninx, 2017; Lwin et al., 2020). This has led some researchers to hypothesize a causal link from depression to poor physical health in older adults. However, whether the association between depression and physical health is driven by reciprocal influences or common causes remains widely debated (Mayerl et al., 2020).

Many theoretical explanations for why depression would affect physical health have been proposed (Penninx et al., 1999; Bruce, 2001). It could be that depressive symptoms such as sleep disturbance or lost appetite have a direct effect on functional decline and disability. There could also be indirect effects through intermediate behaviors (Bruce, 2001). For example, depressive symptoms could reduce motivation and lead to reduced medical care or poor health behaviors (e.g., smoking, poor nutrition, reduced physical activity). Other proposed mechanisms include antidepressant use (Lakey et al., 2012), increased allostatic load (McEwen, 2003), or other neuronal, hormonal, and/or immunological alterations (Bruce, 2001).

Beyond health effects, late-life depression could also influence an individual's economic outcomes. For example, depression among older adults increases health service utilization and costs (Luppa et al., 2012). Standard consumer theory suggests this could have a negative contemporaneous effect on consumption expenditures. Consumption could also decrease with depression due to reduced productivity and earnings (Lerner and Henke, 2008) or even decreased utility from goods and services that are complements to good mental health. On the other hand, the life-cycle hypothesis suggests that an unexpected depressive episode could *increase* contemporaneous consumption if there is an associated decline in life expectancy. Moreover, there could be additional dynamic effects that persist over time, for example if depression leads to an early retirement (Doshi et al., 2008; Rice et al., 2011). In the presence of such dynamic

effects, cross-sectional correlations between depression and other health and economic outcomes would only reveal part of the larger story.

In this paper, we adopt a life-cycle approach to better quantify the welfare cost of late-life depression when incorporating persistence and dynamic spillover effects. We extend the panel VAR model proposed by Miller and Bairoliya (2022) to simulate the joint evolution of health and economic outcomes, adapted to include the onset and persistence of late-life depression. We estimate the model using longitudinal data from the Health and Retirement Study (HRS) spanning more than twenty years. Using the observed joint distribution of outcomes at age sixty as initial conditions, we show that model simulations are able to closely match the empirically observed evolution of depressive symptoms, physical health, labor supply, and consumption.

Equipped with our simulation model, we next estimate the welfare cost of late-life depression. As the causal relationships between depressive symptoms and other health and economic outcomes remains unsettled, we take a bounds analysis approach. First, we estimate a lower bound on welfare costs by assuming there are no spillover effects on other model outcomes. More specifically, we leave all expected paths of comorbidities, mortality, and economic outcomes at their baseline levels and only remove the health utility penalty associated with depression at age sixty-two and older. We follow this with an upper bound estimate calculated by running a second set of counterfactual simulations starting from the same initial conditions but removing any possibility of depressive symptoms after age sixty. We consider this an upper bound as it assumes all the statistical relationships estimated in the restricted VAR model are entirely causal.

These analyses provide bounds on the expected costs of late-life depression in terms of quality-adjusted life years (QALYs), labor years, and dollars of consumption. We combine these differing costs using standard expected utility theory by calculating an ex-ante compensating variation (CV) measure of welfare. The welfare concept is akin to asking how much an individual would be willing to pay at age sixty to avoid any possibility of depressive symptoms over their remaining life. As our measure integrates multiple health and economic outcomes, it gives a more comprehensive view of well-being loss than the direct utility cost of depression alone. As it incorporates individual expectations over the entirety of remaining life from age sixty, it also provides a useful single metric of the ex-ante welfare cost of late-life depression.

1.1 Contributions

This study makes several contributions to our understanding of the welfare or utility burden of depression in older adults. First, previous studies have focused on estimating lost quality of life in older populations using cross-sectional observation, clinical settings, and/or limited longitudinal data (Sivertsen et al., 2015). Our estimates capture both contemporaneous and dynamic spillover effects on the evolution of depression, health, and economic outcomes over the entirety of remaining life. This provides a more complete measure of the total welfare burden of depression as it incorporates the cumulative burden of disease over time. We also provide an estimate that combines the

impact of depression on health-related quality-of-life, leisure, and consumption into a single measure grounded in economic and public health theory. Moreover, as our simulations are at the individual level within a larger representative sample, we are able to examine the entire distribution of welfare as opposed to only specific sub-samples or summary aggregates. This approach also allows us to examine how the level and distribution of welfare costs changed over birth cohorts, as opposed to cross-sectional changes over time.

Finally, we also contribute to the literature that has attempted to estimate the economic burden of depression to society. Studies have examined the impact of depression on direct medical costs and indirect workplace costs, including absenteeism from work and presenteeism while at work (Wang et al., 2003; Stewart et al., 2003; Lerner and Henke, 2008; Birnbaum et al., 2010; Luppa et al., 2012). Combined with suicide-related mortality costs, Greenberg et al. (2015) estimate the economic burden of major depression disorders in the U.S. was $210.5 billion in 2010. The total direct healthcare costs for treating depressive disorders among those aged 65 and older have been estimated at $9.8 billion in 2016. About $930 million out-of-pocket, $1.4 billion from private insurance, and the remaining $7.4 billion paid for by public insurance programs (Dieleman et al., 2020). While these costs center on the employer or healthcare side, we complement these studies by incorporating economic costs to private individuals. We also focus on older adults and quantify effects from the full range of depressive symptoms as opposed to only major disorders.

2 Data and methods

2.1 Data

The HRS is an ongoing longitudinal survey of U.S. individuals over the age of fifty and their spouses. The survey began in 1992 and data has since been collected every two years with new birth cohorts added periodically. There are currently eight birth cohorts in study—the early HRS cohort (born 1931-36), late HRS cohort (born 1937-41), AHEAD cohort (born before 1924), Children of Depression (born 1924-30), War Babies (born 1942-47), early Baby Boomers (born 1948-53), mid-Baby Boomers (born 1954-59), and late-Baby Boomers (born 1960-65). We use the publicly available RAND HRS Longitudinal File 2016 (V2) to obtain data on depression, health, mortality, and economic outcomes from 1992 to 2016. We also utilize other individual characteristics including age, education, gender, race, birth cohort, region, and occupation.

2.1.1 Depression

Depressive symptoms in the HRS were measured using the eight-item Center for Epidemiologic Studies Depression scale (CESD). The measure ranges from zero (no depressive symptoms) to eight, created by summing the respondent's number of "yes"

answers across eight survey items (with positive items reverse-coded).[1] The CESD is a common measure of depressive symptoms in older adults (Lewinsohn et al., 1997; Turvey et al., 1999; Steffick, 2000; Karim et al., 2015). The CESD was designed to measure a continuum of psychological distress (symptoms of depression), rather than determining the presence or absence of specific psychiatric disorders. However, a longer form CESD scale has been broadly validated against diagnostic interviews for depression and other anxiety disorders (Fechner-Bates et al., 1994; Lewinsohn et al., 1997). The eight-item CESD has also been shown to be a valid and reliable instrument of depression in a large sample of older Europeans (Karim et al., 2015).

2.1.2 Additional health outcomes

In addition to depression, our model incorporates data on comorbidities. These include eight binary indicators for ever having been diagnosed by a doctor with the following health problems—(1) high blood pressure or hypertension; (2) diabetes or high blood sugar; (3) cancer or a malignant tumor of any kind except skin cancer; (4) chronic lung disease except asthma such as chronic bronchitis or emphysema; (5) heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems; (6) stroke or transient ischemic attack (TIA); (7) emotional, nervous, or psychiatric problems; and (8) arthritis or rheumatism. We also include an indicator for ever reported difficulty with any activity of daily living (ADL) such as bathing, getting dressed, or walking across a room. ADL difficulties are a common health metric in older populations.

As a final health measure we use self-rated health status reported on a five-point scale from poor (one) to excellent (five). Self-rated health has been shown to be predictive of mortality in the HRS and other datasets, even after controlling for other health conditions, health behavior, and socioeconomic characteristics (Idler and Benyamini, 1997; Stenholm et al., 2014). This may reflect that people have private information about their health over and above disease diagnosis.

2.1.3 Economic outcomes

As our empirical focus is on individuals nearing the end of working life, we limit labor considerations to retirement. We treat retirement as an absorbing state in our model and define retired individuals as those reporting zero annual hours of paid work in the current or any previous survey wave.[2]

[1] About 12% of observations in our estimation sample were missing CESD score. These were imputed along with consumption and other missing data as detailed in the online appendix.

[2] We could also consider the intensive margin, partial retirement, and/or reentry into the workforce but this comes with additional model complexity. Moreover, retirement is likely to be the largest labor market decision for this age group, but we find relatively small effects of depression on retirement in our empirical analysis. See also the appendix for robustness results where we use full-time employment instead of zero hours for our definition of retirement.

We use consumption data from the Consumption and Activities Mail Survey (CAMS), which was sent to a random sub-sample of HRS respondents in off years of the core survey. We use the RAND 2017 CAMS data file (V1), which contains a constructed estimate of total household consumption from 2001-2015 based on household spending on durables, nondurables, transportation, and housing. We create our measure of individual consumption by subtracting out-of-pocket health spending from household consumption and then dividing by the total number of household members.[3] As consumption data is only available between the core HRS waves, we merge each CAMS wave with the HRS core data from the previous wave.[4]

A challenge to our analysis is that CAMS data is only available for approximately 20% of HRS respondents for the years 2000-2014. We follow Miller and Bairoliya (2022) and use closely related available data such as wealth and income to address missing consumption data by using the multiple imputation method proposed by Honaker and King (2010) for cross-sectional time-series data (see online appendix for details).

2.2 Simulation model

We extend and estimate the forecasting model proposed by Miller and Bairoliya (2022), adapted to include the onset and persistence of late-life depression. The model follows the structure of a panel vector autoregression (VAR) making it useful for microsimulations. Specifically, we use the model to repeatedly simulate potential outcome paths for each individual with and without the prevalence of late-life depression, given a set of initial (age sixty) conditions. Here we discuss the basic structure of the model and identifying assumptions. The online appendix provides additional details on sample selection, descriptive statistics, and model estimation procedures and results.

The general structure of the simulation model is illustrated in Figure 1. At the beginning of each time period, morbidity status is updated based on (correlated) random shocks, which in turn influence an individual's self-rated health status. Morbidities and self-rated health then contemporaneously influence an individual's reported depression status.[5] We choose this outcome sequence because 1) it is consistent with evidence that general health affects depression (Moussavi et al., 2007; Ambresin et al., 2014); 2) it allows block identification of the system for estimation (details below); and 3) it provides a more conservative estimate of the welfare cost of depression. On the last point, there is quite plausibly some contemporaneous reverse causation between

[3]Health spending includes health insurance, medication, health services, and medial supplies. We use the CPI-U to convert all waves to 2010 dollars. Household members include all residents but exclude spouses/parents living in nursing homes. We do not adjust for lower consumption of any resident children given the small number of children in HRS households.

[4]This is the recommended procedure for use of the RAND CAMS data file and is also consistent with the time structure of our simulation model.

[5]Note that we posit each of the morbidity states to contemporaneously influence depression both directly and through changes in self-rated health. For example, a stroke may lower an individual's self-rated health status which in turn may worsen depression. However, a stroke may also influence depression beyond any changes in self-rated health.

depression and general health (Rothermund and Brandtstädter, 2003; Moussavi et al., 2007). However, our later counterfactual simulations will assume that depression does *not* influence current period general health, yielding the more conservative estimate of its total welfare burden (see online appendix for robustness simulations where we relax this assumption). The simulations *will* allow current depression status to influence the evolution of health moving forward through general lagged effects. For example, mild depression today may result in a higher chance of stroke or lower self-rated health the following period. Moreover, higher order lagged effects allow, for example, the recent onset of depression to alter next period self-rated health more than if an individual has been living with depression for an extended period of time.

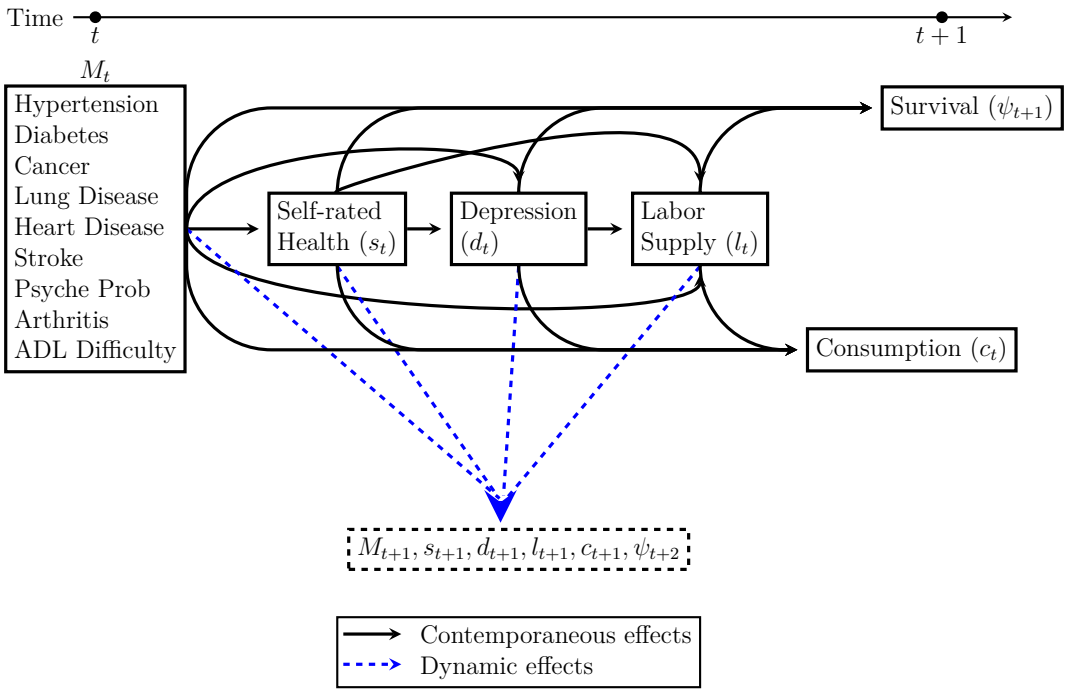

Figure 1: Forecasting model with one period lag

The latter part of the model allows morbidities, self-rated health, and depression to influence labor supply, consumption, and survival to following period of life. The assumed pathway from health to labor supply to consumption is consistent with evidence that health and depression affects the retirement decision (Currie and Madrian, 1999; Doshi et al., 2008; Rice et al., 2011), that consumption declines with retirement due to lower income and changing time constraints (Hall, 2009), and that health impacts economic outcomes, particularly at older ages (Smith, 1999).[6]

[6]In contrast, the effects of economic status on health appear concentrated during childhood and young adulthood when health trajectories are being established (Smith, 1999).

2.2.1 Panel VAR representation

While multiple lags are used in estimation of the simulation model, the following VAR(1) demonstrates the key features of the framework (see online appendix for extension to higher order lags). Let Y_{it} be a vector of outcomes for individual i at time t that includes depression d, log consumption c, retirement indicator r, self-rated health s, and our $n = 9$ morbidity states given by $n \times 1$ vector M. Outcomes are assumed to jointly evolve according to the structural VAR(1) model:

$$AY_{it} = BY_{it-1} + \epsilon_{it},$$

where ϵ is a vector of mean zero shocks that are normally distributed. The shocks are assumed to be independent and identically distributed (iid) across individuals and time and independent across outcomes. The main diagonal terms of matrix A are scaled to one and we assume that all parameters are homogeneous across individuals and time (e.g. $A_{it} = A \quad \forall i, t$). This implies that estimated relationships are assumed constant across individuals in our sample. For example, the marginal impact of depression on consumption does not change over time or across individuals.[7]

We estimate our model in five "blocks" of outcomes—the morbidity block consisting of n outcomes, and the self-rated health, depression, retirement, and consumption blocks, each consisting of one outcome. The unrestricted model can be written in block matrix form as:

$$\begin{array}{c} \overbrace{\phantom{-A_{11}}}^{n} \quad \overbrace{\phantom{-A_{12}\ -A_{13}\ -A_{14}\ -A_{15}}}^{4} \\ \left.\begin{array}{|c|cccc|}\hline -A_{11} & -A_{12} & -A_{13} & -A_{14} & -A_{15} \\ \hline -A_{21} & 1 & -a_{23} & -a_{24} & -a_{25} \\ -A_{31} & -a_{32} & 1 & -a_{34} & -a_{35} \\ -A_{41} & -a_{42} & -a_{43} & 1 & -a_{45} \\ -A_{51} & -a_{52} & -a_{53} & -a_{54} & 1 \\ \hline \end{array}\right\} \begin{array}{l} n \\ \\ \\ \\ \\ 4 \end{array} \end{array} \begin{bmatrix} M_{it} \\ \hline s_{it} \\ d_{it} \\ r_{it} \\ c_{it} \end{bmatrix} = \begin{array}{c} \overbrace{\phantom{B_{11}}}^{n} \quad \overbrace{\phantom{B_{12}\ B_{13}\ B_{14}\ B_{15}}}^{4} \\ \left[\begin{array}{|c|cccc|}\hline B_{11} & B_{12} & B_{13} & B_{14} & B_{15} \\ \hline B_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ B_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ B_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ B_{51} & b_{52} & b_{53} & b_{54} & b_{55} \\ \hline \end{array}\right] \end{array} \begin{bmatrix} M_{it-1} \\ \hline s_{it-1} \\ d_{it-1} \\ r_{it-1} \\ c_{it-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,it} \\ \hline \epsilon_{2,it} \\ \epsilon_{3,it} \\ \epsilon_{4,it} \\ \epsilon_{5,it} \end{bmatrix},$$

where $n \times n$ matrix A_{11} has main diagonal terms scaled to one.

The causal pathways we propose in Figure 1 suggest a block recursive system. Specifically, we assume that $A_{12} = A_{13} = A_{14} = A_{15} = 0$ in the morbidity block, $a_{23} = a_{24} = a_{25} = 0$ in the self-rated health block, $a_{34} = a_{35} = 0$ in the depression block, and $a_{45} = 0$ in the retirement block. In other words, we assume the contemporaneous causal pathway runs from morbidities to self-rated health to depression to retirement to consumption. However, we allow health and retirement to affect future outcomes

[7] As detailed below, all outcomes besides consumption are simulated using nonlinear models (e.g., ordered probit for depression). So while the coefficients are assumed homogeneous in the nonlinear models, obviously marginal effects differ across individuals. For example, the marginal impact of depression on retirement probability will depend on other modeled characteristics like age, calendar year, and co-morbidities.

through lagged effects.[8] Block triangulation of the system eliminates simultaneity across blocks and allows for block-by-block estimation.[9]

2.2.2 Exogenous characteristics

We also allow the evolution of outcomes in the simulation model to depend on a set of exogenous individual characteristics. Denoting the $k \times 1$ vector of exogenous regressors X_{it}, the VAR(1) model may be written:

$$AY_{it} = BY_{it-1} + CX_{it} + \epsilon_{it}. \tag{1}$$

Exogenous characteristics include a linear calendar year trend and dummies for age, education, gender, race, census division, census occupation code, birth cohort, and a post-2008 recession indicator.[10] In order to replicate the observed variance in consumption in the data, we also include a time invariant individual fixed effect π in the consumption equation. The fixed effects acts as person specific drift in the autoregressive process. The modeled exogenous characteristics can be explicitly written as:

$$CX_{it} = \underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} & C_{19} & 0 \\ \hdashline c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} & c_{29} & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} & c_{37} & c_{38} & c_{39} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} & c_{47} & c_{48} & c_{49} & 0 \\ c_{51} & 0 & 0 & 0 & 0 & 0 & 0 & c_{58} & c_{59} & c_{510} \end{bmatrix}}_{(n+4)\times k} \underbrace{\begin{bmatrix} Age_{it} \\ Education_i \\ Gender_i \\ Race_i \\ Division_i \\ Occupation_i \\ Cohort_i \\ Year_t \\ Post_t \\ \pi_i \end{bmatrix}}_{k \times 1}.$$

Note that we normalize $c_{510} = 1$ to allow identification of the unobserved fixed effects. We have also excluded time invariant exogenous characteristics from the consumption equation due to colinearity with the fixed effect. However, we include socioeconomic characteristics instead of additional fixed effects in the health and retirement equations because 1) morbidities and retirement are absorbing states and depression and self-rated health are ordinal, each of which poses difficulties in estimating dynamic

[8] Though we assume there is no such feedback from consumption and set $B_{15} = b_{25} = b_{35} = b_{45} = 0$.

[9] Note this produces the same results as the Cholesky decomposition of shocks from a reduced form VAR.

[10] The inclusion of age, cohort, and calendar year introduces some multicollinearity into the model, so interpreting point estimates on these variables should be done with caution. However, using the estimates for forecasting does not pose an issue (Holford, 1991).

panel models with fixed effects[11] and 2) the simpler model does well in replicating the observed dynamics of health and retirement in the data (see online appendix).

2.2.3 Morbidities

As there are multiple morbidities in the triangulated VAR system, we cannot identify the underlying structural parameters in the morbidity block. Instead we estimate the block as a reduced form VAR. We can premultiply the structural morbidity block by the inverse of matrix $-A_{11}$ to obtain the reduced form system:

$$M_{it} = \hat{B}_1 M_{it-1} + \hat{B}_2 s_{it-1} + \hat{B}_3 d_{it-1} + \hat{B}_4 r_{it-1} + \hat{C} X_{it} + e_{it},$$

where $\hat{B}_j = -A_{11}^{-1} B_{1j}$, $\hat{C} = -A_{11}^{-1} [C_{11}, \ldots, C_{19}]$ and $e_t = -A_{11}^{-1} \epsilon_{1,t}$. In this reduced form system all right hand side variables are predetermined at time t and morbidity states do not a have direct contemporaneous effect on each other. However, the error terms e_t are composites of morbidity specific structural shocks and thus are potentially correlated across morbidity states (i.e. $cov\left(e_{it}, e'_{it}\right) \neq 0$). This allows for contemporaneous correlation in the probability of morbidity states. For example, the onset of heart disease may be correlated with the onset of hypertension or stroke due to correlated contemporaneous shocks. Reduced form morbidity shocks are assumed to follow a standard multivariate normal distribution with an $n \times n$ covariance matrix given by Σ.

As morbidity outcomes are binary, forecasting of the measures is not a true linear VAR process. Instead, we assume a continuous latent variable $m^\star$ underlies each observed outcome such that:

$$\begin{aligned} m_{j,it} &= 0 \quad if \quad m^\star_{j,it} \leq 0 \\ m_{j,it} &= 1 \quad if \quad m^\star_{j,it} > 0 \end{aligned}$$

for $j = 1 \ldots n$. The estimated reduced form VAR can then be written:

$$\begin{bmatrix} m^\star_{1,it} \\ \vdots \\ m^\star_{n,it} \end{bmatrix} = \begin{bmatrix} \hat{b}_{11} & \cdots & \hat{b}_{1n} \\ \vdots & \ddots & \vdots \\ \hat{b}_{n1} & \cdots & \hat{b}_{nn} \end{bmatrix} \begin{bmatrix} m_{1,it-1} \\ \vdots \\ m_{n,it-1} \end{bmatrix} + \hat{B}_2 s_{it-1} + \hat{B}_3 d_{it-1} + \hat{B}_4 r_{it-1} + \hat{C} X_t + \begin{bmatrix} e_{1,it} \\ \vdots \\ e_{n,it} \end{bmatrix}. \tag{2}$$

Note that each latent morbidity variable is determined by lagged values of the other *observed* self-rated health, depression, and morbidity states. As we have assumed joint normality in the error term, this morbidity block of equations is in the form of a multivariate probit model.

[11] For example, it is not possible to estimate fixed effects for individuals that never enter an absorbing state in the data and estimated fixed effects would be needed for our simulations. This is one reason we exclude fixed effects from the health and retirement equations.

2.2.4 Self-rated health

Self-rated health is measured on a five point scale so we assume a continuous latent variable $s^\star$ underlies the observed self-rated health state. The relevant equation from system (1) can then be explicitly written as:

$$s^\star_{it} = A_{21}M_{it} + B_{21}M_{it-1} + b_{22}s_{it-1} + b_{23}d_{it-1} + b_{24}r_{it-1} + [c_{21}, \ldots, c_{29}]\, X_{it} + \epsilon_{2,it}, \quad (3)$$

with the observed self-rated health state defined as:

$$s_{it} = \delta \;\; if \;\; \kappa_{\delta-1} < s^\star_{it} < \kappa_\delta \;\; for \;\; \delta = 1, \ldots, 5$$

for cut-points $(\kappa_0, \ldots, \kappa_5)$. The worst health state (poor) is given by $\delta = 1$ and the best health state (excellent) by $\delta = 5$. We assume ϵ_2 is an iid shock with standard normal distribution so that the evolution of self-rated health follows an ordered probit structure. Unlike the morbidity block, block triangulation of the system allows this equation to be estimated independently of other outcome blocks with all structural parameters identified.

2.2.5 Depression

Similar to other health outcomes, we assume a continuous latent variable $d^\star$ underlies the observed depression state such that the forecasting equation given in system (1) can be written:

$$d^\star_{it} = A_{31}M_{it} + B_{31}M_{it-1} + a_{32}s_{it} + b_{32}s_{it-1} + b_{33}d_{it-1} + b_{34}r_{it-1} + [c_{31}, \ldots, c_{39}]\, X_{it} + \epsilon_{3,it}, \quad (4)$$

with the observed depression state defined as:

$$d_{it} = \delta \;\; if \;\; \kappa_\delta < d^\star_{it} < \kappa_{\delta+1} \;\; for \;\; \delta = 0, \ldots, 8$$

for cut-points $(\kappa_0, \ldots, \kappa_9)$ with $\delta = 0$ representing the no depressive symptoms state and $\delta = 8$ the worst depression state. Note that latent depression is assumed to depend on the lagged value of the *observed* depression category to incorporate the persistence in depression over time. We assume ϵ_3 is an iid shock with standard normal distribution yielding an ordered probit structure for the depression model. Given our block recursive system, this equation may also be estimated independently of other blocks with all structural parameters identified.

2.2.6 Retirement

As retirement is a binary outcome, we again assume a continuous latent variable $r^\star$ underlies the observed outcome such that:

$$r_{it} = 0 \quad if \quad r^\star_{it} \leq 0$$
$$r_{it} = 1 \quad if \quad r^\star_{it} > 0.$$

Conditional on working the previous period, the retirement block equation is given by:

$$r_{it}^{\star} = A_{41}M_{it} + a_{42}s_{it} + a_{43}d_{it} + B_{41}M_{it-1} + b_{42}s_{it-1} + b_{43}d_{it-1} + [c_{41}, \ldots, c_{49}]\, X_{it} + \epsilon_{4,it}. \quad (5)$$

Note that as retirement is an absorbing state, we set $b_{44} = 0$. In addition to exogenous individual characteristics, retirement is influenced by current and lagged values of health (depression, self-rated health, and specific morbidities). We assume ϵ_4 is an iid shock with standard normal distribution implying the retirement model has a standard probit structure.

2.2.7 Consumption

The consumption forecasting equation given in system (1) can be explicitly written as:

$$c_{it} = A_{51}M_{it} + a_{52}s_{it} + a_{53}d_{it} + a_{54}r_{it} + B_{51}M_{it-1} + s_{52}d_{it-1} + b_{53}d_{it-1} + b_{54}r_{it-1} + b_{55}c_{it-1} + c_{51}Age_{it} + c_{58}Year_t + c_{59}Post_t + \pi_i + \epsilon_{5,it}. \quad (6)$$

This equation is in the form of a standard linear dynamic panel data model with lagged dependent variable and individual level fixed effects. Block triangulation of the system also allows this equation to be estimated independently of other blocks with all structural parameters identified including the variance of ϵ_5.

2.2.8 Mortality

Mortality probabilities are estimated independently of the VAR system above as all other outcomes are conditional on survival. Survival from time period $t-1$ to time period t is modeled by:

$$\psi_{it} = I\left(\sum_{k=1}^{K}\left[\gamma_k^M M_{it-k} + \gamma_k^s s_{it-k} + \gamma_k^s d_{it-k} + \gamma_k^r r_{it-k}\right] + \delta X_{it} + u_{it} > 0\right), \quad (7)$$

where $I(.)$ is an indicator function and $\psi = 1$ indicates survival, X the vector of observed individual characteristics previously defined, and u_{it} an iid random shock with standard normal distribution. The model allows K lags of health, depression, and retirement to influence survival probability.

2.3 Welfare measure

We use an ex-ante consumption-compensating variation (CV) measure to quantify the welfare costs of late-life depression using simulations from our VAR model. We first define expected lifetime utility at age j for individual i as:

$$E\left[\sum_{a=j}^{J}\psi_{ia}\beta^{a-j}\phi(h_{ia})\left[\bar{u} + log(c_{ia}) + \nu(l_{ia})\right]\right]$$

where c is consumption, l leisure, h health, and ψ is a survival indicator. Health measure h is a vector of modeled morbidities, self-rated health, and depressive symptoms. Expectations are taken over the uncertain path of all outcomes after age j. This simple formulation yields an additive decomposition of welfare allowing us to add cumulative corrections for the cost of depression on comorbidities, mortality, leisure, and consumption (see online appendix for derivation). We also check the robustness of our results to more general preferences (see robustness section 3.5 for results). We model health in the utility function to map to the large literature on quality-adjusted life years (QALYs). Specifically, we assume utility from consumption and leisure each period is scaled by the health function $\phi(h) \in [0,1]$. Here, $\phi(h) = 1$ represents utility in the "best" health state and $\phi(h) = 0$ represents death. In this form, $\psi\phi(h)$ provides a measure of QALYs. For example, a year spent in the best health state is a single QALY and represented by $\psi\phi(h) = 1$.

Let $U_{ij}(1-\lambda)$ denote the expected lifetime utility at age j from the outcome bundles of individual i if consumption is multiplied by a factor $(1-\lambda)$ at each age and realization of the world:

$$U_{ij}(1-\lambda) = E\left[\sum_{a=j}^{J} \psi_{ia}\beta^{a-j}\phi(h_{ia})\left[\bar{u} + log\left((1-\lambda)c_{ia}\right) + \nu(l_{ia})\right]\right].$$

The consumption-compensating variation measure of welfare for individual i, λ_{ij}, is derived through the condition:

$$U_{mj}(1-\lambda_{ij}) = U_{ij}(1), \tag{8}$$

where U_{mj} refers to the expected lifetime utility from the outcome bundles in the absence of any possible depression after age j. In words, λ_{ij} is the proportion of the individual's (depression-free) consumption they would be willing to give up at every age starting from j (in all possible realizations of the world and holding health and leisure fixed) to eliminate all possibility of depression after age j. For example, if person i expects depression to be a serious problem in late life, they may have a welfare measure $\lambda_{ij} = 0.3$. This implies they would be ex-ante willing to give up to 30% of their consumption in every period from age j to avoid any possibility of depression. Note that this does not imply that depression must have any direct effect on consumption. It could be the entire 30% is due to lost health utility. The welfare measure simply provides a convenient way to combine multiple types of utility costs (e.g., health, mortality, retirement) into a common metric—dollars of consumption. Moreover, as this measure is based on potential outcomes over remaining life, it encompasses the likelihood of persistence and emergence of depression over remaining life.

In order to gain a sense of the aggregate cost of depression, we also calculate the product of an individual's expected remaining lifetime consumption at age sixty (ELC) and our CV measure: $\lambda \times ELC$. This is a similar concept (but not the same) as an individual's willingness-to-pay at age sixty to eliminate all possibility of depression after age sixty. Effectively, it is an individual's *expected* willingness-to-pay or the expected value of consumption they are willing to forgo.

2.3.1 Health utility weights

Analysis using our welfare model requires calibration of preference parameters. This includes parameters of the function $\phi(h)$ mapping health states into flow utility. We assume health utility depends linearly on our health state vector: $\phi(h_t) = \gamma h_t$. Our health utility weights γ are derived from the Health Utilities Index Mark 3 (HUI3) instrument which was collected from approximately 1,200 respondents in the HRS in the year 2000. The HUI3 was developed to produce cardinal utility scores on the conventional utility scale ranging from zero (death) to one (best health) and has been extensively used in the literature on health utilities (Furlong et al., 1998; Feeny et al., 2002; Horsman et al., 2003).[12] We use the HUI multi-attribute utility score (*hui3ou*).

The HUI3 was conceptualized such that $u(h_i) = HUI3_i \times u(h_{best})$ for individual i and general utility function $u(.)$. For example, a year in the best health state is equal in utility to two years with $HUI3 = 0.5$. As another example, three years of life spent with an $HUI3 = 0.33$ is equal to the same utility as about one year in perfect health, or a single QALY. In the context of our model, we assume that the HUI3 measures the relative utility across health states *holding consumption and leisure fixed*:

$$\gamma h_i \left[\bar{u} + log(c_i) + \nu(l_i)\right] = HUI3_i \times h_{best} \left[\bar{u} + log(c_i) + \nu(l_i)\right].$$

This approach is consistent with the HUI3 instrument where the interview script reads: "*when imagining yourself in these health states please remember that where you live, your income, your friends, and family would be the same as now.*" With this assumption, the above equation simplifies to $\gamma h_i = HUI3_i$ when $h_{best} = 1$. The utility weights γ can then be estimated by regressing the HUI3 utility score on depression score, self-rated health, and all morbidity indicators. Results are also robust to relaxing the assumption of holding consumption and leisure fixed (see online appendix).

2.3.2 Calibration of other parameters

Leisure is normalized to one for retired individuals. Leisure for working individuals is set to $0.66 = 1 - (2000/5,840)$, based on an annual time endowment of 5,840 hours (16 hours a day $\times$ 365 days) and 2,000 hours of work. Preferences over leisure are defined by $\nu(l) = -\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}$, where ϵ is a constant Frisch elasticity of labor supply. Note that in addition to retirement, we are assuming that the cost of depression on leisure is captured through the health utility function $\phi(h)$. In other words, depression is assumed to make leisure time less valuable in terms of utility. It is feasible there could be an additional direct time cost of depression, for example due to additional time needed for treatment or personal care. By not explicitly attempting to include a direct time cost, our model is again pushed towards a conservative estimate of the costs of depression. We follow Miller and Bairoliya (2022) and set the disutility weight θ such that the marginal cost of leisure equals the marginal benefit for the median individual

[12] HUI3 scores less than zero are possible and represent current health states worse than death.

in our sample. This gives us a benchmark $\theta = 7.8$. We use a benchmark value of $\epsilon = 1$ and a discount factor $\beta = 0.98$ implying an annual discount rate of one percent (with additional discounting implicit due to mortality risk). We examine robustness of results to each of these parameter values.

Finally, note that with our benchmark preferences, as long as flow intercept $\bar{u}$ plus log consumption is positive (and health utility is positive), a retired individual will prefer life to death in the current period.[13] After normalizing consumption to thousands of 2010 dollars, we set $\bar{u} = -log(3)$, implying that \$3,000 of consumption is needed for a retiree to maintain positive flow utility. This is approximately 10% of mean annual consumption in our sample. Although there is not much evidence on this value, Murphy and Topel (2006) argue 10% as a reasonable parameterization. This value also yields a mean and median value of remaining life for sixty year olds in our simulation sample of \$78,000 and \$47,000 per QALY.[14] In a review of the literature, Ryen and Svensson (2015) estimate mean and median values of life across studies of approximately \$98,000 and \$32,000.[15] Traditional values in the U.S. often range from \$50,000 to \$100,000 (Kaplan and Bush, 1982). In some robustness exercises, we show that using log consumption and a relatively low value of life in our benchmark likely yields conservative estimates of welfare costs.

2.4 Estimation and simulations

Equipped with our simulation model and welfare concept, our empirical analysis involves four steps.

1. We use data from the HRS to estimate the parameters of the simulation model. Here we use data on all individuals aged fifty and older from all available waves of the HRS from 1992-2016 (40,708 unique individuals and 238,091 total individual-year observations). See the online appendix for details on the model estimation sample and procedures.

2. We repeatedly simulate remaining life-cycle paths for all outcomes for a subsample of the HRS respondents using the parameter estimates and age sixty data as initial conditions. This simulation sample includes all individuals with

[13] A current period utility less than zero does not necessarily imply an individual would rather die than continue living, as it does not guarantee an expected remaining lifetime utility less than zero. For example, consumption may be expected to increase from next period onward.

[14] The value of life per QALY at age j is given by $VOL_j/E\left[\sum_{a=j}^{J}\psi_a\beta^{a-j}\phi(h_a)\right]$ where $VOL_j = U_{ij}(1)\,c_j/\phi(h_j)$.

[15] Ryen and Svensson (2015) document substantial variation across estimates of willingness-to-pay for a QALY, most notably with conversions based on revealed preferences of the value of statistical life (VSL) averaging 5-7 times higher than those based directly on stated preferences. The VSL studies reviewed are by definition measuring value of length of life, while stated preference studies elicited willingness-to-pay for pure quality of life improvements, pure length of life, or a mixture of both.

age sixty data and requisite lagged data for simulations. This yields representative results over four birth cohorts—early HRS (EHRS), late HRS (LHRS), War Babies, and (early) Baby Boomers. See the online appendix for details on initial condition descriptives, sampling weights and representativeness, and simulation procedure.

3. We estimate a lower bound of the welfare costs of depression after age sixty for each individual in our simulation sample by assuming there are no spillover effects on other model outcomes. More specifically, we leave all expected paths of morbidities, self-rated health, mortality, labor supply, and consumption at their baseline levels and only remove the health utility penalty associated with depression at age sixty-two and older (i.e., we set all the CESD weights to zero in Table 1). In doing so, we are assuming that all statistical associations estimated in our VAR model do not reflect any causal impact going from depression to other modeled outcomes (hence, no spillover).

4. We estimate an upper bound of the welfare costs of depression by running a new set of simulations starting from the same initial conditions but removing any possibility of depressive symptoms after age sixty. We consider this an upper bound as it assumes all the coefficients estimated in the simulation model are purely causal. We embed the baseline and counterfactual simulated data within our expected utility framework to construct a measure of the ex-ante welfare cost of future depression at age sixty for each individual in our simulation sample.

3 Results

3.1 Model estimates

We begin with estimation results from our simulation model to demonstrate the association between depression and other outcomes in the data. Selected results are provided in Figure 2 while the full set of results are available in the online appendix. The first panel shows the average marginal effects of depressive symptoms on the contemporaneous probability of retirement (controlling for other health outcomes as shown in model (1)). Results indicate that low and mild depression (CESD=0,1,2,3,4) do not have a significant association with the probability of retirement for older adults. However, as the severity of depression increases, a significant association emerges. For example, there is an increase in the probability of retirement of around 2 percentage points (pp) for a CESD score of five, rising to a point estimate of over 5 pp at higher levels of depression. The second panel of Figure 2 shows a small *positive* relationship between the severity of depression and contemporaneous log consumption. For example, a CESD score of five is associated with an increase in consumption of about 0.03 log points. Expectations about the longevity and quality of life may be a plausible explanation for this positive association among older adults. For example, lower expectations about a

long and healthy life might push older depressed adults to discount their future utility and hence consume more in the near term. Moreover, there is an additional indirect effect operating on consumption due to any changes in retirement induced by depression. For example, we saw that a CESD score above five is associated with an increased probability of retirement. Our estimates also show that retirement is associated with an immediate decrease in consumption of about 0.04 log points (see appendix table 5).

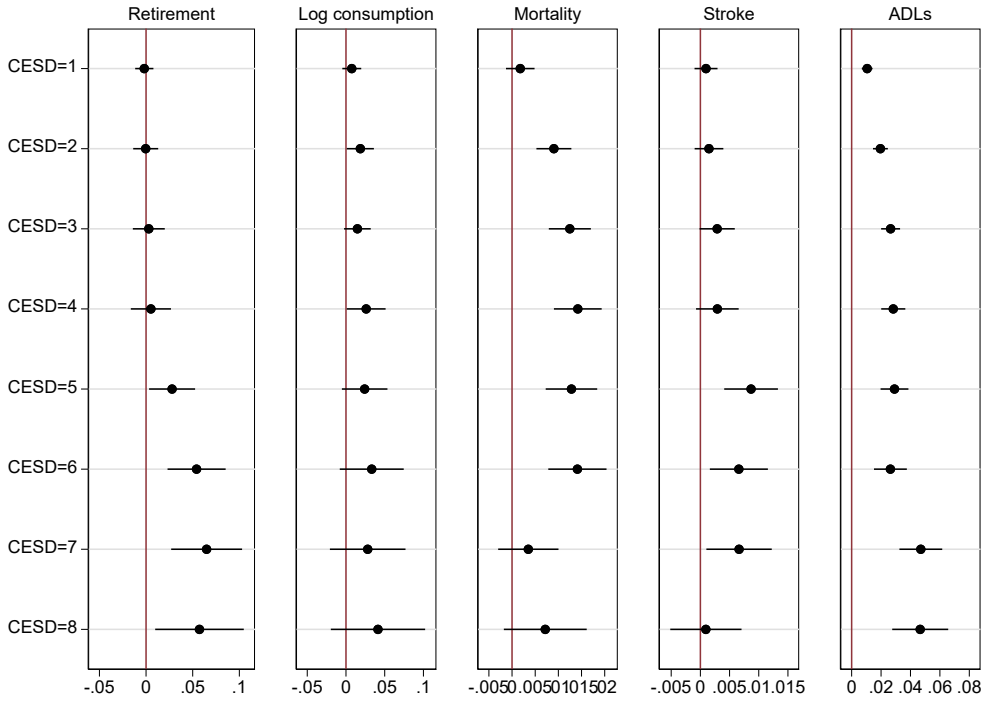

Figure 2: Selected estimation results

Notes: Dependent variables across columns. Average marginal effects on the probability of an outcome reported for probit results—retirement, mortality, stroke, and ADLs. Contemporaneous associations reported for retirement and log consumption as dependent variables. Lagged associations reported for mortality, stroke, and ADLs. CESD=0 (no depression) is the reference group. Spikes indicate 95% confidence intervals.

Panel three of Figure 2 shows a generally positive association between depressive symptoms and mortality. For example, an individual with a CESD score of three has about a 1 pp lower probability of surviving to the next model period compared to if they had a CESD score of zero. Note that at the highest CESD scores the association with mortality is moderately diminished, although results are somewhat noisy given that relatively few individuals have very high levels of depression (about 4% of our sample). In addition to a comparatively small sample with very high depression score, this pattern is also explained by indirect associations between high CESD score and mortality. For example, CESD scores higher than six have a strong correlation with several worse health outcomes like difficulties with ADLs (see the last panel of Figure

2). So part of the association between high CESD score and mortality is captured through poorer health.[16]

The final two panels help illustrate the dynamics of the system by showing the average marginal effects of current depression state on the probability of having a stroke or ADL difficulty the *following* model period. For example, a CESD score of five increases the probability of a stroke the following period by nearly 1 pp and the probability of having difficulty with ADLs by more than 2 pp. Moreover, these relationships continue to propagate dynamically throughout the system influencing the evolution other comorbidities and self-rated health along with future retirement and consumption decisions.

3.1.1 Simulation fit

A comparison between mean simulated CESD scores and those based on available data is shown by age and cohort in Figure 3. Additional comparisons for each outcome by cohort are provided in the online appendix. In both the data and simulations, mean CESD score tends to rise with age in the EHRS cohort.[17] For the LHRS cohort the relationship is U shaped, CESD scores decline until the age of 72, then begin to increase. For the younger War Baby and Baby Boomers cohorts there is less available data, but both cohorts have falling CESD scores over the sixties. Note that by construction, the data and simulations are the same at age sixty. However, using only age sixty data and the estimated model parameters, the simulations continue to match the data reasonably well even up to 24 years later (when the EHRS cohort is age 84). Overall, the simulations match the available aggregated data well suggesting our life-cycle dynamics model provides a good approximation of the underlying data generating processes.

3.1.2 Health utility weights

Table 1 provides our health utility weights γ estimated via a linear regression of HUI3 utility score on health outcomes. Depressive symptoms measured by the CESD scale have a strong and highly significant negative association with utility. For example, moving from no depressive symptoms (the base category) to a CESD score of three lowers flow health utility by 7.9 pp. Moving all the way to a score of eight lowers health utility by 29.3 pp. In addition to depression, self-rated health also has a strong association with health utility. For example, moving from poor health (the base category) to good health improves flow health utility by 25.0 pp. Conditions such as hypertension, diabetes, and cancer have little independent effect on health utility after controlling for their association with self-rated health, depression, and other comorbidities. Other morbidities such as stroke and arthritis have larger (and statistically significant) in-

[16] For example, when we exclude other health outcomes from our mortality equation, CESD scores higher than six show slightly *stronger* associations with mortality than lower CESD scores.

[17] Conditioning data on survival to the end of the simulation period to eliminate mortality bias yields lower CESD scores but similar dynamics for each cohort (see figure in online appendix).

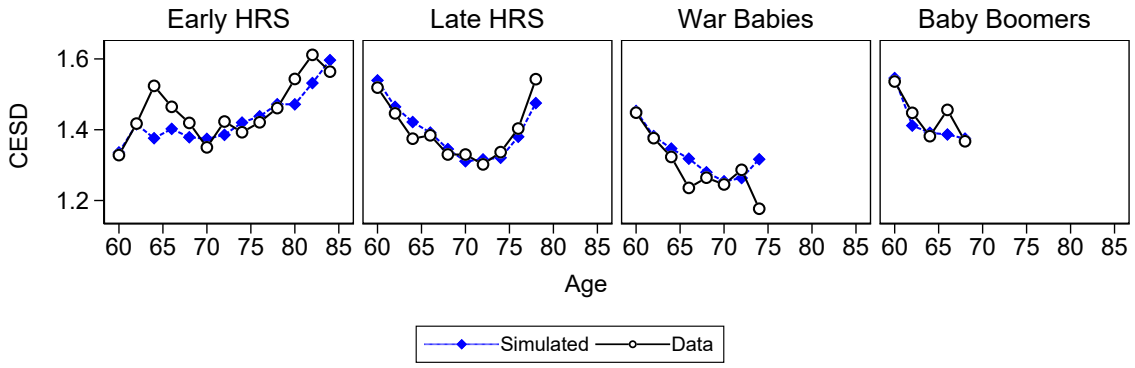

Figure 3: Mean of life-cycle CESD profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

dependent negative effects. The most influential morbidity indicator is difficulty with ADLs, which lowers health utility by an estimated 14.2 pp.

While the eight-point CESD does not map directly into clinical diagnosis of depression disorder, Turvey et al. (1999) propose a CESD score of six or higher to approximate cases of clinical depression. Our weights then imply that a clinically depressed individual in good self-rated health and without other comorbidities would have a health utility score between 0.56-0.68. A clinically depressed individual with poor self-rated health would have a score of 0.31-0.43. In a systematic review, Mohiuddin and Payne (2014) examine results from studies using indirect valuation methods to estimate health utility scores in alternate depressive states. They calculate pooled mean utilities across studies of 0.56 for mild, 0.45 for moderate, and 0.25 for severe depression. By comparison, our results are likely conservative in attributing health utility penalties to depressive states.

3.2 Cost of depression

We start with a detailed examination of the cost of depression in the EHRS cohort as it is the oldest of the four cohorts and contains the longest panel of available data. Table 2 shows the mean cost of depression after age sixty for the EHRS. The first column provides our lower bound estimate where we simply remove the health utility penalty of depressive symptoms but leave all simulated outcomes at their baseline levels. On average, removing the health utility penalty of depression increases quality-adjusted life expectancy by 0.85 years. The mean associated CV welfare measure is 0.084, implying a willingness-to-pay up to 8.4% of annual consumption over remaining life to avoid any possibility of depression. As shown in the final row, this amounts to an expected loss of $45,933 of lifetime consumption.

The final four columns of Table 2 provide results from re-simulating outcomes for

Table 1: Estimated health utility weights (γ)

Measure	Weight	SE
Depression		
CESD=1	-0.021	0.015
CESD=2	-0.087	0.018
CESD=3	-0.079	0.023
CESD=4	-0.094	0.028
CESD=5	-0.138	0.030
CESD=6	-0.172	0.039
CESD=7	-0.225	0.046
CESD=8	-0.293	0.056
Hypertension	0.004	0.012
Diabetes	-0.000	0.017
Cancer	0.007	0.017
Lung disease	-0.024	0.021
Heart disease	-0.034	0.015
Stroke	-0.073	0.022
Psych problem	-0.041	0.020
Arthritis	-0.053	0.012
Diff with ADL	-0.142	0.016
Self-rated health		
Fair	0.179	0.026
Good	0.250	0.027
Very good	0.331	0.028
Excellent	0.338	0.032
Constant	0.610	0.030

Notes: Results from regression of HUI3 score on CESD score, self-rated health, and morbidities. SE denotes standard error. $R^2 = 0.48$. N = 1,089.

Table 2: Mean costs of depression after age sixty

		Cumulative corrections			
	Depression	Comorbidities	Mortality	Leisure	Consumption
Expected loss					
QALYs	0.853	1.241	2.064	2.064	2.064
Labor supply (yrs)				0.095	0.095
Consumption (annual)					-0.631
CV (λ)	0.084	0.108	0.158	0.156	0.148
$\lambda \times$ELC	45.933	63.030	94.128	93.459	89.405

Notes: Estimates use base year respondent analysis weights. ELC denotes expected lifetime consumption. Consumption in $1000s.

each individual after removing any possibility of depression after age sixty. Each column cumulatively adjusts welfare for an additional outcome with our "fully adjusted" upper bound provided in the last column. For example, when accounting for the possible spillover effects of depression on comorbidities (column two), eliminating depression

increases quality-adjusted life expectancy by 1.24 years. The mean associated CV implies a willingness-to-pay up to 10.8% of annual consumption to avoid depression when accounting for these spillovers, or an expected loss of $63,030 of lifetime consumption. Further adjusting for the effect of depression on mortality rates yields an increase in quality-adjusted life expectancy of 2.06 years, a willingness-to-pay of 15.8% of annual consumption, or an expected loss of $94,128 in lifetime consumption.

Moving to leisure time, our simulations suggest that eliminating depression after age sixty could increase labor supply only by an average of about 1.1 months (0.095 years). This relatively small impact is likely due to the fact that many individuals in the simulation sample are already retired by age sixty and the direct effects of depression on retirement are minimal except with quite severe symptoms (recall Figure 2). As increased labor supply alone results in a loss in welfare due to less leisure time, the mean CV falls very slightly to 15.6% of annual consumption. Finally, simulations suggest eliminating depression could lower consumption by up to $631 annually. This is consistent with the positive contemporaneous association between depression and annual consumption shown in Figure 2. The mean associated CV implies the willingness-to-pay falls to 14.8% of annual consumption to avoid depression when accounting for these consumption losses. However, as we demonstrate in the next section, a fall in annual consumption does *not* imply a fall in lifetime consumption.

In order to gain a better sense of how depression influences the dynamics of other outcomes in the system, Figure 4 plots the average percentage change in expected outcomes with the exogenous elimination of all prevalence of depression after age sixty. The first two plots show that the elimination of depression after age sixty is associated with a significant decline in psychiatric problems, difficulty with ADLs, and to some extent lung disease. For example, the elimination of depression is associated with nearly a 30% decrease in the probability of diagnosed psychiatric problems by the late-seventies. Associated effects were fairly small for the other morbidities.[18] These patterns are consistent with previous literature that has also found depressive symptoms to be predictive of increased difficulties with ADLs, such as eating, dressing, and bathing among the elderly (Penninx et al., 1999; Cronin-Stubbs et al., 2000; Kivelá and Pahkala, 2001; Kazama et al., 2011; Sodhi and Al Snih, 2020). A number of studies have also established a strong correlation between depression and chronic lung disease such as Chronic obstructive pulmonary disease (COPD) (Dowson et al., 2001; Van Manen et al., 2002; Mikkelsen et al., 2004; Kunik et al., 2005; Wilson, 2006; Ryerson et al., 2011). Hypothesized mechanisms linking depression to the decline in ADLs and increased lung disease include increased lifestyle risk factors (e.g., smoking, poor nutrition, low physical activity) and less social integration. For example, depression has been shown to significantly reduce the effectiveness of smoking cessation programs (Kinnunen et al., 1996; Cinciripini et al., 2003).

[18]Cancer has relatively little association with depression or other morbidities. However, as eliminating depression improves chances of survival even when sick, there is actually a small increase in the prevalence of cancer (conditional on survival) starting around age eighty.

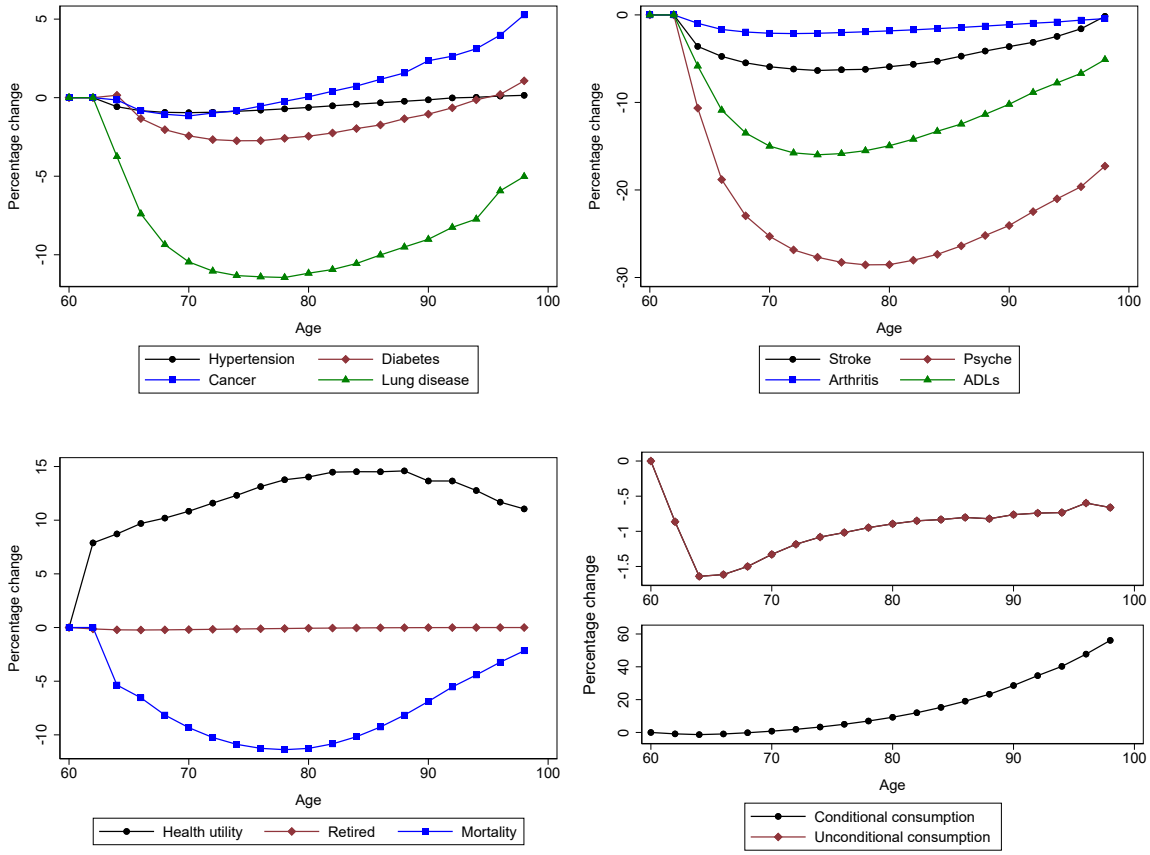

Figure 4: Expected cost of depression by age

Notes: Results plot percentage difference in expected outcomes with the exogenous elimination of all prevalence of depression after age sixty. Sample includes all individuals in the simulation sample. Expected outcomes in first three panels are conditional on survival.

The third plot of Figure 4 shows the upper bound effect of eliminating depression on health utility, labor supply, and mortality. The age-specific mortality rate is estimated to be over 10% lower by the early-seventies and remains more than 5% lower even into the nineties. In contrast, the probability of being retired falls only very slightly. This is consistent with the small association of depression with average labor supply discussed above. There is an immediate increase of about 8% in health utility at age sixty-two, which climbs to nearly 15% by age eighty. The final plot shows the response of consumption (unconditional and conditional on survival) to the elimination of depression. There is about a 1-2% decline in annual consumption conditional on survival throughout the remaining life-cycle. In contrast, when examining expected unconditional consumption (i.e. imputing zero consumption for the dead state), there is a large rise over time, reaching differences of more than 50% by the early nineties. These plots again highlight the small loss in annual consumption but potential gains in lifetime consumption due to an increase in life expectancy.

3.3 Distribution of cost

Another advantage of our approach is that we have individual level data and simulations so we are able to examine the entire distribution of estimated costs. Figure 5 shows the distribution of estimated QALYs lost due to late-life depression. The "depression only" curve plots our lower bound estimate with a mean of 0.853 as shown in Table 2. The distribution demonstrates substantial inequality in the expected direct utility cost of late-life depression—the worst off individuals expect a loss of nearly three QALYs. When adding the estimated spillover effects on comorbidities, the mean shifts to 1.241 and the distribution flattens. This implies the health utility cost of depression becomes even more unequal when accounting for spillovers on other comorbidities. When further adjusting for increased mortality rates associated with depression, the mean reaches 2.064 and inequality in the distribution continues to rise.

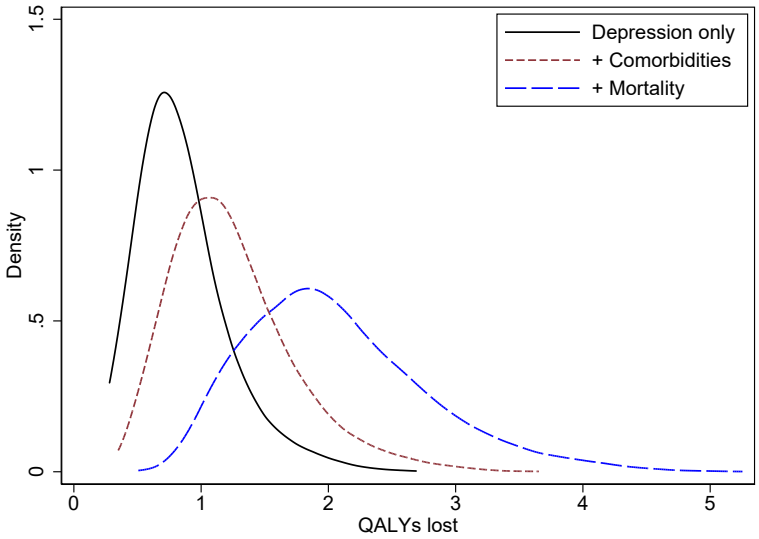

Figure 5: Expected lost quality-adjusted life years (QALYs) after age sixty

Turning to economic outcomes, Figure 6 shows the distribution of the estimated upper bound on the expected loss in consumption and labor supply associated with late-life depression. The first panel shows the change in expected annual consumption (conditional on survival). The negative values again demonstrate the small expected *gain* in annual consumption from late-life depression. For most individuals this gain is less a $1,000, though a small share expect to gain more than $3,000 in annual consumption. Despite this rise in annual consumption with depression, the second panel of Figure 6 shows there is an expected fall in total lifetime consumption for all individuals due to decreased life expectancy. The expected loss in lifetime consumption averages around $16,000, but ranges from almost zero to over $100,000. Finally, the third panel shows the distribution of lost labor supply. The potential loss due to early retirement is quite small—less than a year for all individuals.

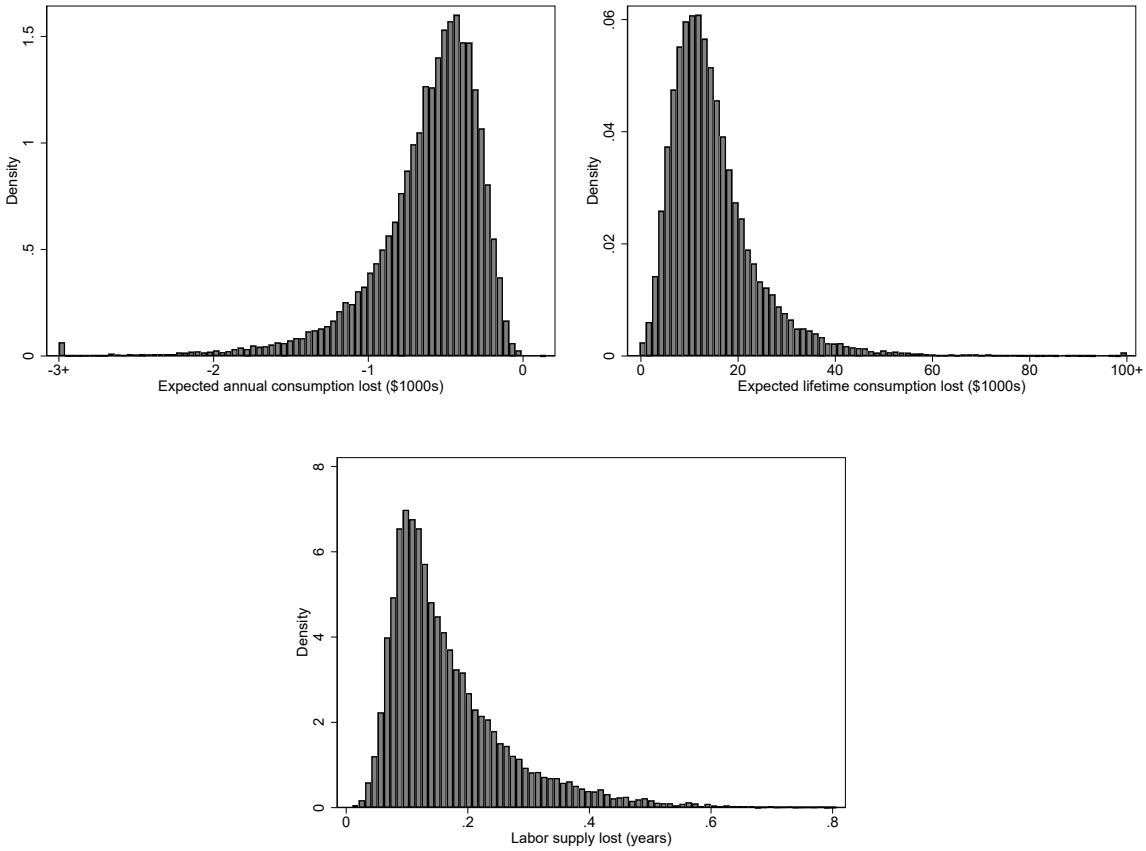

Figure 6: Expected consumption and labor supply loss after age sixty (full model)

Notes: Estimates use base year respondent analysis weights. Labor supply lost conditional on working at age sixty.

Figure 7 shows the distribution of the expected welfare cost of late-life depression (compensating variation) and the approximate expected monetary equivalent. As shown in Table 2, ignoring all spillovers (depression only model) yields a lower bound on the average welfare cost of 8.4% of annual consumption. The first panel of Figure 7 shows that most of the lower bound distribution falls under 10%, though there is a thin right tail suggesting substantially higher costs for a select few. Likewise, the second panel shows this lower bound translates into an expected loss in lifetime consumption of under $100,000 for most individuals in the sample. When adjusting for potential spillover effects of depression on comorbidities, mortality, leisure, and consumption, there is a substantial increase in the mean and inequality of welfare costs. For example, there is now a substantial portion of the distribution willing to pay over 20% of their annual consumption to avoid the possibility of late-life depression.

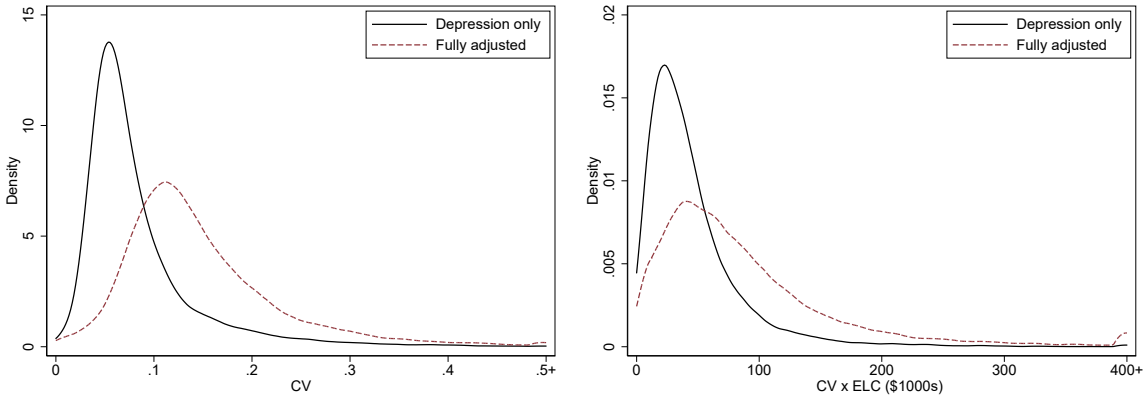

Figure 7: Expected welfare loss after age sixty

Notes: Estimates use base year respondent analysis weights. ELC denotes expected lifetime consumption ($1000s).

3.4 Cost over birth cohorts

Our analysis so far has focused on results only in the EHRS birth cohort. We now compare our estimated welfare costs across the four cohorts in our simulation sample. We begin by examining the predicted mean CESD depression score by age and cohort (see Figure 8). In contrast to each of the younger cohorts, simulations suggest that the EHRS experienced a rising mean CESD score during their sixties. However, based on currently available data, our model predicts that all cohorts have realized or will realize a rising mean CESD score over much of their seventies and eighties. In general, these trends are consistent with the U-shaped pattern over age found in previous studies (Mirowsky and Ross, 1992; Sutin et al., 2013; Tampubolon and Maharani, 2017; Abrams and Mehta, 2019). After the late-eighties, CESD scores are predicted to fall for all cohorts. In terms of levels, the model predicts that after their mid-sixties, mean depression scores will be lower among War Babies and Baby Boomers than the early or late HRS cohorts.

Table 3 reports the estimated costs of late-life depression by cohort. When only considering the direct health utility penalty of depression (depression only model), the expected loss in QALYs is slightly smaller for the younger cohorts. For example, the average expected QALYs lost falls from 0.85 for the EHRS cohort to 0.81 for Baby Boomers. Likewise, the willingness-to-pay to avoid depression falls from 8.4% to 7.5% of annual consumption. A similar general pattern of falling average costs of depression over cohorts remains when adjusting for spillover effects on other modeled outcomes (full model). For example, the expected QALYs lost falls from 2.06 to 1.93 between the EHRS and Baby Boomer cohorts. Similarly, the expected annual consumption gain from depression falls from $631 to $497. There is also a slightly higher gain in labor supply for younger cohorts. In terms of our CV welfare metric, fully-adjusted willingness-to-pay falls from 14.8% to 12.8% of annual consumption. This amounts to

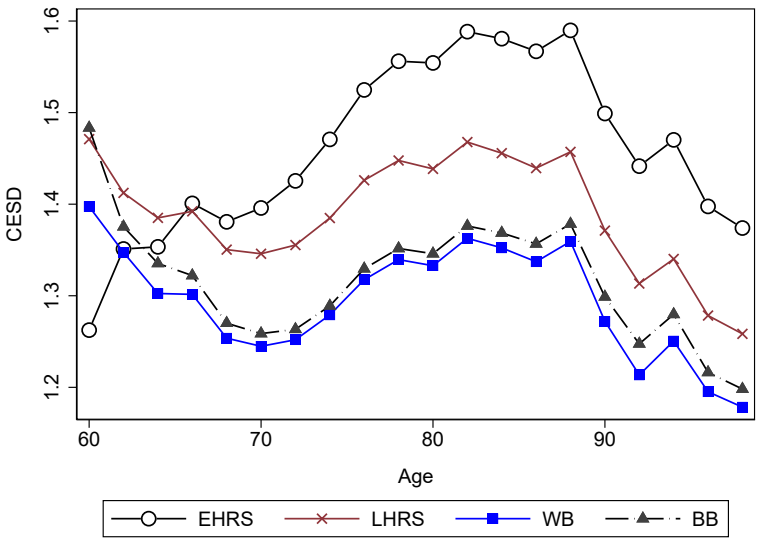

Figure 8: Expected CESD by age and cohort

Notes: Expected CESD score (0-8 scale) conditional on survival.

a fall in the average expected loss of lifetime consumption from $89,405 in the EHRS cohort to $77,338 among Baby Boomers.

Table 3: Mean costs of depression after age sixty by birth cohort

	Depression only				Full model			
	EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB
Expected loss								
QALYs	0.853	0.825	0.791	0.813	2.064	1.975	1.889	1.929
Labor supply (yrs)					0.095	0.097	0.097	0.112
Consumption (annual)					-0.631	-0.615	-0.558	-0.497
CV (λ)	0.084	0.083	0.077	0.075	0.148	0.144	0.134	0.128
λ×ELC	45.933	46.630	44.370	40.131	89.405	90.406	85.584	77.338
CV Gini	0.337	0.359	0.381	0.419	0.280	0.293	0.316	0.347
CV P90	0.154	0.161	0.155	0.163	0.250	0.252	0.246	0.247

Notes: Estimates use base year respondent analysis weights. ELC denotes expected lifetime consumption. Consumption in $1000s.

While we estimate falling *average* costs of depression over cohorts, the final two row of Table 3 reveals another important trend. When looking at the distribution within a cohort, we see that the inequality of depression costs is rising. For example, the Gini coefficient on our fully-adjusted CV measure of welfare has increased from 0.28 in the EHRS cohort to 0.34 for Baby Boomers. Thus, while the average cost has decreased slightly across cohorts, the welfare burden of depression has become significantly more unequal *within* cohorts. The final row sheds further light on these changes by showing

that the CV measure at the ninetieth percentile of the distribution has largely held steady across cohorts. In other words, falling costs of depression over cohorts have not been realized by the most severely affected, even though some improvements appear to have accrued overall.

One potential explanation for the falling average cost of depression over cohorts is the general rise in the use of depression treatments. For example, the share of the U.S. adult population using antidepressants increased from 7.7% in 1999 to 12.7% in 2014 (Pratt et al., 2017). This rise in treatment usage (or the introduction of more effective treatments) may have improved symptoms associated with moderate depression while remaining ineffective against more severe depression, explaining the patterns observed across cohort in our estimates. Alternatively, it could be that antidepressant treatments are improving severe depression as well, but prevalence of severe depression is simultaneously increasing at a comparable rate, hence leaving the welfare costs relatively constant across cohorts at the top end of the distribution.

3.5 Robustness

We examined the robustness of our key welfare numbers under several alternate modeling assumption. The online appendix provides robustness results for preference parameters $(\beta, \epsilon, \theta)$ and using alternate health utility weights. Here we examine the robustness of results to a more general form of flow utility for consumption and leisure given by:

$$u(c,l,h) = \phi(h)\left[\frac{c^{1-\gamma}}{1-\gamma}\left(1-(1-\gamma)\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}\right)^{\gamma} - \frac{\bar{c}^{1-\gamma}}{1-\gamma}\right] \tag{9}$$

which reduces to our benchmark case with $\gamma = 1$ and $\bar{c} = 3$.[19] With $\gamma > 1$ there is more curvature over consumption. These preferences follow those proposed by Trabandt and Uhlig (2011) and Jones and Klenow (2016) which maintain a constant Frisch elasticity of labor supply. We check robustness of results for curvatures up to $\gamma = 3.5$ as there have been a wide range of empirical estimates, with large curvatures arguably more plausible at older ages. Mean welfare results are reported in Table 4.

The first row of Table 4 provides our benchmark results for easy comparison. The second row examines the impact of assuming a higher value for the flow intercept $\bar{c}$. Specifically, we set $\bar{c} = 1.5$, implying that \$1,500 of consumption is needed for a retiree to prefer life to death. This increases the median value of life to about \$66,000 per QALY in the EHRS cohort. It also increases the estimated mean welfare cost of late-life depression in the EHRS from 8-15% to 11-21% of annual consumption. There are similar increases in welfare estimates for later cohorts and we continue to see a small decline in mean welfare costs across cohorts.

The remaining rows of Table 4 provides results with higher curvature values over consumption. As discussed by Murphy and Topel (2006), one problem that arises with

[19] $\bar{c} = 3$ is equivalent to $\bar{u} = -log(3)$ as in our benchmark when $\gamma = 1$.

Table 4: Selected robustness results

γ	$\bar{c}$	VOL	Depression only				Full model			
			EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB
1.0	3.00	46.81	0.084	0.083	0.077	0.075	0.148	0.144	0.134	0.128
1.0	1.50	66.06	0.119	0.117	0.110	0.110	0.212	0.207	0.194	0.190
1.5	3.00	72.87	0.139	0.140	0.132	0.122	0.254	0.253	0.238	0.220
2.0	3.00	104.10	0.209	0.212	0.199	0.180	0.341	0.342	0.324	0.295
3.0	3.00	211.91	0.344	0.354	0.336	0.297	0.457	0.466	0.447	0.403
3.5	3.00	299.34	0.384	0.398	0.380	0.335	0.482	0.494	0.476	0.428
1.5	4.84	46.81	0.094	0.095	0.089	0.080	0.178	0.177	0.167	0.149
2.0	5.49	46.81	0.115	0.118	0.110	0.095	0.204	0.206	0.193	0.168
3.0	5.97	46.81	0.165	0.174	0.162	0.135	0.248	0.257	0.242	0.204
3.5	6.01	46.81	0.186	0.198	0.185	0.152	0.260	0.273	0.257	0.216

Notes: Median value of life per QALY (in thousands of dollars) for EHRS cohort denoted by VOL. All other columns report mean CV (λ). Estimates use base year respondent analysis weights. War Babies denoted by WB and Baby Boomers by BB.

higher curvature in this framework is that as γ rises, the implied value of life grows rapidly. In order to gain a sense of this issue, the first column in Table 4 shows the median value of life per QALY with higher curvatures. With $\gamma = 2$ and $\bar{c}$ held at its benchmark value, the median value of life is high but not completely implausible at \$104,000 per QALY. The bound on the estimated mean welfare cost of late-life depression in the EHRS rises to 20-34% of annual consumption. When $\gamma = 3.5$, the value of life reaches about \$300,000 per QALY and the welfare bound reaches 38-48%. Only three out 23 value of life studies surveyed by Ryen and Svensson (2015) estimated a mean value of life over \$150,000. The likely overstated value of life at higher curvatures suggests caution should be taken when interpreting robustness results with high (but empirically plausible) curvature values. In light of this concern, the final four rows show results at the same curvature values but with intercept $\bar{c}$ adjusted to maintain the same median value of life as our benchmark estimates. These results give us a sense of how sensitive welfare results are to higher values of life versus risk aversion to consumption fluctuations. For example, with $\gamma = 3.5$, the intercept rises to about \$6,000 and the welfare cost of late-life depression in the EHRS is about 18-26%. This is still larger than the benchmark but significantly smaller than welfare estimates with higher curvature and constant intercept. Overall, the higher curvature values provide a sense of the robustness of key results and the conservative nature of our benchmark welfare estimates.

3.6 Policy simulations

Finally, in this section we conduct several additional counterfactual simulations to gain a sense of policy lessons we might learn from our results. First, we estimate the impact of setting the CESD score to zero at age sixty but allowing depression to emerge after

age sixty. This gives a sense of how results are driven by initial (age sixty) conditions compared to the development of depressive symptoms after age sixty. Table 5 provides welfare results over cohorts for this experiment. As expected, the welfare cost of depression at age sixty is lower than our benchmark which captures the cost from age sixty onward. In the EHRS cohort, for example, the fully-adjusted willingness-to-pay is 12.2% compared to 14.8% in the benchmark. So while the costs are lower, depressive symptoms at age sixty still explain a large share of the total costs. Comparing to the benchmark numbers across cohorts, age sixty depression explains roughly 75-80% of the total welfare costs of late-life depression. This suggests that targeting depression interventions earlier in the life-cycle may have substantial benefits later in life. Of course, this does not imply that depression cannot be addressed at older ages, only that early interventions could have significant dynamic benefits later in life.

Table 5: Mean welfare costs of depression: additional counterfactuals

	Depression only				Full model			
	EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB
Benchmark	0.084	0.083	0.077	0.075	0.148	0.144	0.134	0.128
Age sixty only	0.065	0.063	0.059	0.057	0.122	0.116	0.108	0.103
Clinical only	0.047	0.046	0.042	0.043	0.078	0.076	0.069	0.067

Notes: Estimates use base year respondent analysis weights. All results are mean CV (λ). Age sixty only experiment eliminates depression at sxity but allows depression to emerge after age sixty. Clinical only experiment estimates costs of a CESD score over three.

It is also useful to consider how far policy could realistically go towards mitigating the costs of late-life depression through promoting diagnosis and treatment. Our results so far have concerned the costs associated with reporting any symptoms on the CESD depression scale. In practice, low level symptoms may be very hard to diagnose and/or treat and could to some extent even reflect noise in the data. It could be argued that broader population level policies could potentially address low level depressive symptoms, for example through promoting general health and social interactions among older populations. Nonetheless, narrowing in on clinically-relevant depressive symptoms may give a better sense of the costs of depression that could realistically be addressed with direct diagnosis and treatment policy interventions. The final row of Table 5 provides results when we ignore the costs of low level symptoms by setting the utility penalty to zero for CESD scores under four. We chose a cutoff of four as it roughly corresponds to clinically-relevant symptoms of depression or "caseness" in the HRS (Steffick, 2000). In all cohorts, clinically-relevant depressive symptoms account for 50-60% of the total cost of late-life depression. For example, based on the full model, the EHRS cohort would be willing to pay up to 7.6% of annual consumption over remaining life to avoid any possibility of clinically-relevant depression (compared to 14.4% to avoid all depressive symptoms). So while low level symptoms are clearly important, the costs associated with clinically-relevant depression suggests substantial scope for direct policy intervention. Of course, it is worth noting that treatment

inventions do not guarantee the mitigation of depression. For example, although antidepressant medications can be effective in treating major depressive disorder, they fail to achieve remission in roughly a third of patients (Souery et al., 2006; Voineskos et al., 2020). Optimal strategies to deal with such treatment resistant depression is an active area of research, and our results suggest substantial potential welfare gains through continuing to improve treatment outcomes.

4 Conclusion

We estimated a panel VAR model of mental and physical health, labor supply, and consumption using longitudinal data from the Health and Retirement Study. We used the estimated model to repeatedly simulate life–cycle paths for older Americans, with and without the prevalence of late–life depression, given a set of initial age sixty conditions. We estimated an average loss of labor supply of up to 1.1 months, lifetime consumption of up to $16,000, and quality–adjusted life expectancy of between 0.85 and 2.1 years per person in the EHRS birth cohort. Combining into a single welfare metric, we estimated a bound on the expected welfare loss of depression of 8-15% of annual consumption after age sixty. This amounts to an expected loss in lifetime consumption (or expected willingness-to-pay to avoid late-life depression) of approximately $46,000-$91,000 per person. In a hypothetical world populated by identical cohorts of size four million at age sixty, this produces a back-of-the-envelope estimate of aggregate welfare loss on the order of $180-360 billion annually. We also found substantial heterogeneity in the estimated cost with some individuals willing to give up well over 20% of annual consumption to avoid late-life depression. Moreover, while we found a small general decline in average costs over birth cohorts, the welfare burden of depression appears to have become significantly more unequal within cohorts.

From a policy perspective, our results suggest there are substantial potential benefits from reducing depression among older Americans. If enacted early enough, such policies could be workplace based. For example, consistent with broader recommendations of the CDC, there could be customized and intensive employee assistance programs (EAPs) for pre-retirees that include employee classes or seminars. A stronger focus could be put on information and referral services for employees with symptoms of depression, and the EAPs could be responsive to events, stressors, and changes in the lives of employees. Work or community based health promotion programs for older adults such as physical activity could also be used as adjunct therapies in preventing and treating depression. From the standpoint of the state, there could be more active steps to address depression and major depressive disorders among older adults by promoting public awareness, prescriptions and therapies to treat depression, and treatment centers. States could legislate and execute mental health crisis intervention training and other evidence-based programs for people most likely to encounter older adults on the job. In addition, states could subsidize the cost of mental health screenings, promote "telepsychology" (providing mental health services via remote technology),

and offer incentives for providers to screen retired individuals. States could also review programs such as Medicaid to ensure that older adults living on fixed incomes have access to the drugs they need to treat their depression in the most effective manner. States could also take a lead from the federal government and require depression and other mental health conditions to be treated and covered at the same level as physical health conditions as is the case with the Affordable Care Act.

This study is not without limitations. Our estimates only include private costs of depression and do not capture public expenses (e.g., Medicare costs) or general equilibrium effects. We also do not capture potential spillover effects. For example, depression could have substantial costs for a partner or children. On the other hand, we have included lost lifetime consumption in our upper bound cost estimate. But if some of this consumption is shifted to children or charities through higher bequests, some of the cost may be recovered by society at large. Moreover, while our statistical model does well in replicating the observed patterns in the data, point estimates cannot be viewed as necessarily causal nor as adhering to any particular unobserved mechanism. This leaves us with only an estimated bound on feasible costs. Our compensating variation measure is also quantitatively sensitive to the choice of curvature in the utility function. Nonetheless, this study's novelty is in estimation of a more comprehensive measure that incorporates life-cycle dynamics to improve our understanding of the welfare costs of late-life depression.

References

Abrams, L. R. and Mehta, N. K. (2019). Changes in depressive symptoms over age among older americans: differences by gender, race/ethnicity, education, and birth cohort. *SSM-population health*, 7:100399.

Ambresin, G., Chondros, P., Dowrick, C., Herrman, H., and Gunn, J. M. (2014). Self-rated health and long-term prognosis of depression. *The Annals of Family Medicine*, 12(1):57–65.

Barry, L. C., Abou, J. J., Simen, A. A., and Gill, T. M. (2012). Under-treatment of depression in older persons. *Journal of affective disorders*, 136(3):789–796.

Birnbaum, H. G., Kessler, R. C., Kelley, D., Ben-Hamadi, R., Joish, V. N., and Greenberg, P. E. (2010). Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. *Depression and anxiety*, 27(1):78–89.

Blazer, D. G. (2003). Depression in late life: review and commentary. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(3):M249–M265.

Bruce, M. L. (2001). Depression and disability in late life: directions for future research. *The American Journal of geriatric psychiatry*, 9(2):102–112.

Bruce, M. L., Leaf, P. J., Rozal, G. P. M., Florio, L., and Hoff, R. A. (1994). Psychiatric status and 9-year mortality data in the new haven epidemiologic catchment area study. *The American journal of psychiatry*.

Chiang, H.-H., Guo, H.-R., Livneh, H., Lu, M.-C., Yen, M.-L., and Tsai, T.-Y. (2015). Increased risk of progression to dialysis or death in ckd patients with depressive symptoms: A prospective 3-year follow-up cohort study. *Journal of psychosomatic research*, 79(3):228–232.

Chu, W., Chang, S.-F., Ho, H.-Y., and Lin, H.-C. (2019). The relationship between depression and frailty in community-dwelling older people: a systematic review and meta-analysis of 84,351 older adults. *Journal of Nursing Scholarship*, 51(5):547–559.

Chui, H., Gerstorf, D., Hoppmann, C. A., and Luszcz, M. A. (2015). Trajectories of depressive symptoms in old age: Integrating age-, pathology-, and mortality-related changes. *Psychology and Aging*, 30(4):940.

Ciechanowski, P. S., Katon, W. J., and Russo, J. E. (2000). Depression and Diabetes: Impact of Depressive Symptoms on Adherence, Function, and Costs. *Archives of Internal Medicine*, 160(21):3278–3285.

Cinciripini, P. M., Wetter, D. W., Fouladi, R. T., Blalock, J. A., Carter, B. L., Cinciripini, L. G., and Baile, W. F. (2003). The effects of depressed mood on smoking cessation: mediation by postcessation self-efficacy. *Journal of consulting and clinical psychology*, 71(2):292.

Cronin-Stubbs, D., De Leon, C. F. M., Beckett, L. A., Field, T. S., Glynn, R. J., and Evans, D. A. (2000). Six-year effect of depressive symptoms on the course of physical disability in community-living older adults. *Archives of internal medicine*, 160(20):3074–3080.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of Labor Economics*, 3:3309–3416.

De Groot, M., Anderson, R., Freedland, K. E., Clouse, R. E., and Lustman, P. J. (2001). Association of depression and diabetes complications: a meta-analysis. *Psychosomatic medicine*, 63(4):619–630.

Dieleman, J. L., Cao, J., Chapin, A., Chen, C., Li, Z., Liu, A., Horst, C., Kaldjian, A., Matyasz, T., Scott, K. W., et al. (2020). Us health care spending by payer and health condition, 1996-2016. *Jama*, 323(9):863–884.

Doshi, J. A., Cen, L., and Polsky, D. (2008). Depression and retirement in late middle-aged us workers. *Health services research*, 43(2):693–713.

Dowson, C., Laing, R., Barraclough, R., Town, I., Mulder, R., Norris, K., and Drennan, C. (2001). The use of the hospital anxiety and depression scale (hads) in patients with chronic obstructive pulmonary disease: a pilot study. *New Zealand medical journal*, 114(1141):447.

Fechner-Bates, S., Coyne, J. C., and Schwenk, T. L. (1994). The relationship of self-reported distress to depressive disorders and other psychopathology. *Journal of consulting and clinical psychology*, 62(3):550.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., and Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care*, 40(2):113–128.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., DePauw, S., Zhu, Z., Denton, M., Boyle, M., et al. (1998). Multiplicative multi-attribute utility function for the health utilities index mark 3 (hui3) system: a technical report. Technical report, Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada.

Geerlings, S., Beekman, A., Deeg, D., Twisk, J., and Van Tilburg, W. (2001). The longitudinal effect of depression on functional limitations and disability in older adults: an eight-wave prospective community-based study. *Psychological Medicine*, 31(8):1361.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., and Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the united states (2005 and 2010). *The Journal of clinical psychiatry*, 76(2):155–162.

Hall, R. E. (2009). Reconciling cyclical movements in the marginal value of time and the marginal product of labor. *Journal of Political Economy*, 117(2):281–323.

Holford, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual review of public health*, 12(1):425–457.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Horsman, J., Furlong, W., Feeny, D., and Torrance, G. (2003). The health utilities index (hui®): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, 1(1):54.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, pages 21–37.

Jones, C. I. and Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *The American Economic Review*, 106(9):2426–2457.

Kaplan, R. M. and Bush, J. W. (1982). Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychology*, 1(1):61.

Karim, J., Weisz, R., Bibi, Z., and ur Rehman, S. (2015). Validation of the eight-item center for epidemiologic studies depression scale (ces-d) among older adults. *Current Psychology*, 34(4):681–692.

Kazama, M., Kondo, N., Suzuki, K., Minai, J., Imai, H., and Yamagata, Z. (2011). Early impact of depression symptoms on the decline in activities of daily living among older japanese: Y-hale cohort study. *Environmental health and preventive medicine*, 16(3):196–201.

Kinnunen, T., Doherty, K., Militello, F. S., and Garvey, A. J. (1996). Depression and smoking cessation: characteristics of depressed smokers and effects of nicotine replacement. *Journal of consulting and clinical psychology*, 64(4):791.

Kivelá, S.-L. and Pahkala, K. (2001). Depressive disorder as a predictor of physical disability in old age. *Journal of the American Geriatrics Society*, 49(3):290–296.

Kok, R. M. and Reynolds, C. F. (2017). Management of depression in older adults: a review. *Jama*, 317(20):2114–2122.

Kunik, M. E., Roundy, K., Veazey, C., Souchek, J., Richardson, P., Wray, N. P., and Stanley, M. A. (2005). Surprisingly high prevalence of anxiety and depression in chronic breathing disorders. *Chest*, 127(4):1205–1211.

Lakey, S. L., LaCroix, A. Z., Gray, S. L., Borson, S., Williams, C. D., Calhoun, D., Goveas, J. S., Smoller, J. W., Ockene, J. K., Masaki, K. H., et al. (2012). Antidepressant use, depressive symptoms, and incident frailty in women aged 65 and older from the women's health initiative observational study. *Journal of the American Geriatrics Society*, 60(5):854–861.

Lerner, D. and Henke, R. M. (2008). What does research tell us about depression, job performance, and work productivity? *Journal of occupational and environmental medicine*, 50(4):401–410.

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., and Allen, N. B. (1997). Center for epidemiologic studies depression scale (ces-d) as a screening instrument for depression among community-residing older adults. *Psychology and aging*, 12(2):277.

Luppa, M., Sikorski, C., Motzek, T., Konnopka, A., Konig, H.-H., and G Riedel-Heller, S. (2012). Health service utilization and costs of depressive symptoms in late life-a systematic review. *Current pharmaceutical design*, 18(36):5936–5957.

Lwin, M. N., Serhal, L., Holroyd, C., and Edwards, C. J. (2020). Rheumatoid arthritis: The impact of mental health on disease: A narrative review. *Rheumatology and therapy*, 7(3):457–471.

Mayerl, H., Stolz, E., and Freidl, W. (2020). Frailty and depression: Reciprocal influences or common causes? *Social Science & Medicine*, 263:113273.

McEwen, B. S. (2003). Mood disorders and allostatic load. *Biological psychiatry*, 54(3):200–207.

Mikkelsen, R. L., Middelboe, T., Pisinger, C., and Stage, K. B. (2004). Anxiety and depression in patients with chronic obstructive pulmonary disease (copd). a review. *Nordic journal of psychiatry*, 58(1):65–70.

Miller, R. and Bairoliya, N. (2022). Health, longevity, and welfare inequality of older americans. *(Forthcoming) The Review of Economics and Statistics*.

Mirowsky, J. and Ross, C. E. (1992). Age and depression. *Journal of health and social behavior*, pages 187–205.

Mohiuddin, S. and Payne, K. (2014). Utility values for adults with unipolar depression: systematic review and meta-analysis. *Medical Decision Making*, 34(5):666–685.

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., and Ustun, B. (2007). Depression, chronic diseases, and decrements in health: results from the world health surveys. *The Lancet*, 370(9590):851–858.

Murphy, K. M. and Topel, R. H. (2006). The value of health and longevity. *Journal of political Economy*, 114(5):871–904.

Penninx, B. W. (2017). Depression and cardiovascular disease: epidemiological evidence on their linking mechanisms. *Neuroscience & Biobehavioral Reviews*, 74:277–286.

Penninx, B. W., Leveille, S., Ferrucci, L., Van Eijk, J. T., and Guralnik, J. M. (1999). Exploring the effect of depression on physical disability: longitudinal evidence from the established populations for epidemiologic studies of the elderly. *American journal of public health*, 89(9):1346–1352.

Pratt, L. A., Brody, D. J., and Gu, Q. (2017). Antidepressant use among persons aged 12 and over: United states, 2011-2014. nchs data brief. number 283. *National Center for Health Statistics*.

Rice, N. E., Lang, I. A., Henley, W., and Melzer, D. (2011). Common health predictors of early retirement: findings from the english longitudinal study of ageing. *Age and ageing*, 40(1):54–61.

Rothermund, K. and Brandtstädter, J. (2003). Depression in later life: cross-sequential patterns and possible determinants. *Psychology and Aging*, 18(1):80.

Rubio-Guerra, A. F., Rodriguez-Lopez, L., Vargas-Ayala, G., Huerta-Ramirez, S., Serna, D. C., and Lozano-Nuevo, J. J. (2013). Depression increases the risk for uncontrolled hypertension. *Experimental and clinical cardiology*, 18(1):10–2.

Ryen, L. and Svensson, M. (2015). The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Economics*, 24(10):1289–1301.

Ryerson, C. J., Berkeley, J., Carrieri-Kohlman, V. L., Pantilat, S. Z., Landefeld, C. S., and Collard, H. R. (2011). Depression and functional status are strongly associated with dyspnea in interstitial lung disease. *Chest*, 139(3):609–616.

Sivertsen, H., Bjørkløf, G. H., Engedal, K., Selbæk, G., and Helvik, A.-S. (2015). Depression and quality of life in older persons: a review. *Dementia and geriatric cognitive disorders*, 40(5-6):311–339.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 13(2):144.

Sodhi, J. and Al Snih, S. (2020). Effects of pain and depression on adl disability over 6 years of follow-up among older adult americans. *Innovation in Aging*, 4(Suppl 1):370.

Souery, D., Papakostas, G. I., Trivedi, M. H., et al. (2006). Treatment-resistant depression. *Journal of Clinical Psychiatry*, 67:16.

Soysal, P., Veronese, N., Thompson, T., Kahl, K. G., Fernandes, B. S., Prina, A. M., Solmi, M., Schofield, P., Koyanagi, A., Tseng, P.-T., et al. (2017). Relationship between depression and frailty in older adults: A systematic review and meta-analysis. *Ageing research reviews*, 36:78–87.

Steffick, D. E. (2000). Documentation of affective functioning measures in the health and retirement study. *HRS/AHEAD Documentation Report*, DR-005.

Stenholm, S., Pentti, J., Kawachi, I., Westerlund, H., Kivimäki, M., and Vahtera, J. (2014). Self-rated health in the last 12 years of life compared to matched surviving controls: the health and retirement study. *PloS one*, 9(9):e107879.

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R., and Morganstein, D. (2003). Cost of lost productive work time among us workers with depression. *Jama*, 289(23):3135–3144.

Stuck, A. E., Walthert, J. M., Nikolaus, T., Büla, C. J., Hohmann, C., and Beck, J. C. (1999). Risk factors for functional status decline in community-living elderly people: a systematic literature review. *Social science & medicine*, 48(4):445–469.

Sutin, A. R., Terracciano, A., Milaneschi, Y., An, Y., Ferrucci, L., and Zonderman, A. B. (2013). The trajectory of depressive symptoms across the adult life span. *JAMA psychiatry*, 70(8):803–811.

Tampubolon, G. and Maharani, A. (2017). When did old age stop being depressing? depression trajectories of older americans and britons 2002–2012. *The American Journal of Geriatric Psychiatry*, 25(11):1187–1195.

Trabandt, M. and Uhlig, H. (2011). The Laffer curve revisited. *Journal of Monetary Economics*, 58(4):305–327.

Turvey, C. L., Wallace, R. B., and Herzog, R. (1999). A revised ces-d measure of depressive symptoms and a dsm-based measure of major depressive episodes in the elderly. *International psychogeriatrics*, 11(2):139–148.

Van Manen, J., Bindels, P., Dekker, F., IJzermans, C., Van der Zee, J., and Schade, E. (2002). Risk of depression in patients with chronic obstructive pulmonary disease and its determinants. *Thorax*, 57(5):412–416.

Vaughan, L., Corbin, A. L., and Goveas, J. S. (2015). Depression and frailty in later life: a systematic review. *Clinical interventions in aging*, 10:1947.

Voineskos, D., Daskalakis, Z. J., and Blumberger, D. M. (2020). Management of treatment-resistant depression: challenges and strategies. *Neuropsychiatric disease and treatment*, 16:221.

Wang, P. S., Simon, G., and Kessler, R. C. (2003). The economic burden of depression and the cost-effectiveness of treatment. *International Journal of Methods in Psychiatric Research*, 12(1):22–33.

Wilson, I. (2006). Depression in the patient with copd. *International journal of chronic obstructive pulmonary disease*, 1(1):61.

Online Appendix for: "The Welfare Cost of Late-life Depression"

Contents

List of Tables

List of Figures

A Imputation of consumption and other missing data

The CAMS collected consumption data for approximately 20% of the HRS sample starting from 2001. In order to estimate our dynamic panel models and construct simulated life-cycle paths for the remaining sample, we follow Miller et al. (2019) and Miller and Bairoliya (2022) and multiply impute the missing consumption data. We use the computationally attractive EM-bootstrapping algorithm allowing for cross-sectional time-series data proposed by Honaker and King (2010) and implemented through the freely available Amelia II software program (Honaker et al., 2011). This approach provides m separate complete datasets in which all analyses are conducted independently. Results are then combined into a single estimate.[1] We follow Miller and Bairoliya (2022) and set $m = 12$.

There are two primary assumptions underlying the proposed imputation method. First, the complete data is assumed to be multivariate normal. While this may seem somewhat restrictive, it has been shown that multivariate normal imputation models provide an adequate approximation to the true underlying distribution in a variety of settings, even in the presence of categorical or mixed data (Schafer, 1997). Second is the standard required assumption that data is missing at random (MAR)—any nonrandom pattern of missingness can be accounted for by the observed data included in the model. Note this is less restrictive than the requirement data be missing completely at random (MCAR). In practice, we know that missing data is not at random, at least for years falling outside of the CAMS window (1992-1998 and 2016). However, by including a rich set of related covariates in the imputation model, we argue that missing data can be treated as MAR in the statistical sense. While there is no way to empirically test this assumption, we run a number of diagnostic tests to check the credibility of the imputation model in search of any obvious deficiencies.

Variables from the RAND HRS Longitudinal File 2016 (V2) included in our imputation model are number of household members (HHRES), age (AGEY_E), aged squared, cubed root of total wealth (ATOTA), log household income (ITOT)[2], and dummy indicators for cohort (COHBYR), labor force status (LBRF), gender (RAGENDER), race (RARACEM), education (RAEDUC), marital status (MSTAT), census division (CENDIV), 1980 census occupation code for longest reported tenure (JLOCC)[3], CESD score (CESD), self-reported health (SHLT), ADLs (ADLA), and eight doctor diagnosed health conditions (HIBPE, DIABE, CANCRE, LUNGE, HEARTE, STROKE, PSYCHE, ARTHRE). The model also included our constructed indicator for retirement and hours worked. In order to allow for the time-series structure of the data, lags and leads of consumption, wealth, income, and hours worked are included in the imputation model. While we are primarily imputing consumption data, Amelia II also provides imputed values for all other missing variables included in the model.[4]

A useful check of the viability of the imputation model is to compare the distributions of the imputed values against the observed data. While there is no need for these distributions to be the

[1] Assuming asymptotically normally distributed statistics implies a simple average across datasets (Rubin, 2004).

[2] Transformed wealth bounded at values of -4 and 16 and log income bounded above at a value of 6. Both variables in thousands of 2010 dollars.

[3] We treat missing and "other" occupations as one category.

[4] If the observed data used in the imputation model has a poorly behaved likelihood, the convergence of the EM algorithm could be sensitive to the staring values chosen. We found no evidence of local convergence issues using the overdispersed start values diagnostic test proposed by Honaker et al. (2011).

same, the comparison gives a sense of the plausibility of imputations (Honaker et al., 2011). Figure 1 plots the density of observed and imputed values of consumption. The imputed values are taken as the mean across the m imputed datasets. The comparison suggests no unusual pattern in the distribution of imputed values, providing cursory support of model plausibility.

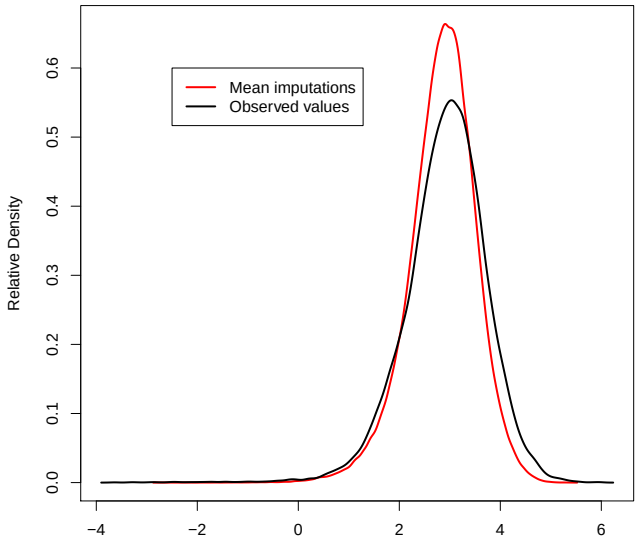

Figure 1: Distributions of observed and imputed values of consumption

Another diagnostic tool proposed by Honaker et al. (2011) is *overimputing*. While it is impossible to examine if the imputed values are close to the missing values they are attempting to recover, *observed* values can be used to test the accuracy of the imputation process. Overimputing sequentially treats each of the observed consumption values as if they were missing and then imputes their values several hundred times. This provides a mean imputed value and confidence interval that can be compared to the actual observed data. Figure 2 plots all observed consumption values against the mean of their imputed values and the associated 95% confidence interval. A visual inspection of the diagnostic plot suggests the model does fairly well predicting values other than the lowest values. However, few individuals lie in this extreme end of the distribution—less than 0.3% of the observations fall below zero ($1,000 annual consumption). Honaker et al. (2011) suggest a good imputation model should have at least 90% of the confidence intervals containing the true values (i.e. 90% of the confidence intervals should cross the $y = x$ line). In our case, 94% of the observed values are within the confidence bounds.

As a final examination of the imputation model we try to get a sense of how it predicts missing values in a time series. While it is infeasible to examine the imputed time trends for each individual in the sample, Figure 3 provides time series for a random sub-set of ten individuals with at least one observed consumption value. The mean of the imputed values are plotted in red with 95% confidence bounds (based on 100 imputations). The isolated black points without bounds are observed data. Broadly, the imputed values fall in line with the observed data and no egregious outliers emerge. Note that prior to wave five (2000) and for wave thirteen (2016) all values are

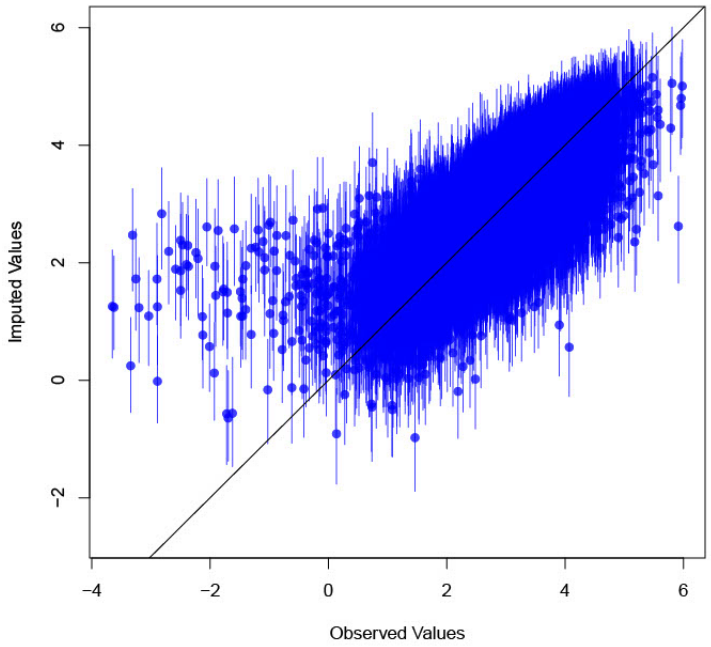

Figure 2: Overimputed values of consumption

imputed as these waves are outside of our CAMS data window.

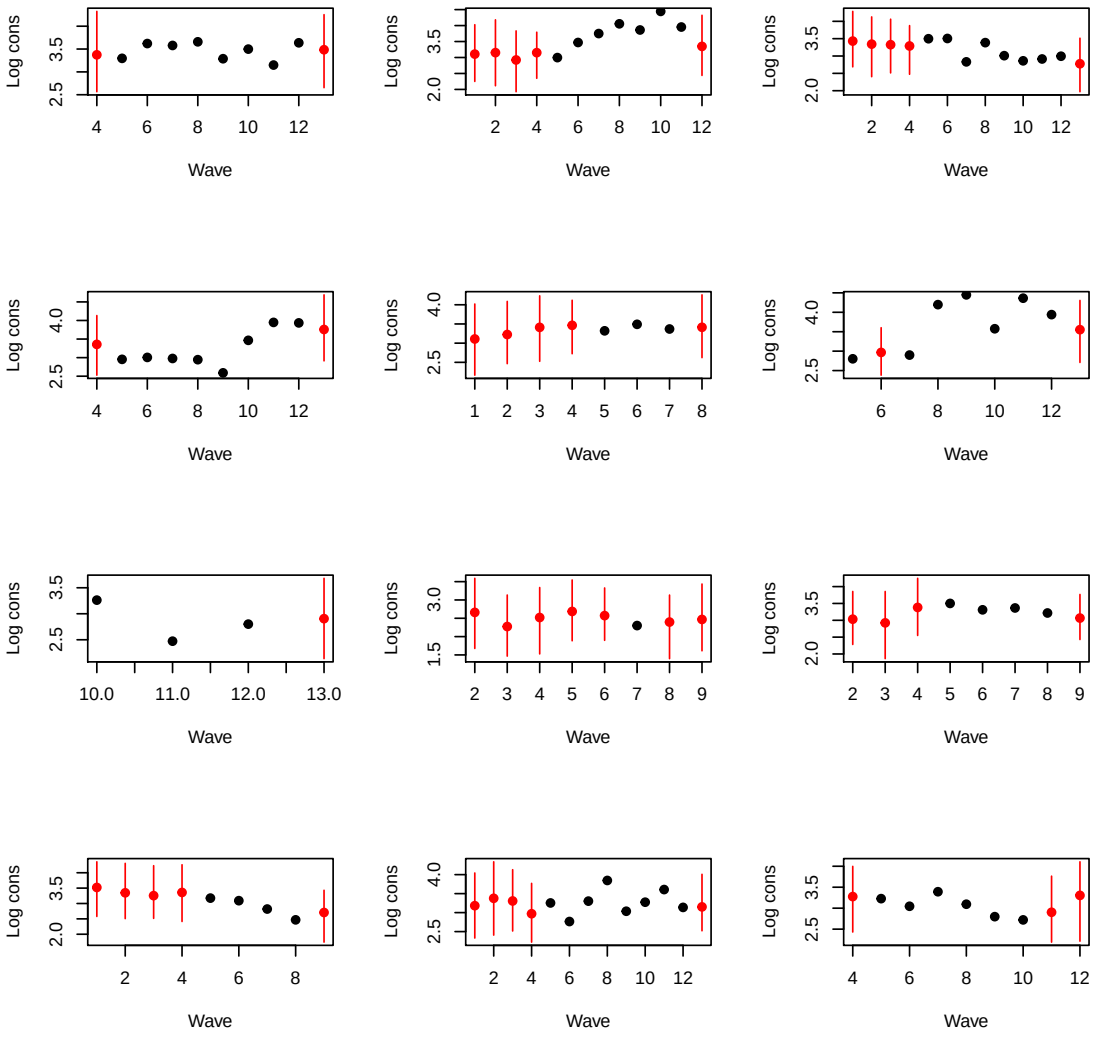

Figure 3: Observed and imputed consumption over time for a random sub-sample

B Simulation model

In this appendix we provide additional detail of our estimation and simulation procedures. For additional applications of this framework see Miller et al. (2019); Miller and Bairoliya (2022). As the HRS is collected biennially, a model period corresponds to two calendar years and individuals are grouped in two-year age intervals.

B.1 Higher order lags

Including additional outcome lags may be necessary to ensure there is no autocorrelation in the structural error terms of the system. The VAR(1) model extends easily to higher orders. For example, a VAR(2) version of our model takes the following form:

$$AY_{it} = BY_{it-1} + DY_{it-2} + CX_{it} + \varepsilon_{it},$$

with the block matrix form of DY_{it-2} given by:

$$\left[\begin{array}{c|cccc} D_{11} & D_{12} & D_{13} & D_{14} & D_{15} \\ \hline D_{21} & d_{22} & d_{23} & d_{24} & d_{25} \\ D_{31} & d_{32} & d_{33} & d_{34} & d_{35} \\ D_{41} & d_{42} & d_{43} & d_{44} & d_{45} \\ D_{51} & d_{52} & d_{53} & d_{54} & d_{55} \end{array}\right] \left[\begin{array}{c} M_{it-2} \\ \hline s_{it-2} \\ d_{it-1} \\ r_{it-2} \\ c_{it-2} \end{array}\right].$$

Here, for example, coefficient vector D_{31} allows the second lag of the morbidity state vector to directly affect current depression. Note that it is not strictly required that the number of lags included be identical for each outcome. For example, excluding the second lag of self-rated health on consumption simply implies setting $d_{52} = 0$.

B.2 Estimation

The pooled sample used to estimate the simulation model includes all individuals born prior to 1960 and aged fifty and over at the time of the survey. This consists of 40,708 unique individuals and 238,091 total individual-year observations. Table 1 shows descriptive statistics for modeled outcomes for each cohort in the HRS. Prevalence of depressive symptoms was substantial among respondents, allowing for relatively precise estimates of their effects on dynamic processes. The share of observations reporting no depressive symptoms (CESD = 0) ranged from nearly 48% among War Babies to 33% among the oldest AHEAD cohort. The most severe depression state (CESD = 8) ranged from 1% in the EHRS cohort to 2.44% among mid-Baby Boomers. There was also substantial variation across diagnosed morbidities and reported self-rated health. In terms of labor supply, the share of retired individuals ranged from 34% in the most recent Baby Boomer cohort to 95% in the (much older) AHEAD cohort. Annual real consumption averaged between \$19-\$27,000 across cohorts. Younger cohorts were also more educated and racially diverse.

B.2.1 Methods

As there is no simultaneity across blocks in the system, we follow Miller and Bairoliya (2022) and estimate the model block-by-block. The consumption block is comprised only of equation (6), which is a standard single equation linear dynamic panel data model with lagged dependent variables and individual level fixed effects. The equation is estimated via OLS. We use the bootstrap-based method of Everaert and Pozzi (2007) to correct for the so-called Nickell (1981) bias that is known to arise from OLS estimates of such models.[5] Including a single period lag (two calendar years) of retirement and health on consumption and two lags (four years) for consumption on itself is sufficient to ensure that shocks are serially uncorrelated in the consumption equation.[6]

For consistency with the model of consumption, we use two lags of outcomes in all retirement, depression, health, and survival equations (i.e. we estimate a VAR(2) system and set $K = 2$ in the survival model). The ordered probit modes of self-rated health (3) and depression (4) are each estimated independently of other VAR blocks using maximum likelihood.[7] The retirement equation (5) and mortality equation (7) are estimated independently using standard probit regressions.

This leaves the morbidity block. The morbidity model (2) is structured as a multivariate probit with correlated shocks. Note that this approach does not allow for identification of the variance in structural errors in vector ε_1, but only of the variance in composite errors in vector e. Thus, while this approach is not sufficient for evaluating outcome responses to structural morbidity shocks, identification of composite errors is sufficient for forecasting outcomes as desired in our analysis. We follow Miller and Bairoliya (2022) and estimate this model using a chain of bivariate probit estimators as proposed by Mullahy (2016) due to the large number of outcomes and large number of observations in the HRS. With no additional assumptions, this approach allows for consistent estimation via maximum likelihood as opposed to relying on more computationally intensive simulation based methods. However, a potential estimation issue arises in the morbidity block because morbidity states are absorbing (e.g. *ever* been diagnosed with heart disease). This means, for example, diagnosed heart disease at time t perfectly predicts heart disease at time $t+1$ and we have quasi-complete separation. This implies the effective coefficients on the lagged dependent variables in the morbidity block are infinity (i.e. $\hat{b}_{11}, \hat{b}_{22}, \ldots, \hat{b}_{nn} = \infty$ in system (2)). In a simple univariate probit model, the obvious solution is to condition on not being diagnosed with the morbidity at time t. However, estimation of the bivariate probit involves maximization of the joint likelihood function, so the model is estimated while exogenously constraining infinite coefficients to large values ($\hat{b}_{11}, \hat{b}_{22}, \ldots, \hat{b}_{nn} = 10$), instead of conditioning on time t morbidity status. This restriction to include all observations in the bivariate probit does not effect the likelihood or estimates of the remaining (non-infinite) coefficients. For example, conditioning the bivariate probit

[5] We implement the bootstrap with De Vos et al. (2015) Stata routine $xtbcfe$. We use the deterministic initialization as our benchmark where initial conditions are set equal to those observed.

[6] Following Miller and Bairoliya (2022), we exclude second lags of retirement and health outcomes (included depression) as they were insignificant and noisy. This is equivalent to estimating the VAR(2) system with $D_{51} = d_{52} = d_{53} = d_{54} = 0$. First order autocorrelation was tested for consumption using the approach of Born and Breitung (2016) and implemented in Stata with Wursten et al. (2016). Under the null hypothesis of no autocorrelation, p-values were all greater than 0.22 regardless of imputed dataset used for the test.

[7] Note there is no incidental parameters or initial conditions problem in this case as there is no permanent unobserved heterogeneity or serial correlation in the self-rated health or depression (or retirement and morbidity) model. The standard (ordered) probit estimator is consistent and provides asymptotically valid test statistics and standard errors.

on not having been previously diagnosed with heart disease results in nearly identical estimates for parameters in the heart disease equation as the unconditional bivariate probit with constrained lagged effect.

B.3 Simulations

We use the estimated panel VAR model to construct expected remaining lifetime utility for a subset of sixty year old from the HRS. Note that as the HRS began in 1992, age sixty data is not available for the older AHEAD or CODA cohorts, so these are excluded from our welfare analysis. Moreover, the mid- and late-Baby Boomers were only recently added to the survey and do not have the requisite lagged data to estimate welfare. This leaves four cohorts for welfare analyses—the EHRS, LHRS, War Babies, and early Baby Boomers. Our forecasting model requires lagged outcomes implying data is needed from age fifty-eight as part of age sixty "initial" conditions. However, the oldest respondents in the EHRS cohort were already sixty when first interviewed in 1992, so they are dropped from the simulation sample. Effectively, this drops those born in 1931 from the EHRS and leaves the cohort as those born 1932-1936.

The HRS provides respondent level analysis weights for each wave designed to produce representative cohort samples of the non-institutionalized US population. We use base year weights corresponding to when the cohort is approximately age sixty to examine the welfare distribution. Specifically, we use 1996 analysis weights for the EHRS, 2000 for the LHRS, 2006 for War Babies, and 2008 for Baby Boomers. As any missing data was imputed among respondents (see appendix A), no individuals were dropped from the simulation due to missing item response. However, individuals were dropped if they were not interviewed at ages 58-59 and 60-61.[8] For example, a member of the EHRS cohort interviewed at age 60 in 1996 but missing from the previous survey round would be excluded from the simulation sample (but included in the 2000 nationally representative sample). Table 2 provides a comparison of time invariant characteristics between the weighted representative sample and the sample used in our simulations after dropping these missing cases. The simulation sample is slightly more female, educated, and white relative to the representative sample. However, the differences are small and generally move in same directions for all cohorts.

Table 3 provides a summary of initial outcome conditions in the simulation sample. By most measures, there was an average decline in age sixty health over cohorts. For example, there were some declines in self-rated health, particularly movements from "excellent" to "very good" health. However, morbidities seem to have shifted even more than self-rated health. For example, prevalence of diabetes increased from 11% to 20% and psychiatric problems increased from 7% to 21%. The shift in CESD depression scale is more mixed across cohorts. The EHRS reported the highest percent of respondents with no depressive systems at age sixty (51%) while the LHRS reported to lowest share (44%). On the other hand, more severe depression at sixty showed a generally increasing trend across cohorts, particularly for Baby Boomers. Average age sixty consumption and retired share increased slightly between the EHRS and LHRS cohorts. However, both declined for War Babies and fell even more for Baby Boomers, presumably due to the timing of the great recession, which hit when Baby Boomers were in their late fifties. Increased longevity (and hence

[8] Due to the timing of the interviews across the calendar year, some respondents were 59 in one wave of the survey and 62 in the next. We treat these age 59 data as age 60 data for our simulations.

savings motive) could also potentially explain some of the decline in flow consumption for later cohorts.

B.3.1 Procedure

Using age sixty data as initial $(t = 0)$ conditions[9], we simulate the remaining life outcomes for each individual (i) as follows:

1. Survival shock u_{i1} is drawn and survival to time $t = 1$ (age 62) is determined according to equation (7). If individual survives, move to step two.

2. Morbidity shock vector e_{i1} is drawn from a standard multivariate normal distribution with estimated covariance matrix Σ (see Table 4). This shock vector along with the model outlined in equation (2) is used to compute simulated age 62 morbidity vector M_{i1}.

3. Given age 62 morbidities (M_{i1}), general health shock $\varepsilon_{2,i1}$ is drawn and age 62 self-rated health (s_{i1}) is computed using equation (3).

4. Given age 62 self-rated health (s_{i1}) and morbidities (M_{i1}), depression shock $\varepsilon_{3,i1}$ is drawn to determine age 62 depression state (d_{i1}) using equation (4).

5. Given age 62 depression (d_{i1}), self-rated health (s_{i1}) and morbidities (M_{i1}), retirement shock $\varepsilon_{4,i1}$ is drawn to determine age 62 retirement (r_{i1}) using equation (5).

6. Given all other age 62 outcomes $(r_{i1}, d_{i1}, s_{i1}, M_{i1})$, consumption shock $\varepsilon_{5,i1}$ is drawn to determine age 62 consumption (c_{i1}) using equation (6).[10]

7. Steps 1-6 are repeated for $t = 2, 3, \ldots$ until death or $t = 30$ (age 120).

8. Steps 1-7 are repeated 5,000 times for each individual.

A comparison between the average simulated life-cycle profiles and those based on available data is shown by cohort in Figures 4-8. Overall, the simulations match the available aggregated data well suggesting our life-cycle dynamics model provides a reasonable approximation of the underlying data generating processes. The simulations also match the standard deviation of consumption and health utility quite well (Figure 8). Note that by construction, the data and simulations are the same at age 60. However, using only age 60 data and the estimated model parameters, the simulations continue to match the data reasonably well even up to 24 years later (when the EHRS cohort is age 84).

[9] Initial conditions also include unobserved endowments $\hat{\pi}$ estimated from model (6) using the prediction method of De Vos et al. (2015).

[10] ε_5 is drawn from the normal distribution with mean zero and standard deviation determined to match the empirical error distribution of each cohort. Specifically, standard deviations used for EHRS, LHRS, WB, and BB cohorts are 0.49, 0.48, 0.48, and 0.40. Clustering by cohort provides a slightly better fit to the data, but main results change negligibly with use of a common standard deviation.

B.4 Figures and Tables

Table 1: Estimation sample descriptive statistics by cohort

	AHEAD	CODA	EHRS	LHRS	WB	BB	MBB	LBB
Individuals	7,758	4,233	5,368	5,138	3,628	4,802	5,131	4,650
Observations	37,177	28,535	46,201	46,623	29,037	25,719	18,761	6,038
Age (mean)	81.75	75.23	67.64	62.74	60.47	58.45	55.34	52.70
Hypertension (%)	54.70	57.22	53.42	50.63	49.67	49.78	47.58	44.96
Diabetes (%)	15.46	18.89	19.38	18.16	18.65	20.49	19.80	20.04
Cancer (%)	16.83	17.81	14.02	11.16	10.69	8.52	7.60	6.86
Lung disease (%)	9.45	10.17	9.55	8.51	7.22	6.98	7.58	7.72
Heart disease (%)	35.36	31.02	23.18	19.16	16.86	14.79	12.57	10.74
Stroke (%)	15.30	12.15	7.43	6.06	5.75	5.07	4.39	4.49
Psyche problem (%)	11.85	11.69	11.08	12.87	16.95	19.35	19.41	20.28
Arthritis (%)	55.99	60.17	57.43	52.48	51.82	46.19	39.51	33.06
Difficulty with ADLs (%)	40.50	28.87	24.07	21.73	22.18	21.42	19.50	15.18
Depression (%)								
CESD=0	33.48	41.13	45.39	46.33	46.75	44.26	40.30	36.54
CESD=1	21.25	21.64	20.56	20.36	21.36	21.29	23.14	24.64
CESD=2	14.30	12.55	11.54	11.04	10.94	10.56	11.24	12.56
CESD=3	10.18	8.61	7.53	7.14	6.65	6.68	6.73	7.51
CESD=4	7.46	5.55	5.16	4.85	4.47	4.42	4.77	5.23
CESD=5	5.50	4.18	3.84	3.66	3.32	3.89	3.94	4.29
CESD=6	4.11	3.14	2.91	3.00	2.84	3.40	4.01	3.63
CESD=7	2.43	2.12	2.04	2.27	2.26	3.30	3.44	3.29
CESD=8	1.28	1.09	1.03	1.35	1.40	2.21	2.44	2.31
Self-rated health (%)								
Poor	14.26	10.37	9.29	7.81	6.63	7.76	7.25	7.43
Fair	25.76	21.73	19.37	18.77	16.93	19.64	21.23	22.83
Good	30.86	32.26	31.62	30.88	30.72	30.31	31.27	31.20
Very good	21.35	26.40	28.10	28.96	31.99	30.15	29.44	27.00
Excellent	7.77	9.25	11.63	13.58	13.72	12.15	10.81	11.54
Retired (%)	95.32	90.78	75.68	62.30	56.79	47.18	38.93	34.11
Annual consumption ($1000s, mean)	22.22	24.94	24.84	26.19	26.67	23.63	19.58	17.95
Male (%)	37.59	46.81	45.05	45.29	37.56	42.41	42.52	39.96
Education (%)								
<HS	41.68	32.47	31.19	28.27	21.36	20.10	21.96	22.36
HS	29.62	31.42	32.71	32.85	30.91	24.60	24.97	23.55
Some College	16.34	17.81	18.57	20.51	24.45	28.36	29.27	29.20
College	12.36	18.30	17.53	18.38	23.28	26.94	23.80	24.89
Race (%)								
White	84.95	86.88	80.28	79.92	80.15	67.56	60.15	53.41
Black	12.95	9.69	16.32	15.99	14.89	21.39	26.17	27.28
Other	2.11	3.44	3.40	4.09	4.96	11.05	13.68	19.31

Notes: Children of the Depression denoted by CODA, War Babies by WB, early Baby Boomers by BB, and mid Baby Boomers by MBB. Consumption is reported in real 2010 dollars. Source: HRS.

Table 2: Representative and simulation sample comparison

	EHRS		LHRS		WB		BB	
	Rep	Sim	Rep	Sim	Rep	Sim	Rep	Sim
	0	1	2	3	4	5	6	7
Individuals	3,160	3,091	3,816	3,607	2,697	2,572	3,015	2,737
Male (%)	47.20	46.36	46.82	46.61	47.89	47.92	48.25	47.53
Education (%)								
<HS	29.08	28.88	25.32	25.43	18.73	18.47	14.88	14.94
HS	33.61	33.80	32.04	32.28	30.45	30.27	24.80	24.92
Some College	19.28	19.25	21.56	21.42	24.35	24.45	29.19	28.90
College	18.04	18.08	21.09	20.87	26.46	26.81	31.13	31.24
Race (%)								
White	86.31	86.55	86.15	86.54	85.48	85.95	81.57	81.76
Black	10.38	10.24	10.00	9.97	9.73	9.23	10.81	10.61
Other	3.31	3.20	3.85	3.49	4.79	4.82	7.63	7.63

Notes: War Babies denoted by WB and Baby Boomers by BB. EHRS cohort inclues those under age 60 in 1992. "Rep" indicates representative sample based on HRS respondent analysis weights. "Sim" indicates simulation sample weighted by the same analysis weights.

Table 3: Simulation sample initial conditions by cohort

	EHRS	LHRS	WB	BB
Age (mean)	60	60	60	60
Hypertension (%)	38.00	41.68	47.45	50.48
Diabetes (%)	11.80	12.64	16.26	20.22
Cancer (%)	6.81	8.24	10.72	9.39
Lung disease (%)	7.06	6.74	7.08	7.95
Heart disease (%)	13.72	14.60	15.81	16.04
Stroke (%)	2.88	3.88	5.17	4.52
Psyche problem (%)	7.30	11.73	17.04	20.98
Arthritis (%)	44.61	47.94	51.44	51.81
Difficulty with ADLs (%)	11.75	19.53	22.40	22.40
Depression (%)				
CESD=0	50.98	44.72	49.00	47.92
CESD=1	21.38	21.55	19.45	21.27
CESD=2	9.51	11.45	10.39	8.80
CESD=3	5.92	7.48	6.99	6.45
CESD=4	3.31	4.77	4.44	4.00
CESD=5	3.11	3.65	3.19	3.19
CESD=6	2.55	2.79	2.63	2.79
CESD=7	2.03	2.53	2.28	3.27
CESD=8	1.24	1.05	1.64	2.32
Self-rated health (%)				
Poor	7.31	6.68	6.60	7.26
Fair	15.20	16.71	16.60	17.14
Good	28.32	30.11	31.09	29.39
Very good	31.66	30.80	31.72	34.16
Excellent	17.51	15.70	13.99	12.05
Retired (%)	48.63	50.43	48.06	47.54
Annual consumption ($1000s, mean)	27.59	29.50	28.96	25.86

Notes: Mean and percentage estimates use base year respondent analysis weights. War Babies denoted by WB and Baby Boomers by BB. Consumption is reported in real 2010 dollars. Source: HRS.

Table 4: Morbidity shock covariance matrix (Σ)

	Hyper	Diabetes	Cancer	Lung	Heart	Stroke	Psych	Arthritis	ADLs
Hyper	1.00	0.27	0.05	0.08	0.28	0.29	0.14	0.09	0.10
Diabetes	0.27	1.00	0.06	0.05	0.10	0.14	0.08	0.04	0.07
Cancer	0.05	0.06	1.00	0.13	0.04	0.06	0.12	0.05	0.13
Lung	0.08	0.05	0.13	1.00	0.24	0.11	0.18	0.08	0.18
Heart	0.28	0.10	0.04	0.24	1.00	0.28	0.17	0.09	0.14
Stroke	0.29	0.14	0.06	0.11	0.28	1.00	0.21	0.11	0.39
Psych	0.14	0.08	0.12	0.18	0.17	0.21	1.00	0.16	0.28
Arthritis	0.09	0.04	0.05	0.08	0.09	0.11	0.16	1.00	0.26
ADLs	0.10	0.07	0.13	0.18	0.14	0.39	0.28	0.26	1.00

Table 5: Model estimates for self-rated health, depression, retirement, consumption, and mortality

	Health		Depression		Retirement		Consumption		Mortality	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Hyper	-0.271	0.014	0.015	0.015	0.053	0.034	0.001	0.013	0.101	0.026
Diab	-0.258	0.018	-0.031	0.019	0.113	0.042	-0.002	0.017	0.098	0.032
Cancer	-0.684	0.019	0.065	0.019	0.255	0.046	0.030	0.012	0.656	0.026
Lung	-0.461	0.022	0.106	0.023	0.032	0.064	0.002	0.018	0.412	0.031
Heart	-0.484	0.015	0.044	0.016	0.188	0.040	-0.002	0.018	0.197	0.024
Stroke	-0.482	0.021	0.060	0.023	0.439	0.065	-0.073	0.030	0.240	0.029
Psych	-0.399	0.021	0.527	0.022	0.252	0.054	-0.055	0.019	0.220	0.029
Arthritis	-0.222	0.014	0.090	0.015	0.091	0.032	0.017	0.015	-0.022	0.024
ADL	-0.654	0.013	0.368	0.014	0.376	0.035	-0.063	0.015	0.315	0.019
CESD=1					-0.001	0.019	0.007	0.006	0.018	0.017
CESD=2					0.016	0.026	0.019	0.009	0.089	0.019
CESD=3					0.048	0.033	0.015	0.009	0.121	0.023
CESD=4					0.042	0.041	0.026	0.013	0.137	0.025
CESD=5					0.125	0.044	0.024	0.015	0.125	0.027
CESD=6					0.190	0.051	0.033	0.021	0.136	0.030
CESD=7					0.233	0.060	0.028	0.025	0.036	0.034
CESD=8					0.189	0.076	0.041	0.031	0.071	0.047
Health 2			-0.351	0.013	-0.558	0.041	0.053	0.012	-0.320	0.017
Health 3			-0.646	0.014	-0.712	0.042	0.070	0.014	-0.517	0.019
Health 4			-0.838	0.015	-0.737	0.044	0.085	0.017	-0.624	0.023
Health 5 (best)			-0.937	0.019	-0.733	0.049	0.094	0.022	-0.615	0.032
Lag Hyper	0.149	0.019	-0.051	0.022	0.010	0.046	-0.005	0.013	-0.047	0.026
Lag Diab	0.082	0.025	0.031	0.027	-0.112	0.063	-0.005	0.015	0.072	0.033
Lag Cancer	0.533	0.028	-0.087	0.029	-0.190	0.072	-0.009	0.017	-0.444	0.028
Lag Lung	0.213	0.033	-0.031	0.034	0.121	0.095	-0.011	0.018	-0.115	0.034
Lag Heart	0.281	0.022	-0.027	0.022	-0.148	0.061	0.000	0.013	-0.034	0.025
Lag Stroke	0.349	0.031	-0.015	0.033	-0.340	0.109	-0.004	0.024	-0.047	0.031
Lag Psych	0.257	0.030	-0.242	0.030	-0.063	0.079	0.029	0.019	-0.132	0.031
Lag Arthritis	0.121	0.018	-0.031	0.019	-0.020	0.042	-0.006	0.013	-0.075	0.023
Lag ADL	0.344	0.018	-0.184	0.019	-0.210	0.051	0.008	0.012	-0.113	0.019
Lag CESD=1	-0.072	0.008	0.429	0.008	0.008	0.019	0.006	0.007	-0.033	0.016
Lag CESD=2	-0.124	0.010	0.633	0.011	0.008	0.027	0.017	0.010	-0.016	0.020
Lag CESD=3	-0.134	0.013	0.775	0.012	-0.015	0.034	0.019	0.013	-0.037	0.022
Lag CESD=4	-0.151	0.015	0.890	0.015	-0.014	0.042	0.021	0.015	-0.002	0.027
Lag CESD=5	-0.143	0.017	1.024	0.017	0.007	0.047	0.026	0.016	-0.033	0.029
Lag CESD=6	-0.176	0.019	1.131	0.019	0.018	0.055	0.041	0.021	-0.071	0.031
Lag CESD=7	-0.222	0.023	1.302	0.022	0.093	0.066	0.032	0.019	-0.109	0.038
Lag CESD=8	-0.236	0.029	1.534	0.026	-0.026	0.084	0.028	0.029	-0.138	0.047
Lag Health 2	0.597	0.014	0.045	0.013	-0.078	0.053	0.022	0.012	-0.040	0.018
Lag Health 3	1.080	0.015	0.062	0.015	-0.108	0.054	0.031	0.015	-0.073	0.020
Lag Health 4	1.603	0.016	0.043	0.016	-0.140	0.056	0.043	0.016	-0.105	0.023
Lag Health 5	2.216	0.018	0.047	0.019	-0.202	0.060	0.065	0.018	-0.134	0.031
Time	0.018	0.003	-0.024	0.003	-0.003	0.009	-0.000	0.011	-0.015	0.005
2008+	0.006	0.011	0.006	0.012	-0.058	0.030	-0.046	0.008	0.045	0.021
CODA	0.027	0.016	0.020	0.016	0.063	0.068			-0.008	0.024
Early HRS	0.014	0.022	-0.008	0.022	0.096	0.081			-0.047	0.033
Late HRS	0.001	0.028	-0.006	0.029	0.030	0.095			-0.060	0.043
War Babies	-0.016	0.034	0.004	0.035	0.057	0.111			-0.114	0.054
Boomers	-0.086	0.041	0.040	0.043	0.002	0.132			-0.138	0.066
Mid Boomers	-0.129	0.049	0.066	0.051	-0.083	0.151			-0.179	0.082
Late Boomers	-0.139	0.069	0.107	0.073	-0.074	0.187			-0.077	0.160
Black	-0.044	0.008	0.036	0.009	0.051	0.021			0.060	0.016
Other race	-0.089	0.014	0.048	0.014	-0.053	0.033			-0.075	0.029
Female	0.044	0.007	0.074	0.007	0.126	0.016			-0.225	0.012
HS grad	0.080	0.008	-0.039	0.008	-0.010	0.021			0.042	0.014
Some college	0.112	0.009	-0.070	0.009	-0.045	0.023			0.027	0.017
College grad	0.185	0.010	-0.127	0.011	-0.087	0.026			0.018	0.020
Retired							-0.039	0.013	0.205	0.029
Lag Retired	-0.023	0.013	-0.007	0.013			-0.023	0.014	-0.021	0.025
Lag2 Retired	-0.015	0.012	-0.001	0.013						
Lag Con							0.169	0.004		
Lag2 Con							0.082	0.005		
Constant					-1.146	0.194			-1.820	0.246

Notes: Dependent variable across columns. Standard (ordered) probit results reported for self-rated heath, depression, mortality, and retirement as dependant outcomes. Linear dynamic panel estimates reported for consumption as outcome. All regressions also include dummies for age. Regressions for self-rated health, depression, mortality, and retirement also include dummies for occupation and census division. Regressions for self-rated health and depression also includes second lag for all health outcomes.

Table 6: Model estimates for morbidities

	Hypertension		Diabetes		Cancer		Lung disease		Heart disease	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Lag Hyper			0.261	0.033	-0.028	0.038	0.082	0.039	0.115	0.033
Lag Diab	0.259	0.049			0.055	0.045	0.040	0.049	0.042	0.042
Lag Cancer	-0.042	0.051	0.023	0.051			0.032	0.057	-0.076	0.049
Lag Lung	0.093	0.058	0.079	0.056	0.085	0.059			0.259	0.050
Lag Heart	0.096	0.043	0.076	0.040	0.010	0.041	0.229	0.039		
Lag Stroke	0.089	0.066	-0.034	0.062	-0.015	0.058	0.010	0.059	0.074	0.053
Lag Psych	0.045	0.054	0.061	0.051	-0.063	0.058	0.077	0.055	0.076	0.048
Lag Arthritis	0.086	0.029	-0.005	0.032	-0.005	0.034	0.134	0.037	0.069	0.029
Lag ADL	0.042	0.032	0.025	0.034	0.013	0.034	0.070	0.035	0.067	0.029
Lag CESD=1	0.002	0.018	-0.010	0.020	0.006	0.020	0.029	0.025	0.001	0.018
Lag CESD=2	0.028	0.024	0.011	0.026	0.013	0.026	0.072	0.030	-0.005	0.024
Lag CESD=3	0.010	0.030	0.015	0.031	-0.011	0.032	0.086	0.034	0.018	0.028
Lag CESD=4	0.030	0.037	-0.039	0.038	0.021	0.037	0.088	0.042	0.026	0.033
Lag CESD=5	0.075	0.040	0.000	0.040	-0.038	0.045	0.051	0.049	0.061	0.039
Lag CESD=6	0.095	0.046	-0.000	0.045	0.038	0.047	0.039	0.051	0.076	0.041
Lag CESD=7	0.100	0.052	-0.011	0.053	0.049	0.054	0.173	0.051	-0.016	0.052
Lag CESD=8	0.009	0.069	0.064	0.066	-0.039	0.073	-0.001	0.071	-0.115	0.071
Lag Health 2	0.023	0.035	0.007	0.032	-0.036	0.033	-0.060	0.032	-0.108	0.029
Lag Health 3	0.023	0.036	-0.012	0.034	-0.060	0.035	-0.123	0.035	-0.162	0.031
Lag Health 4	-0.020	0.038	-0.092	0.037	-0.088	0.038	-0.263	0.039	-0.241	0.034
Lag Health 5	-0.097	0.042	-0.219	0.045	-0.121	0.045	-0.401	0.055	-0.288	0.040
Lag2 Hyper			0.032	0.032	0.058	0.038	-0.089	0.039	0.057	0.032
Lag2 Diab	-0.090	0.052			-0.061	0.048	-0.119	0.052	0.106	0.044
Lag2 Cancer	0.022	0.055	-0.019	0.055			0.062	0.061	0.077	0.052
Lag2 Lung	-0.154	0.064	-0.082	0.061	0.054	0.063			-0.106	0.055
Lag2 Heart	-0.050	0.046	0.013	0.042	0.025	0.043	-0.098	0.041		
Lag2 Stroke	-0.036	0.074	0.057	0.067	0.013	0.064	0.014	0.065	0.062	0.058
Lag2 Psych	-0.026	0.057	-0.065	0.054	0.064	0.061	0.050	0.058	-0.020	0.050
Lag2 Arthre	-0.043	0.029	-0.009	0.032	0.048	0.034	-0.031	0.036	0.032	0.029
Lag2 ADL	-0.050	0.035	0.035	0.036	-0.018	0.036	-0.010	0.036	0.010	0.031
Lag2 CESD=1	0.006	0.019	0.027	0.021	0.013	0.020	0.046	0.024	0.012	0.018
Lag2 CESD=2	-0.030	0.025	0.049	0.026	-0.005	0.026	0.029	0.029	0.040	0.023
Lag2 CESD=3	0.027	0.029	0.058	0.033	0.065	0.031	0.077	0.034	-0.055	0.029
Lag2 CESD=4	0.034	0.036	0.060	0.038	0.018	0.039	0.022	0.041	-0.015	0.035
Lag2 CESD=5	-0.040	0.044	0.021	0.044	0.019	0.045	-0.024	0.050	0.040	0.038
Lag2 CESD=6	0.028	0.048	-0.014	0.048	-0.002	0.051	0.048	0.050	-0.004	0.043
Lag2 CESD=7	0.046	0.056	0.034	0.052	0.100	0.054	0.019	0.057	0.085	0.048
Lag2 CESD=8	0.043	0.070	-0.039	0.072	-0.179	0.088	0.133	0.067	-0.005	0.066
Lag2 Health 2	-0.018	0.037	-0.066	0.033	-0.053	0.035	-0.082	0.034	0.003	0.032
Lag2 Health 3	-0.015	0.038	-0.060	0.035	-0.009	0.037	-0.115	0.037	-0.009	0.033
Lag2 Health 4	-0.031	0.039	-0.107	0.038	0.002	0.039	-0.159	0.040	-0.034	0.036
Lag2 Health 5	-0.060	0.043	-0.138	0.044	0.009	0.045	-0.250	0.052	-0.081	0.041
Time	0.040	0.007	0.027	0.008	0.004	0.008	0.013	0.009	-0.005	0.007
2008+	-0.057	0.027	-0.064	0.029	0.016	0.029	-0.008	0.034	-0.049	0.026
CODA	-0.031	0.038	-0.022	0.042	-0.020	0.039	0.002	0.044	-0.014	0.035
Early HRS	-0.086	0.052	-0.043	0.057	-0.078	0.053	-0.054	0.061	0.021	0.048
Late HRS	-0.080	0.066	-0.042	0.072	-0.100	0.069	-0.005	0.078	0.040	0.061
War Babies	-0.091	0.081	0.007	0.089	-0.071	0.085	-0.012	0.098	0.069	0.076
Boomers	-0.170	0.099	0.033	0.108	-0.127	0.105	-0.036	0.119	0.105	0.092
Mid Boomers	-0.305	0.117	0.053	0.127	-0.097	0.125	0.022	0.142	0.133	0.110
Late Boomers	-0.327	0.156	0.211	0.163	-0.158	0.196	-0.009	0.210	0.116	0.165
Black	0.188	0.021	0.079	0.021	-0.035	0.022	-0.145	0.025	-0.133	0.020
Other race	0.064	0.030	0.210	0.031	-0.175	0.040	-0.090	0.041	-0.103	0.033
Female	0.016	0.015	-0.106	0.017	-0.202	0.017	-0.051	0.020	-0.174	0.015
HS grad	-0.039	0.018	-0.076	0.019	0.007	0.020	-0.079	0.022	0.026	0.018
Some college	-0.074	0.021	-0.074	0.022	0.048	0.023	-0.073	0.026	0.050	0.021
College grad	-0.117	0.024	-0.137	0.026	0.050	0.026	-0.183	0.032	-0.020	0.024
Lag Retired	-0.015	0.029	0.024	0.030	0.036	0.032	0.070	0.038	0.016	0.029
Lag2 Retired	-0.004	0.028	-0.019	0.030	-0.014	0.031	-0.002	0.036	-0.016	0.029
Constant	-1.582	0.087	-2.055	0.093	-1.951	0.094	-2.108	0.108	-1.752	0.084

Notes: Multivarite probit results with dependent variable across columns. Regressions also include dummies for age, occupation, and census division.

Table 7: Model estimates for morbidities (continued)

	Stroke		Psych		Arthritis		ADLs	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Lag Hyper	0.099	0.040	0.129	0.037	0.072	0.032	0.040	0.030
Lag Diab	0.035	0.052	0.015	0.050	0.055	0.044	0.083	0.039
Lag Cancer	-0.030	0.058	-0.043	0.059	0.065	0.050	0.024	0.043
Lag Lung	0.015	0.062	0.089	0.062	0.157	0.064	0.160	0.048
Lag Heart	0.184	0.039	0.079	0.041	0.090	0.041	0.076	0.033
Lag Stroke			0.249	0.051	-0.027	0.061	0.370	0.046
Lag Psych	0.124	0.053			0.254	0.053	0.274	0.043
Lag Arthritis	-0.005	0.036	0.109	0.035			0.206	0.025
Lag ADL	0.164	0.032	0.149	0.032	0.152	0.036		
Lag CESD=1	0.022	0.023	0.081	0.024	0.050	0.018	0.095	0.017
Lag CESD=2	0.033	0.028	0.180	0.029	0.042	0.024	0.167	0.021
Lag CESD=3	0.062	0.034	0.226	0.034	0.010	0.033	0.218	0.025
Lag CESD=4	0.062	0.039	0.352	0.038	0.083	0.037	0.230	0.031
Lag CESD=5	0.171	0.043	0.362	0.042	0.055	0.043	0.235	0.036
Lag CESD=6	0.135	0.048	0.341	0.050	0.071	0.049	0.214	0.043
Lag CESD=7	0.138	0.054	0.442	0.054	0.035	0.059	0.353	0.047
Lag CESD=8	0.022	0.072	0.537	0.070	0.024	0.077	0.353	0.062
Lag Health 2	-0.114	0.031	-0.147	0.031	-0.069	0.037	-0.182	0.030
Lag Health 3	-0.187	0.034	-0.185	0.033	-0.092	0.038	-0.388	0.031
Lag Health 4	-0.217	0.039	-0.254	0.038	-0.131	0.040	-0.537	0.033
Lag Health 5	-0.336	0.050	-0.330	0.050	-0.226	0.044	-0.604	0.040
Lag2 Hyper	0.042	0.039	-0.072	0.037	0.028	0.033	-0.015	0.030
Lag2 Diab	0.091	0.054	0.014	0.053	-0.046	0.047	0.024	0.041
Lag2 Cancer	0.006	0.062	0.071	0.063	-0.018	0.054	-0.010	0.046
Lag2 Lung	0.045	0.067	0.028	0.067	-0.066	0.071	-0.018	0.053
Lag2 Heart	-0.029	0.041	-0.064	0.043	-0.019	0.044	-0.075	0.035
Lag2 Stroke			-0.152	0.058	0.018	0.068	-0.173	0.052
Lag2 Psych	-0.031	0.056			-0.116	0.057	-0.136	0.046
Lag2 Arthre	-0.014	0.035	-0.025	0.034			0.044	0.025
Lag2 ADL	-0.089	0.034	-0.071	0.034	-0.053	0.041		
Lag2 CESD=1	-0.011	0.023	0.059	0.023	0.022	0.018	0.091	0.016
Lag2 CESD=2	0.014	0.028	0.071	0.031	0.031	0.026	0.101	0.021
Lag2 CESD=3	-0.023	0.035	0.183	0.034	0.041	0.032	0.096	0.027
Lag2 CESD=4	-0.036	0.045	0.212	0.038	0.028	0.038	0.152	0.032
Lag2 CESD=5	0.015	0.045	0.137	0.048	0.001	0.047	0.168	0.037
Lag2 CESD=6	0.015	0.051	0.205	0.051	0.001	0.054	0.176	0.043
Lag2 CESD=7	-0.044	0.058	0.250	0.055	-0.033	0.063	0.162	0.052
Lag2 CESD=8	0.044	0.072	0.348	0.076	0.110	0.077	0.219	0.067
Lag2 Health 2	-0.052	0.034	-0.024	0.033	0.037	0.040	-0.136	0.033
Lag2 Health 3	-0.053	0.037	-0.028	0.036	0.056	0.041	-0.229	0.033
Lag2 Health 4	-0.031	0.040	-0.092	0.040	0.030	0.043	-0.321	0.035
Lag2 Health 5	-0.071	0.049	-0.135	0.049	-0.035	0.046	-0.385	0.041
Time	-0.025	0.008	0.010	0.008	-0.032	0.007	-0.048	0.006
2008+	0.018	0.032	-0.104	0.033	0.021	0.028	0.040	0.025
CODA	0.012	0.038	0.067	0.042	-0.104	0.038	0.089	0.031
Early HRS	0.003	0.053	0.066	0.058	-0.108	0.053	0.134	0.044
Late HRS	0.015	0.069	0.108	0.074	-0.017	0.066	0.137	0.057
War Babies	0.099	0.086	0.227	0.091	0.103	0.081	0.167	0.070
Boomers	0.086	0.105	0.309	0.111	0.135	0.099	0.250	0.086
Mid Boomers	0.152	0.128	0.300	0.131	0.159	0.117	0.322	0.101
Late Boomers	0.549	0.195	0.040	0.195	0.209	0.153	0.309	0.148
Black	0.051	0.023	-0.207	0.025	-0.002	0.020	0.084	0.018
Other race	-0.128	0.045	-0.049	0.038	-0.047	0.031	0.045	0.031
Female	-0.071	0.019	0.104	0.019	0.161	0.015	-0.029	0.014
HS grad	0.044	0.021	-0.047	0.021	-0.027	0.019	-0.078	0.016
Some college	0.059	0.025	0.030	0.025	0.009	0.021	-0.025	0.019
College grad	0.053	0.030	0.004	0.029	-0.024	0.024	-0.070	0.022
Lag Retired	0.068	0.040	0.052	0.037	0.003	0.028	0.153	0.026
Lag2 Retired	-0.003	0.037	-0.009	0.036	-0.032	0.028	-0.071	0.025
Constant	-2.575	0.112	-2.234	0.099	-1.342	0.088	-1.239	0.078

Notes: Multivarite probit results with dependent variable across columns. Regressions also include dummies for age, occupation, and census division.

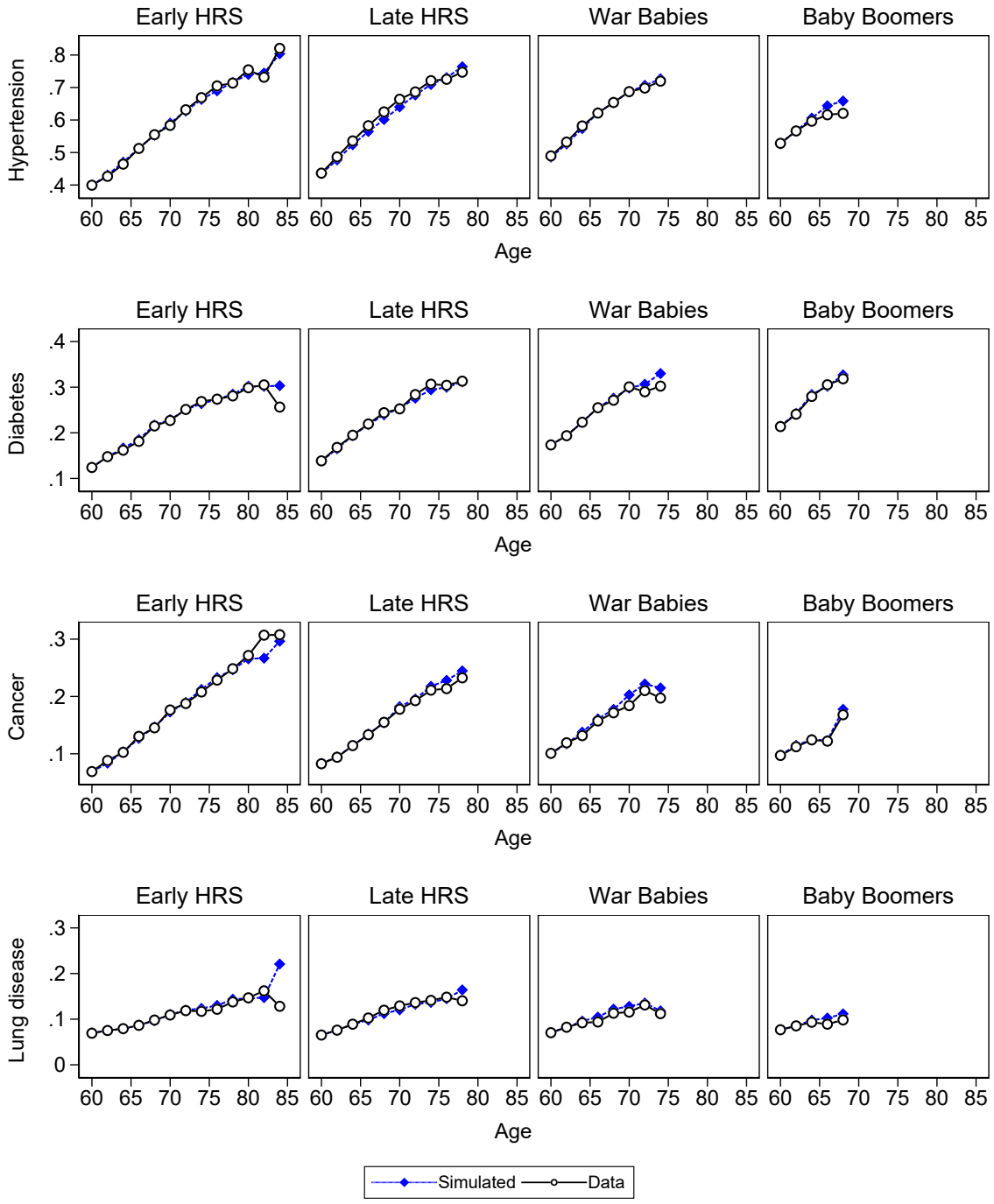

Figure 4: Mean of life-cycle morbidity profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

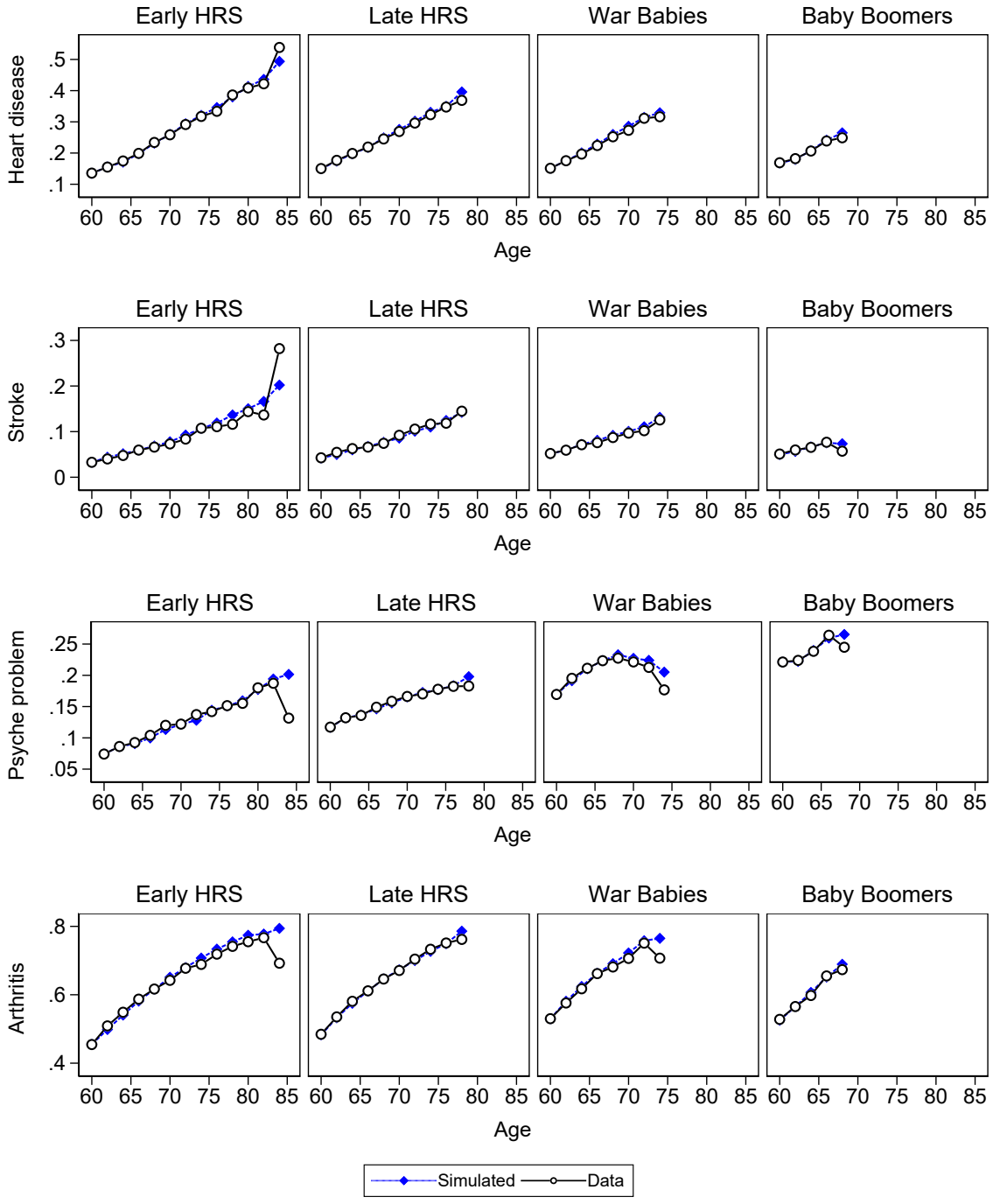

Figure 5: Mean of life-cycle morbidity profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

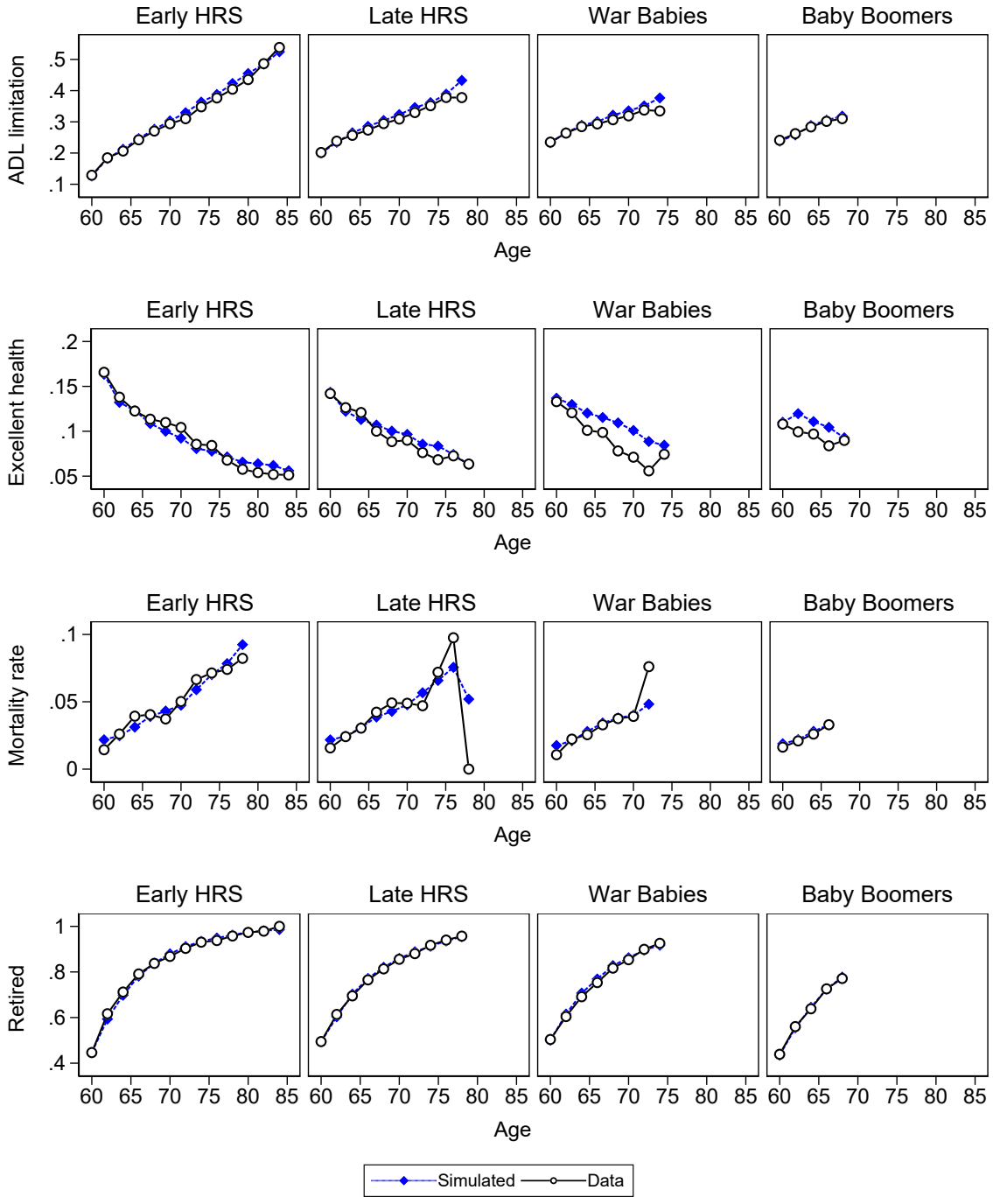

Figure 6: Mean of life-cycle health, mortality, and retirement profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

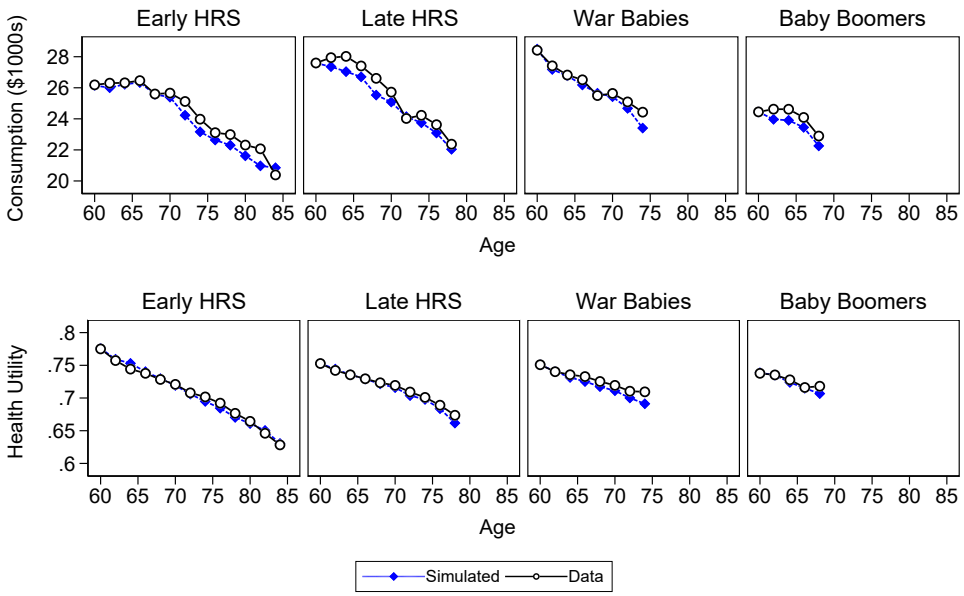

Figure 7: Mean of life-cycle consumption and health utility profiles by cohort

Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

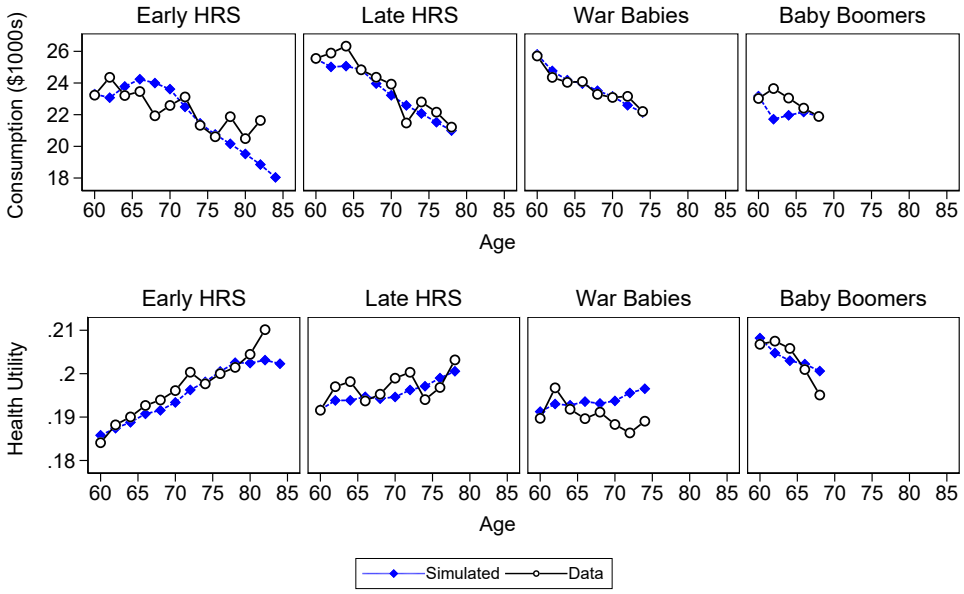

Figure 8: Standard deviation of consumption and health utility life-cycle profiles by cohort

Notes: "Data" plots standard deviation of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of standard deviations of simulated outcome (i.e. the mean of standard deviations calculated for each of the 5,000 simulation runs).

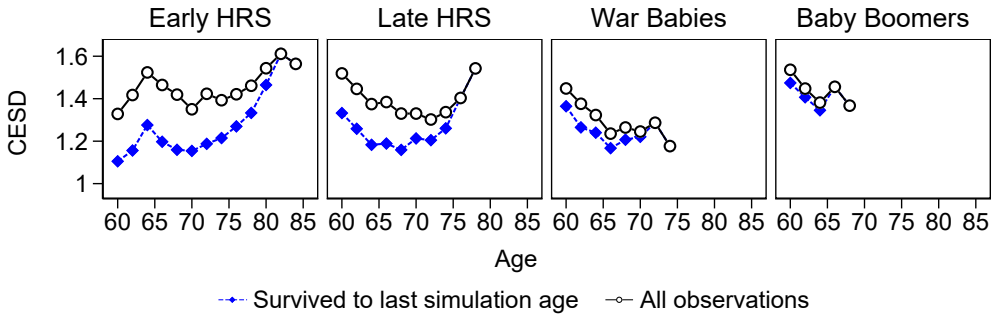

Figure 9: Mean of life-cycle CESD profiles by cohort

Notes: "All observations" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Survived to last simulation age" plots mean data for all individuals that survived to the last simulation age.

C Welfare decomposition

Welfare condition (8) may be rewritten:

$$log\left(1-\lambda_{ij}\right)=\frac{U_{ij}\left(1\right)-U_{mj}\left(1\right)}{E\left[\sum_{a=j}^{J}\psi_{ma}\beta^{a-j}\phi\left(h_{ma}\right)\right]}.$$

Let u_{ia} denote flow utility unadjusted for health at age a given outcome bundles i: $u_{ia}=\bar{u}+log\left(c_{ia}\right)+\nu\left(l_{ia}\right)$. Moreover, denote the expected value conditional on survival with subscript ψ: $E_{\psi}\left[u_{ia}\right]=E\left[u_{ia}\mid\psi_{ia}=1\right]$. Note that the welfare condition under our benchmark preference specification is then given as:

$$\begin{aligned}log\left(1-\lambda_{ij}\right)&=\tilde{\psi}\left(U_{ij}-E\left[\sum_{a=j}^{J}\psi_{ma}\beta^{a-j}\phi\left(h_{ma}\right)\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right]\right)\\&=\tilde{\psi}\left(E\left[\sum_{a=j}^{J}\psi_{ia}\beta^{a-j}\phi(h_{ia})u_{ia}\right]-E\left[\sum_{a=j}^{J}\psi_{ma}\beta^{a-j}\phi\left(h_{ma}\right)\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right]\right)\\&=\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left(E\left[\psi_{ia}\phi(h_{ia})u_{ia}\right]-E\left[\psi_{ma}\phi\left(h_{ma}\right)\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right]\right)\end{aligned}$$

where $\tilde{\psi}$ is the reciprocal of the reference discounted quality-adjusted life expectancy (QALE):

$$\tilde{\psi}=\frac{1}{E\left[\sum_{a=j}^{J}\psi_{ma}\beta^{a-j}\phi\left(h_{ma}\right)\right]}.$$

Also note that $E\left[\psi_{ia}u_{ia}\right]=E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right]$ from the definition of conditional probability where $E_{\psi}\left[u_{ia}\right]$ denotes the expected flow utility conditional on survival. Then adding and subtracting the term:

$$\begin{aligned}\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Bigg(&E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E_{\psi}\left[u_{ia}\right]+E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[u_{ia}\right]-E_{\psi}\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right)\\&+\left(E\left[\psi_{ia}\right]-E\left[\psi_{ma}\right]\right)E_{\psi}\left[\phi\left(h_{ma}\right)\right]E_{\psi}\left[u_{ia}\right]\\&+\left(E_{\psi}\left[\phi\left(h_{ia}\right)\right]-E_{\psi}\left[\phi\left(h_{ma}\right)\right]\right)E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right]\Bigg).\end{aligned}$$

from the right hand side of the above welfare condition gives:

$$
\begin{aligned}
log\left(1-\lambda_{ij}\right) = \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left(E\left[\psi_{ia}\phi\left(h_{ia}\right)u_{ia}\right]-E\left[\psi_{ma}\phi\left(h_{ma}\right)\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right]\right) \\
+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E_{\psi}\left[u_{ia}\right]+E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[u_{ia}\right]-E_{\psi}\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right)\Big) \\
+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(\left(E\left[\psi_{ia}\right]-E\left[\psi_{ma}\right]\right)E_{\psi}\left[\phi\left(h_{ma}\right)\right]E_{\psi}\left[u_{ia}\right]\Big) \\
+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(\left(E_{\psi}\left[\phi\left(h_{ia}\right)\right]-E_{\psi}\left[\phi\left(h_{ma}\right)\right]\right)E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right]\Big) \\
-\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E_{\psi}\left[u_{ia}\right]+E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[u_{ia}\right]-E_{\psi}\left[\bar{u}+log\left(c_{ma}\right)+\nu\left(l_{ma}\right)\right]\right)\Big) \\
-\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(\left(E\left[\psi_{ia}\right]-E\left[\psi_{ma}\right]\right)E_{\psi}\left[\phi\left(h_{ma}\right)\right]E_{\psi}\left[u_{ia}\right]\Big) \\
-\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\Big(\left(E_{\psi}\left[\phi\left(h_{ia}\right)\right]-E_{\psi}\left[\phi\left(h_{ma}\right)\right]\right)E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right]\Big)
\end{aligned}
$$

Rearranging the terms of the above equation and using the definition of $E\left[\psi_{ia}u_{ia}\right]$ yields the following additive decomposition of welfare:

$$
\begin{aligned}
log\left(1-\lambda_{ij}\right) = &+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[\left(E_{\psi}\left[\phi\left(h_{ia}\right)\right]-E_{\psi}\left[\phi\left(h_{ma}\right)\right]\right)E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right]+\Phi\right]. && \textit{Health} && (1)\\
&+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left(E\left[\psi_{ia}\right]-E\left[\psi_{ma}\right]\right)E_{\psi}\left[\phi\left(h_{ma}\right)\right]E_{\psi}\left[u_{ia}\right] && \textit{Mortality} && (2)\\
&+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[\nu\left(l_{ia}\right)\right]-E_{\psi}\left[\nu\left(l_{ma}\right)\right]\right) && \textit{Leisure} && (3)\\
&\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[log\left(c_{ia}\right)\right]-E_{\psi}\left[log\left(c_{ma}\right)\right]\right)\right] && \textit{Consumption} && (4)
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi = &\left(E\left[\psi_{ia}\phi\left(h_{ia}\right)u_{ia}\right]-E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E_{\psi}\left[u_{ia}\right]\right) \\
&-\left(E\left[\psi_{ma}\phi\left(h_{ma}\right)\nu\left(l_{ma}\right)\right]-E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]E_{\psi}\left[\nu\left(l_{ma}\right)\right]\right).
\end{aligned}
$$

The first term in (1) is the expected utility gain from eliminating depression due only to gains in health utility—holding life expectancy, leisure, and consumption at their baseline levels. The Φ term is an adjustment for uncertainty over the life-cycle (the quantitative value of this term is generally quite small). Combined, these provide an individual's consumption-equivalent welfare before adjusting for expected differences in life expectancy, leisure, or consumption. The correction term (2) is the difference in life expectancy weighted by how much a life year is worth—the expected flow utility from outcome bundles of individual i in the baseline. The term (3) is the welfare adjustment for leisure—the expected utility difference in leisure weighted by the depression-free quality-adjusted life expectancy. Finally, term (4) corrects for expected consumption differences from eliminating depression over remaining life.

D Bootstrap standard errors

In order to gain a sense of how uncertainty in the underlying simulation model translates into uncertainty in our main welfare results, we estimate bootstrap standard errors and confidence intervals. This is computationally intensive so we pooled estimates across our imputed data sets. Specifically, for each of the $m = 12$ imputed data sets, we drew 30 bootstrap samples yielding a total of $m \times 30 = 360$ data sets. We then estimated our main welfare numbers in each data set. Finally, we pooled estimates across all 360 to estimate standard errors (see Schomaker and Heumann (2018) for validation of this approach with multiple imputation). Table 8 provides estimated bootstrap standard errors for the depression only and full model results. Overall, standard errors are quite small relative to point estimates and all major conclusions from our main analyses hold.

Table 8: Bootstrap estimated mean costs of depression after age sixty by birth cohort

	Depression only				Full model			
	EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB
Expected loss								
QALYs	0.853	0.826	0.793	0.816	2.059	1.972	1.884	1.924
	(0.016)	(0.019)	(0.025)	(0.035)	(0.093)	(0.100)	(0.100)	(0.115)
Labor supply (yrs)					0.093	0.097	0.094	0.111
					(0.031)	(0.032)	(0.031)	(0.036)
Consumption (annual)					-0.138	-0.144	-0.133	-0.125
					(0.086)	(0.086)	(0.080)	(0.076)
CV (λ)	0.084	0.083	0.078	0.076	0.148	0.144	0.134	0.129
	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.007)	(0.007)	(0.007)
$\lambda \times$ELC	47.164	48.099	45.779	42.953	91.501	92.989	87.962	82.352
	(1.562)	(1.330)	(1.972)	(2.483)	(5.047)	(5.395)	(5.638)	(6.365)
CV Gini	0.337	0.358	0.380	0.416	0.281	0.292	0.315	0.343
	(0.006)	(0.006)	(0.007)	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)

Notes: Bootstrap standard errors in parentheses. Estimates use base year respondent analysis weights. ELC denotes expected lifetime consumption. Consumption in $1000s.

Figure 10 further provides bootstrap standard errors for the outcome profiles in Figure 4 of the main text. Given the computational costs of bootstrapping we show standard errors only for overall health utility and not for each morbidity separately. Again standard errors are small relative to the point estimates provided in Figure 4 of the paper.

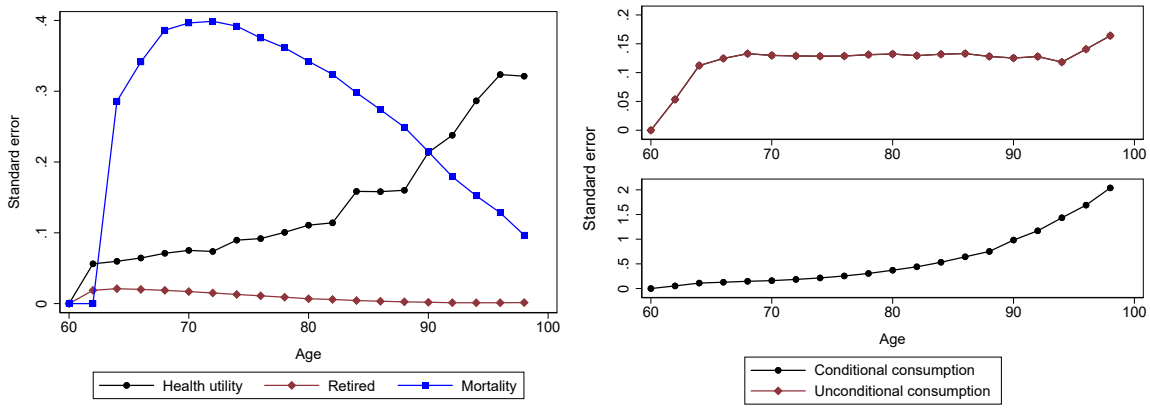

Figure 10: Standard errors of expected cost of depression by age

Notes: Results plot bootstrap standard errors for percentage difference in expected outcomes with the exogenous elimination of all prevalence of depression after age sixty. Sample includes all individuals in the simulation sample. Expected outcomes in first panel are conditional on survival.

E Robustness results

Here we present and discuss some additional robustness results (see Table 9).

Table 9: Additional robustness results

	Depression only				Full model			
	EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB
Benchmark	0.084	0.083	0.077	0.075	0.148	0.144	0.134	0.128
$\beta = 0.90$	0.078	0.079	0.073	0.072	0.121	0.119	0.110	0.105
$\varepsilon = 0.5$	0.085	0.084	0.079	0.077	0.151	0.147	0.137	0.131
$\varepsilon = 2$	0.082	0.081	0.075	0.073	0.144	0.140	0.130	0.124
$\theta = 16$	0.082	0.081	0.075	0.074	0.145	0.141	0.131	0.125
Health utility weights	0.091	0.090	0.083	0.081	0.153	0.148	0.137	0.131
NH consumption	0.098	0.096	0.090	0.089	0.152	0.148	0.140	0.136
No imputed CESD	0.084	0.083	0.078	0.076	0.148	0.144	0.135	0.128
No imputed data	–	0.077	0.063	0.060	–	0.102	0.084	0.076
Depression impact health	0.084	0.083	0.077	0.075	0.189	0.184	0.171	0.162
Full-time job	0.084	0.084	0.078	0.076	0.150	0.146	0.136	0.130

Notes: Mean CV (λ) reported. Estimates use base year respondent analysis weights. War Babies denoted by WB and Baby Boomers by BB.

E.1 Preference parameters

Rows 2-5 in Table 9 indicate sensitivity of our results to preference parameters β, ε and θ. With a lower time discount rate $\beta = 0.90$, the estimated mean welfare cost of late-life depression in the EHRS falls slightly to 7-12% of annual consumption. Raising or lowering the Frisch elasticity of labor supply ε or increasing disutility weight on labor supply θ have very minimal impacts on our welfare estimates as labor supply plays a relatively small role overall. In all cases, there are similar changes in welfare estimates for later cohorts and we continue to see a small decline in mean welfare costs across cohorts.

E.2 Health utility weights

In our calibration of health utility weights we assumed that the HUI3 measures relative utility across health states holding consumption and leisure fixed. While this approach is consistent with the interview instructions of the survey, there is some uncertainty around if respondents were fully capable of conceptualizing changing health states without changes in other aspects of life (Feeny et al., 2018). For example, if respondents considered changes in consumption and leisure in addition to health, the appropriate representation of the HUI3 instrument would be:

$$\gamma h \left[\bar{u} + log\left(c\right) + \nu\left(l\right)\right] = HUI3 \times h_{best} \left[\bar{u} + log\left(c_{best}\right) + \nu\left(l_{best}\right)\right].$$

Rearranging terms and setting $h_{best} = 1$ yields:

$$\gamma h = HUI3 \frac{\bar{u} + log\left(c_{best}\right) + \nu\left(l_{best}\right)}{\bar{u} + log\left(c\right) + \nu\left(l\right)}. \tag{5}$$

Generally, this formulation poses a problem because we do not observe the counterfactual consumption and leisure bundles that would be realized in the best health state. However, as we have already developed an independent forecasting model, we can predict the expected value for c_{best} and l_{best} for each individual in the sample. With these predictions in hand, we estimated the right hand side of (5) for each HUI3 respondent. We then regressed this value on CESD scale, self-rated health, and all morbidity indicators to obtain alternate utility weights γ (see results in Table 10). The sixth row in Table 9 shows that using these alternate utility weights very slightly increases the estimated mean welfare cost of late-life depression in all cohorts.

Table 10: Estimated alternate health utility weights (γ)

Measure	Weight	SE
Depression		
CESD=1	-0.014	0.023
CESD=2	-0.107	0.028
CESD=3	-0.095	0.034
CESD=4	-0.117	0.040
CESD=5	-0.082	0.045
CESD=6	-0.176	0.051
CESD=7	-0.285	0.066
CESD=8	-0.236	0.071
Hypertension	0.005	0.018
Diabetes	0.006	0.025
Cancer	0.021	0.024
Lung disease	-0.018	0.029
Heart disease	-0.031	0.020
Stroke	-0.038	0.031
Psych problem	-0.026	0.028
Arthritis	-0.041	0.018
Diff with ADL	-0.100	0.022
Self-rated health		
Fair	0.199	0.035
Good	0.251	0.037
Very good	0.324	0.039
Excellent	0.318	0.044
Constant	0.524	0.041

Notes: Results from regression of adjusted HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.17$. N = 760.

E.3 Nursing home consumption

For our measure of individual consumption we use household consumption from CAMS (minus health spending) divided by the number of household members. However, the number of household members excludes those residing in nursing homes. This means we are implicitly imputing average household consumption for members residing in nursing homes. For example, assume a couple has one member in a nursing home and one residing at home. Assume the member at home reported consumption of $20,000 on the CAMS survey. In this case, both members would be given $20,000

of consumption in our benchmark model. We believe this to be a reasonable imputation as much of the reported consumption is likely to be for the individual residing at home while the other member receives in-kind consumption from the nursing home (which could vary with nursing home quality). Also, note that if both members of a couple (or single person) are in a nursing home, the household does not receive a CAMS survey and consumption is imputed through multiple imputation.

Here we try to get a sense of how sensitive results are to our assumptions about nursing home consumption. Specifically, instead of assuming average household consumption or using multiple imputation to assign consumption for those in nursing homes, we simply set nursing home consumption to $6,000. We choose this amount because it is roughly equivalent to average annual Supplementary Security Income (SSI) reported in our sample. Many nursing home residents rely on Medicaid which is often tightly connected to SSI. The seventh row in Table 9 provides results from this robustness exercise. With the generally lower nursing home consumption of $6,000, the estimated welfare costs of depression increase slightly. For example, the fully-adjusted measure for the EHRS cohort increased from a benchmark value of 0.148 up to 0.152. This suggests a likely positive correlation between nursing home residence and depression. However, this correlation is not so strong that our overall results are likely to be highly biased by choice of consumption imputation for nursing home residents.

E.4 Non-imputed data

We also checked the sensitivity of results to the use of imputed data. As depression is the focus of our analyses, we begin by estimating results after dropping the approximately 12% of observations with missing CESD score instead of using multiple imputation. Row eight of Table 9 shows that results change very little when dropping missing CESD scores. The following row provides results when dropping all missing data instead of imputing. Consumption is by far the variable with the most missing cases. Recall that consumption data is only available for about 20% of the sample after 2001. As the EHRS cohort was already over age sixty by this time, simulations cannot be run for the cohort as they are all missing initial age sixty consumption. Moreover, only about 15% of the simulation sample remains for the younger cohorts and these observations are unlikely to be representative of the larger older adult population. One of the main benefits of multiple imputation is that the sample remains representative. Moreover, our imputation procedure utilizes a lot related information (e.g., income and wealth) that is thrown out when dropping missing consumption. When excluding all cases with missing data, the depression cost estimates are somewhat lower. For example, the lower bound falls from 0.083 to 0.077 for the LHRS cohort. The decline in the upper bound is somewhat more substantial—for example, falling from 0.144 to 0.102 in LHRS cohort. Nonetheless, there are still substantial estimated costs that decline somewhat over birth cohorts as in the benchmark.

E.5 Depression impacting health

In our benchmark estimation we assumed self-rated health impacted depression but depression did not contemporaneously impact self-rated health. This assumption was made for block identification of the VAR system. We argued this also likely yields conservative estimates of the welfare

costs of depression. Here is we test this argument by instead allowing depression to impact contemporaneous self-rated health while assuming self-rated health does not contemporaneously impact depression. Table 9 provides the results of this simulation. The lower bound estimates change very little while the upper bounds increase somewhat as expected. For example, the upper bound estimate for the EHRS cohort increased from a benchmark value of 0.148 up to 0.189. These results support the argument of conservative cost estimates in our benchmark results.

E.6 Labor hours

In our benchmark we classified retirement as working zero hours. Here we check the robustness of results to defining retirement as not working full-time. In practice, retirement can take many levels in-between, but comparing the full-time definition to our zero hours benchmark should give a sense of how sensitive results might be to intermediate possibilities like partial retirement. We follow the RAND data file and define full-time employment as working at least 1,260 hours annually (35+ hours a week for 36+ weeks a year). We combine data on weekly hours worked and weeks worked per year to estimate annual hours worked. The last row of Table 9 shows that results are largely insensitive to this change in retirement definition. This is perhaps unsurprising given the limited role found for retirement in our benchmark results.

References

Born, B. and Breitung, J. (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews*, 35(7):1290–1316.

De Vos, I., Everaert, G., Ruyssen, I., et al. (2015). Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal*, 15(4):986–1018(33).

Everaert, G. and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, 31(4):1160–1184.

Feeny, D., Furlong, W., and Torrance, G. W. (2018). What were they thinking when providing preference measurements for generic health states? the evidence for hui3. *Health and quality of life outcomes*, 16(1):166.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., Blackwell, M., et al. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Miller, R. and Bairoliya, N. (2022). Health, longevity, and welfare inequality of older americans. *(Forthcoming) The Review of Economics and Statistics*.

Miller, R., Bairoliya, N., and Canning, D. (2019). Health disparities and the socioeconomic gradient in elderly life-cycle consumption. *The Journal of the Economics of Ageing*.

Mullahy, J. (2016). Estimation of multivariate probit models via bivariate probit. *Stata Journal*, 16(1):37–51.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall, London.

Schomaker, M. and Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in medicine*, 37(14):2252–2266.

Wursten, J. et al. (2016). Xtqptest: Stata module to perform Born & Breitung bias-corrected LM-based test for serial correlation. *Statistical Software Components*.