Health, Longevity, and Welfare Inequality of the Elderly*

Ray Miller[†] Neha Bairoliya[‡]

September 13, 2019

Abstract

We estimate the distribution of well-being among the U.S. elderly using an expected utility framework that incorporates differences in consumption, leisure, health, and mortality. We find large disparities in welfare that have increased over time. Incorporating the cost of living with poor health into elderly welfare substantially increases the overall inequality. Disparity measures based on cross-sectional income or consumption underestimate aggregate welfare inequality. Moreover, health is a better indicator of an individual's relative welfare position than income or consumption.

JEL classifications: D63, I14, I31, J14
Keywords: welfare, inequality, elderly, health, morbidity, aging, mortality

*Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number P30AG024409 and by the National Institutes of Health (NIH, Grant No.: 5R01AG048037-02). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.
[†]Corresponding author. Colorado State University. Email: ray.miller@colostate.edu
[‡]University of Southern California

1 Introduction

Inequality has been a subject of great interest to researchers and policymakers on grounds of both fairness and potential consequences.[1] However, the most widely used disparity measures are often based on either income or consumption which provide an incomplete metric of social welfare inequality. Leisure, health, social interactions, political and natural environments, and other factors have all been linked to individual well-being. Moreover, strong socioeconomic gradients have been found in related metrics such as life expectancy (e.g. Kitagawa and Hauser, 1973; Chetty et al., 2016). Given the potential correlation across these factors, a more comprehensive understanding of social welfare and its distribution has significant implications for policy evaluation and prioritization.

We provide a framework to understand how economic circumstances, health, and mortality jointly influence the dispersion of welfare in a given population. Using standard expected utility theory and microsimulations from a forecasting model of life-cycle dynamics, we construct a measure of well-being at the individual level—measured as an ex-ante consumption equivalent. This allows us to analyze the entire distribution of welfare. Our measure is based on comparing expected lifetime utility across individuals of a given age. We incorporate differences in the uncertain evolution of consumption, leisure, health, and mortality over remaining life, providing a more complete measure of well-being than consumption or life expectancy alone.

We apply our methods to estimate the welfare distribution among sixty year olds in the U.S. using data from the Heath and Retirement Study (HRS). As our measure at sixty incorporates individual expectations about outcomes over the entirety of remaining life, it provides a useful single metric of ex-ante elderly well-being. For example, we intend to understand questions such as: how much better do we expect remaining life to be for the median sixty year old in the U.S., compared to the sixty year old who is the worst off? Moreover, how much of the difference in well-being is driven by expected gaps in consumption versus gaps in leisure or health? With these questions in mind, we refer to our measure of well-being as elderly welfare, though strictly speaking we are referring to ex-ante welfare at age sixty.

We conduct our analysis on multiple cohorts in the HRS to examine how the distribution of elderly welfare has changed over time. While income and consumption inequality have increased in the United States over the past three decades, the implications for the distribution of individual welfare are unclear.[2] Growing economic inequality may overstate welfare disparities if, for example, some of the effects are mit-

[1] Fairness is tied to the importance of luck in determining well-being (see Rawls (1971); Dworkin (1985); Roemer (1998)). Inequality has been directly tied to a wide range of outcomes including education, crime, economic growth and mobility, civic engagement, and political influence and polarization. See Kenworthy (2008) for a comprehensive literature review.

[2] See Piketty and Saez (2014); Heathcote et al. (2010); Autor et al. (2008); Katz et al. (1999); Gottschalk et al. (1994) for evidence on income inequality and Attanasio and Pistaferri (2014); Attanasio et al. (2014, 2010); Cutler and Katz (1992) for consumption.

igated through improvements in public health and health equity. The opposite may be true if health gains accrue disproportionately to the financially well-off, making changes in welfare inequality larger than what would be suggested by economic variables alone.

The influence of health on welfare inequality is particularly relevant among the elderly where most health differences are concentrated (Deaton and Paxson, 1998). Compression of morbidity in the elderly U.S. population (Cutler et al., 2013) has been accompanied by evidence of a strong socioeconomic gradient in disability incidence rates in later life (Minkler et al., 2006). More generally, recent evidence suggests that there has been a widening gap in life expectancy and an increase in the socioeconomic gradient of mortality rates.[3] Our focus on the elderly is further motivated by the rapid aging of the U.S. population—more than 20% of people are estimated to be aged 65 and older by 2050 (Colby et al., 2015).

Our approach to welfare analysis can be summarized in three broad steps. First, we propose a welfare model for evaluating individual well-being using expected utility theory. We extend the standard theory to allow for the importance of health by borrowing from the literature on quality-adjusted life years (QALYs). This framework accounts for the impact of consumption, leisure, health, and mortality on well-being and provides a simple analytic decomposition of the contribution of each channel. Next, we use HRS data to estimate a forecasting model of the joint evolution of outcomes over the elderly life-cycle. This forecasting model can be conceptualized as a panel vector autoregression (VAR). While the panel VAR falls short of a fully specified structural model, the equations can be viewed as approximations of the underlying decision rules mapping state variables to individual choices. Finally, using the forecasting model and data from HRS respondents as initial conditions, we repeatedly simulate potential outcome paths. These paths are embedded in the welfare model to compute an ex-ante measure of well-being for each individual in our sample at age sixty.

We measure welfare of a *given* individual in consumption equivalents; how much consumption would have to increase/decrease across the remaining lifetime of a reference person to yield the same expected level of utility as that obtained by the current and potential future outcome bundles of the *given* individual. In our empirical application, we compare consumption equivalents computed for each individual at the age of sixty using the individual with the median utility ranking as our reference person. This measure incorporates all expected inequalities in outcomes across individuals over their remaining lives. It also accounts for welfare costs of uncertainty in outcomes after sixty, providing a useful metric of ex-ante elderly welfare.

The most salient findings of our analysis can be summarized as follows:

1. There is substantial variation in the ex-ante welfare of individuals at age sixty. The Gini coefficient for consumption-equivalent welfare in our benchmark cohort is 0.54. Those at the ninetieth percentile of the welfare distribution have 16 times higher welfare than those at the tenth percentile.

[3] See, for example, Chetty et al. 2016; Currie and Schwandt 2016; National Academies of Sciences, Engineering, and Medicine 2015; Pijoan-Mas and Ríos-Rull 2014; Meara et al. 2008

2. The largest drivers of elderly welfare inequality are health and mortality gaps followed by gaps in consumption. Differences in leisure play a comparatively minor role.

3. Welfare inequality among the elderly has increased over time. Compared to the cohort of individuals reaching age sixty between 1992-96, the welfare Gini rose 11% for those reaching sixty between 1997-2001, 18% for 2002-07, and 23% for 2008-13.

4. Ignoring dynamic uncertainty and the persistence in outcomes over the life-cycle greatly underestimates welfare inequality. The Gini of age sixty flow utility is only 43% of that based on our dynamic welfare measure.

A key implication of our results is that cross-sectional distributions of income and consumption underestimate aggregate welfare inequality at age sixty. This occurs for two primary reasons. First, cross-sectional measures ignore dynamic uncertainty and the persistence of inequality over remaining life. Second, there is a positive correlation between health and consumption, implying those with high consumption also enjoy better health and longer lives on average. However, even in cases where economic outcomes provide a reasonable approximation to aggregate welfare inequality, our results suggest they may still provide a poor ranking of individual well-being. For example, the rank correlation between consumption and welfare is a relatively modest 0.57 for our benchmark cohort. Moreover, we find cross-sectional health utility at age sixty to be a better predictor of remaining lifetime welfare rank, despite the fact that it drastically underestimates aggregate welfare inequality.

Our paper builds on a large body of work attempting to extend measures of welfare beyond income (see Fleurbaey (2009) for an extensive review). Recent examples include Becker et al. (2005) who combine national income and expected longevity in a utility framework to examine the changes in cross-country inequality over time. Fleurbaey and Gaulier (2009) extend this work by examining level differences across countries and incorporating leisure, health-adjusted life expectancy, and aggregate inequality. Our welfare framework builds on recent work by Jones and Klenow (2016) who construct an alternate cross-country measure of economic well-being. Our work is different from these papers along many dimensions. Most notably, by using longitudinal data and estimating the joint dynamic process of outcomes, we are able to construct welfare at the individual as opposed to aggregate level. Moreover, we explicitly allow for health to affect individual welfare by mapping subjective and objective measures directly into utility. We also focus on the U.S. elderly and examine inequality evolution over birth cohorts as opposed to cross-sectional changes over time.

Our paper is also more broadly tied to the literature measuring economic inequality at older ages. Using HRS data, Hurd and Rohwedder (2007) find that income based poverty rates underestimate economic well-being of the elderly compared to consumption based measures. Crystal and Shea (1990) argue that despite the increased presence of social safety nets at older ages, economic inequalities are exacerbated with aging due to the accumulation of "economic advantages and disadvantages" over the entire life-course. Using longitudinal data, Crystal and Waehrer (1996) likewise find that income inequality rises within cohorts as they age, but also document considerable mobility in relative income position. More recently, Bosworth et al. (2016) document that income inequality has increased for the elderly over the past three decades, though more slowly than among the non-elderly perhaps due to wider availability of social safety nets at older ages. Finally, in Miller et al. (2019), we use a simplified version of our panel VAR model to estimate the education gradient in elderly lifetime consumption and document the importance of health and educational differences in explaining overall consumption inequality. In that paper we estimate that 11-12% of the education gradient in remaining lifetime consumption at age sixty could be closed by eliminating elderly health differences. This paper further contributes to this line of research by providing estimates of elderly welfare that incorporates consumption, leisure, and health into a single measure rooted in economic and public health theory.

Several limitations to our approach warrant mentioning at the outset. First, we do not explicitly account for morbidity spillover effects such as the cost of caregiver time and the numerous costs associated with the loss of a spouse. Likewise, we abstract from other potentially important inputs into elderly welfare such as social interactions, bequests, and end-of-life care. Second, we estimate welfare based on common preferences. Considering heterogeneity across individuals' preferences could reduce the welfare costs of inequality along some components. Finally, we assume institutions and relevant policies remain fixed moving forward and past trends in elderly health, retirement, and consumption continue into the future. For example, significant anticipated changes to Social Security or Medicare programs or exponential advances in medicine could alter our welfare measure, particularly for the younger cohorts we study.

The remainder of the paper is organized as follows. Section 2 outlines our welfare model while Section 3 provides details of the data and empirical methods used in our analysis. Section 4 discusses our welfare results including robustness to alternate modeling assumptions. Finally, section 5 provides concluding remarks.

2 Welfare model

Our welfare concept aims to compare well-being across individuals of a given age j. These individuals may differ along many dimensions due to their childhood environment, education, occupation, previous health behaviors, and numerous other factors. We define individual welfare based on observed outcomes at age j and the potential realization of outcomes in the future based on these multi-dimensional differences. Although individuals are heterogeneous, we make welfare comparisons through a common preference specification. These preferences are defined by an expected lifetime utility at age j for individual i given by:

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

where c is consumption, l leisure, h health, and ψ is a survival indicator. Expectations are taken with respect to the uncertainty in the evolution of consumption, leisure, health, and mortality after age j.

We use a consumption-equivalent variation (EV) measure to quantify welfare differences across individuals. This approach requires choosing a reference person as the basis for our welfare comparisons. Welfare λ_{ij} is then the proportion of the reference individual's consumption that must be maintained at every age starting from j (in all possible realizations of the world and holding health and leisure fixed) that would make them indifferent to facing the current and potential future outcome bundles of individual i. For example, if person i is relatively poor and unhealthy, we may have a welfare measure $\lambda_{ij} = 0.3$. This implies the reference individual would be ex-ante indifferent between maintaining 30% of their own consumption every period from age j or receiving the outcome bundle of person i at age j and facing person i's stochastic evolution of consumption, leisure, health, and mortality profiles moving forward. As this measure is based on potential outcomes over the remaining life, it encompasses the cross-sectional inequality in outcomes at age j as well as the likelihood of persistence and emergence of inequalities in future outcomes.

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

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

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

$$U_{mj}\left(\lambda_{ij}\right)=U_{ij}\left(1\right) \tag{1}$$

where U_{mj} refers to the expected lifetime utility from the outcome bundles of the reference individual (m).

In our benchmark model we assume that preferences are additively separable between consumption and leisure and non-separable between health and the consumption-leisure composite. The flow utility takes the following form:

$$u\left(c,l,h\right)=\phi\left(h\right)\left[\bar{u}+log\left(c\right)+\nu\left(l\right)\right]. \tag{2}$$

The additive separability between consumption and leisure allows for a simple decomposition of welfare effects but we check the robustness of our results to more general preferences. Our treatment of health in the utility function follows from a large literature on quality-adjusted life years (QALYs) going back to the works of Klarman and Rosenthal (1968); Fanshel and Bush (1970); Torrance et al. (1972); and Zeckhauser and Shepard (1976). The central assumption is that health utility is a function of both the length and quality of life. The QALY literature provides a framework to combine these two aspects of health in a single index. Accordingly, preferences over health are chosen such that period utility from whatever is regarded as the best possible health state or "full health" equals one. In our framework, health function $\phi(h) \in [0,1]$ scales the utility from consumption and leisure such that $\phi(h) = 1$ represents utility in the perfect health state and $\phi(h) = 0$ represents the dead state. At the same time, $\psi\phi(h)$ represents a measure of QALYs. For instance, $\psi\phi(h) = 1$ represents a year of life with no morbidity; a single QALY.

Under preferences given in (2), welfare condition (1) may be rewritten:

$$log(\lambda_{ij}) = \tilde{\psi}\left(U_{ij}(1) - U_{mj}(1)\right) \tag{3}$$

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

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

Let u_{ia} denote flow utility unadjusted for health at age a given outcome bundles i: $u_{ia} = \bar{u} + log(c_{ia}) + \nu(l_{ia})$. Moreover, denote the expected value conditional on survival with subscript ψ: $E_\psi[u_{ia}] = E[u_{ia} \mid \psi_{ia} = 1]$. Adding and subtracting

$$\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[E[\psi_{ia}\phi(h_{ia})]E_\psi[u_{ia}] + E[\psi_{ma}\phi(h_{ma})](E_\psi[u_{ia}] - E_\psi[u_{ma}])\right]$$

from the right-hand side of (3) and rearranging terms yields the following additive decomposition of welfare:

$$\begin{aligned}
log(\lambda_{ij}) =& \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[(E[\psi_{ia}\phi(h_{ia})] - E[\psi_{ma}\phi(h_{ma})])E_\psi[u_{ia}] + \Phi\right] && QALE \quad (4)\\
&+ \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E[\psi_{ma}\phi(h_{ma})](E_\psi[log(c_{ia})] - E_\psi[log(c_{ma})]) && Cons. \quad (5)\\
&+ \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E[\psi_{ma}\phi(h_{ma})](E_\psi[\nu(l_{ia})] - E_\psi[\nu(l_{ma})]). && Leisure \quad (6)
\end{aligned}$$

where

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

The first term in (4) is the the difference in quality-adjusted life expectancy weighted by how much a healthy life year is worth—the expected flow utility from outcome bundles of individual i. The Φ term is an adjustment for the uncertainty of health and mortality over the life-cycle. Combined these provide the approximate contribution to welfare of health and mortality relative to the reference individual. The final two terms give the expected utility difference in consumption (5) and leisure (6) weighted by the quality-adjusted life expectancy of the reference individual. These provide the approximate contributions of consumption and leisure to welfare.

3 Data and methods

This section details the data used in our analysis, the dynamic forecasting model used to simulate outcomes, and calibration of parameters used in the welfare model.

3.1 Data

The HRS is a longitudinal panel study surveying individuals in the U.S. over the age of fifty and their spouses on a biennial basis. The study consists of seven primary birth cohorts—the early HRS cohort (born 1931-1936), late HRS cohort (born 1937-1941), AHEAD cohort (born before 1924), Children of Depression (born 1924-1930), War Babies (born 1942-1947), early Baby Boomers (born 1948-1953), and mid-Baby Boomers (born 1954-1959). The core survey was conducted for the early and late HRS cohorts every other year starting from 1992 with the other cohorts added periodically over subsequent waves of the survey. As such, a model period corresponds to two calendar years and individuals are grouped in two-year age intervals. To be clear, our welfare measure is constructed for each individual at age sixty *or* sixty-one, though our model makes no distinction between the two and for brevity we refer to this as welfare at age sixty.

3.1.1 Health outcomes, labor supply, and other characteristics

We use the RAND HRS data file (v.P), available through the HRS website, to obtain data on labor supply, health, and mortality from 1992 to 2014.[4] We also utilize other individual characteristics including age, education, gender, race, birth cohort, region, and occupation. We define leisure as time not spent in the labor market. As our empirical focus is on individuals nearing the end of working life, we limit labor considerations to the extensive margin by assuming an absorbing retirement decision.[5]

[4] Data available at `http://hrsonline.isr.umich.edu`.

[5] An extended model may include the intensive margin, partial retirement, and/or reentry into the workforce but this comes with additional model complexity. Moreover, we find relatively small effects of leisure on welfare in our empirical analysis and retirement is likely to be the first-order leisure effect for this age group.

We define retired individuals as those reporting less than 500 annual hours of work in the current or any previous survey wave.[6]

We measure health h as a vector of objective morbidity indicators and a subjective health assessment. Our primary morbidity measures 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). Difficulty with ADLs are a commonly used health metric among the elderly and include activities such as walking across the room, bathing, and getting dressed.

As a final health measure we include 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). We hypothesize that people have private information about their health over and above disease diagnosis. This may reflect the severity of the disease or additional information on other health conditions. Finally, mortality data is also available from the HRS, with participants' death dates either reported by spouses or coming from an exit interview.

3.1.2 Health utility

We use the Health Utilities Index Mark 3 (HUI3) instrument as the conceptual basis of our health utility function $\phi(h)$ (see Horsman et al. (2003) for a detailed discussion on the HUI3). This choice is motivated by two main features of the HUI3. First, it provides a comprehensive description of health status that has been shown to be responsive to changes in health over time (Barr et al., 1997; Furlong et al., 2001; Blanchard et al., 2003). Second, it provides a direct estimate of QALYs which is our preferred measure of health in this analysis.

The HUI3 questionnaire was included as a module for a subset of approximately 1,200 of the participants in the 2000 wave of the HRS with the associated utility scores available through the HRS website.[7] We use the HUI multi-attribute utility score (*hui3ou*) in our analysis (see Furlong et al. (1998); Feeny et al. (2002) for details on construction). It is a health related quality of life measure where death is defined by a score of zero and perfect health by a score of one.[8]

---

[6] We combine data on weekly hours worked and weeks worked per year to estimate annual hours worked.

[7] Researcher contribution file HUI3 (v.1.0).

[8] Negative scores are possible and represent health states that are worse than death.

3.1.3 Consumption

Consumption data comes from the Consumption and Activities Mail Survey (CAMS). Starting in 2001, the CAMS was sent to a random sub-sample of HRS participants during off years of the core survey. The CAMS collected household spending data on durables, nondurables, transportation, and housing. We use cleaned RAND 2015 CAMS data file (v.2), which contains a constructed estimate of total household consumption derived from the available spending data from 2001-2013.[9] Broadly, household consumption was derived by estimating the per-period "usage" from consumer durables, automobiles, and housing expenditures using a similar method as proposed by Hurd and Rohwedder (2007).[10] We subtract out-of-pocket health spending to create an adjusted measure of household consumption.[11] We use this adjusted household consumption divided by the number of household members as our individual consumption measure. 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.[12]

A serious challenge in our analysis is missing consumption data. CAMS data is only available for approximately 20% of HRS participants for the years 2000-2012. However, closely related data is available across all survey waves including detailed measures of wealth and income. We use this additional information to address missing consumption data by using the multiple imputation method proposed by Honaker and King (2010) for cross-sectional time-series data (see appendix A for details). We also conduct robustness tests on our welfare results limiting the amount of imputed data used.

3.2 Forecasting outcomes

Our expected lifetime utility approach to welfare requires the knowledge of all *possible* life-cycle path realizations for an individual. As only the realized outcome path is observable in any longitudinal data set, we estimate a dynamic *forecasting* model to approximate the joint evolutionary process of consumption, leisure, health, and mortality over time. Our forecasting model can be conceptualized as a panel vector autoregression (VAR) of order p. In our application to the U.S. elderly population, we use the model to repeatedly simulate potential outcome paths for each individual, given a set of initial (age sixty) conditions. Here we discuss the primary features of the forecasting

[9] Data and consumption construction details available at `http://hrsonline.isr.umich.edu`.

[10] The annual service flow for durables was roughly estimated as $C \times p$ where C was the total cost and p the probability of purchase. Adjustments for interest payments, depreciation, and insurance costs were done for estimating transportation consumption. The consumption of housing was estimated as the sum of the rental equivalent of the owned house, property tax, homeowners insurance, and any additional rent payments.

[11] Health spending includes health insurance, medication, health services, and medial supplies. We use the CPI-U to convert all waves to 2010 dollars.

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

VAR model and identifying assumptions. Appendix B lays out all specification and estimation details.

Our basic forecasting model of health, labor, and consumption is illustrated in Figure 1. At the beginning of each time period, morbidity status is updated based on (correlated) random shocks and exogenous characteristics of an individual. Given these morbidity conditions (and other exogenous characteristics), an individual then updates their self-rated health status. Morbidities and self-rated health then affect current period labor supply (i.e. the retirement decision). Finally, health outcomes and labor supply impact current period consumption. This specification is consistent with evidence that health affects the retirement decision (Currie and Madrian, 1999), that consumption declines with retirement (Hall, 2009), and that health impacts economic outcomes, particularly at older ages (Smith, 1999).[13] Note that we posit each of the morbidity states to contemporaneously influence labor supply and consumption both directly and through changes in self-rated health. For example, diabetes may affect an individual's self-rated health status which in turn may lower contemporaneous consumption. However, diabetes may also influence consumption independently of changes in self-rated health (e.g. due to higher anticipated future medical costs). Similarly, self-rated health may influence consumption directly or indirectly by impacting labor supply.

Labor supply and all health outcomes are permitted to influence the probability of survival to the following period of life. Moreover, we allow current outcomes to influence the evolution of outcomes moving forward (conditional on survival) through the inclusion of general lagged effects. For example, retirement today may influence self-rated health the following period. Moreover, lagged effects allow, for example, the recent onset of heart disease to alter consumption more than if an individual has been living with a heart disease diagnosis for an extended period of time.

It is important to recognize that we have opted to model life-cycle dynamics as a statistical process to be estimated directly from the data. Alternatively, modeling explicit dynamic maximization of lifetime utility would allow for a richer set of counterfactual policy analyses when outcomes are endogenous. The major difficulty with the latter approach arises from specifying and solving an intertemporal structural model of endogenous savings, labor supply, and multivariate health behaviors and investments. Convincingly mapping behaviors and investments to the endogenous and stochastic evolution of health and labor productivity is particularly challenging. Given our focus on inequality, incorporating sufficient heterogeneity across individuals is also of first order importance and can be limited by computational concerns in a structural approach. As our goal is to construct a welfare measure broadly reflective of observed data, we believe a statistical forecasting approach is appropriate. In contrast, counterfactual policy experiments (e.g. Medicare reform) would likely require a credible estimation of underlying structural parameters. So while our model is not a fully specified structural

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

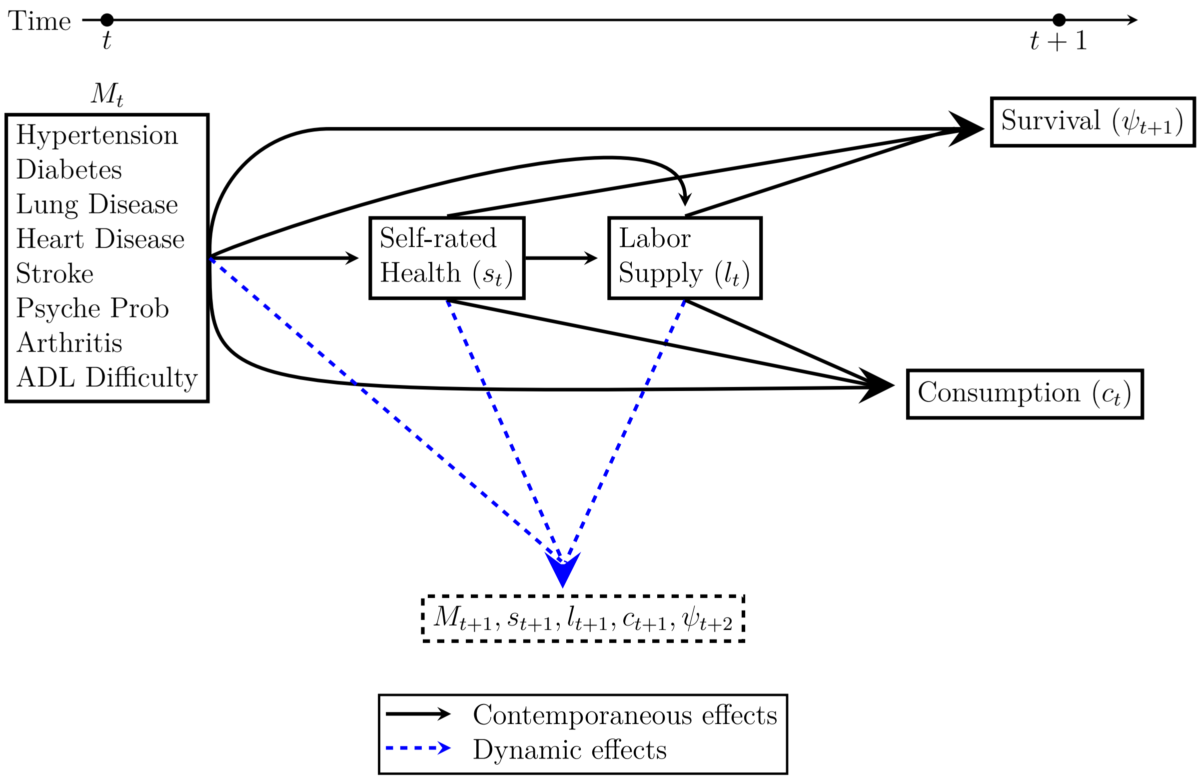

Figure 1: Forecasting model with one period lag

model, it can be viewed as approximating the underlying decision rules mapping state variables to individual choices under a realized policy regime.

3.2.1 Simulations

Equipped with our forecasting model, our empirical analysis involves three steps.

1. We use data from the HRS to estimate the parameters of the forecasting model. Here we use data on all individuals aged fifty and older from all available waves of the HRS from 1992-2014. See appendix B.2 for details.

2. Using the parameter estimates and age sixty data as initial conditions, we repeatedly simulate remaining life-cycle paths for mortality, health, consumption, and leisure for a sub-sample of the HRS respondents. This simulation sample includes all individuals with age sixty data and requisite lagged data for simulations. See appendix B.3 for details.

3. We embed the simulated data within our expected utility framework to construct a measure of ex-ante welfare at age sixty for each individual in our simulation sample.

Our choice of age sixty for welfare comparisons is primarily driven by empirical considerations. It provides a large enough sub-sample for analysis across four birth cohorts included in the HRS (the EHRS, LHRS, War Babies, and early Baby Boomers.) after accounting for the sampling design of the survey and lagged data requirements of our dynamic model. We compute the distribution of elderly welfare within a birth cohort as well as compare welfare distributions across cohorts to examine how it has changed over time. It is important to highlight that welfare differences across cohorts will arise for two reasons. First, the distribution of age sixty initial conditions will differ across cohorts. Second, the forecasting model includes a linear calendar year trend and independent cohort effects on all outcomes.

3.3 Calibration of welfare model

Analysis using our welfare model requires calibration of preference parameters. These include parameters of the functions $\phi(h)$ and $\nu(l)$ mapping health states and leisure into flow utility. Additional parameters include the discount rate β and flow intercept $\bar{u}$. Here we detail our calibration strategy and estimates.

We assume health utility depends linearly on our health state vector: $\phi(h_t) = \gamma h_t$. However, we bound $\phi(h_t) \in [0, 1]$ to be consistent with our QALY framework. The vector of utility weights γ is estimated by regressing the HUI3 utility score on self-rated health and all morbidity indicators. Table 1 provides the linear regression results.[14] Self-rated health has a strong and highly significant positive association with utility. For example, moving from poor health (the reference category) to fair health improves flow utility by 22.6 percentage points. Moving from very good to excellent health increases utility only 1.8 percentage points. Conditions such as hypertension, diabetes, and cancer have little independent effect on health utility after controlling for their association with self-rated health and other co-morbidities. Other morbidities such as stroke and arthritis have larger (and statistically significant) independent negative effects. The most influential morbidity indicator is difficulty with ADLs, which lowers utility an estimated 16.1 percentage points.

Leisure is normalized to one for retired individuals. We assume workers supply 2,000 annual hours to the labor market and set associated leisure to $0.66 = 1 - (2000/5,840)$ where 5,840 = 16 hours a day $\times$ 365 days. 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 (the elasticity of labor supply with respect to wage, holding the marginal utility of consumption fixed). Empirical studies of the Frisch elasticity vary considerably, with estimates ranging from 0.5 to nearly 2 (Chetty, 2012; Hall, 2009). We follow Jones and Klenow (2016) and choose a benchmark value of $\epsilon = 1$ while testing the sensitivity of our results to alternate values. Likewise, we choose a discount factor $\beta = 0.98$ corresponding to an annual discounting of one percent. Note that there is additional implicit discounting in the expected utility framework due to mortality.

[14] Results are insensitive to use of a Tobit regression.

Table 1: Estimated health utility weights (γ)

Measure	Weight	SE
Self-rated health		
Fair	0.226	0.026
Good	0.313	0.026
Very good	0.403	0.027
Excellent	0.421	0.031
Hypertension	0.003	0.012
Diabetes	-0.001	0.018
Cancer	0.010	0.017
Lung disease	-0.020	0.022
Heart disease	-0.032	0.015
Stroke	-0.076	0.022
Psych problem	-0.073	0.020
Arthritis	-0.062	0.012
Diff with ADL	-0.161	0.016
Constant	0.517	0.028

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

The standard first-order condition for the labor-leisure decision equates the marginal cost and benefit of leisure: $u_l = w\,(1-\tau)\,u_c$, where w is the wage rate and τ is the marginal tax rate. Given our functional forms, the implied disutility weight on labor supply $\theta = w\,(1-\tau)\,(1-l)^{-1/\epsilon}/c$. Using earnings and hours worked data from the HRS and a marginal tax rate of 0.38 from 2002 (Barro and Redlick, 2011), we calculate the implied θ for each working sixty year old in the sample. Selecting the median of these values yields our benchmark $\theta = 8.37$.[15]

Finally, we set the intercept in flow utility $\bar{u}$ so that the median value of remaining life for sixty year olds in our simulation sample is \$50,000 per QALY.[16] In a review of the literature, Ryen and Svensson (2015) estimate mean and median values across studies of approximately \$98,000 and \$32,000. Traditional values in the U.S. often range from \$50,000 to \$100,000 (Kaplan and Bush, 1982). Using \$50,000 as our benchmark and normalizing consumption to thousands of 2010 dollars gives $\bar{u} = -0.34$.[17]

[15] As noted by Jones and Klenow (2016), this calibration strategy implicitly invokes wedges (i.e. labor market frictions) to explain individual deviations from the static first-order condition. Note also that each imputed data set was calibrated separately and received unique values for θ and $\bar{u}$. The values reported here are the average across the imputed data sets.

[16] 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 $VOL_j = U_{ij}\,(1)\,c_j$.

[17] With log consumption utility, it is possible for an individual to obtain negative expected remaining lifetime utility in this framework but this occurs for less than 0.4% of our sample. Welfare results are qualitatively insensitive to excluding these individuals.

4 Welfare results

Our simulation sample for welfare analysis includes four birth cohorts—early HRS (EHRS), late HRS (LHRS), War Babies, and Baby Boomers. We use the EHRS cohort as our benchmark group as it is the earliest of the four and contains the longest panel of available data. The reference person for all welfare calculations is the individual with the age sixty initial conditions that yield the median expected lifetime utility within the EHRS cohort. The same reference person is used to calculate welfare for the later LHRS, War Babies, and Baby Boomers cohorts. This approach allows for direct comparison of welfare across cohorts as the reference person is held fixed. At the end of this section, we check robustness of welfare estimates to the choice of reference individual as well as other modeling assumptions.

4.1 Elderly welfare inequality

We begin by examining the distribution of our consumption-equivalent measure of welfare across the sample of sixty year olds from the EHRS cohort. The first row in Table 2 provides different summary measures of welfare inequality in our fully specified "benchmark" model. In order to assess the importance of poor health on our welfare calculations, we also provide measures where we exclude the utility cost of less than perfect health from preferences (i.e. $\phi(h) = 1\ \forall h$). The results from this model are labeled as "no health" in the table.

Table 2: Summary measures of welfare inequality at age sixty for initial HRS cohort

Measure	Gini	10/50 ratio	90/50 ratio	ρ
Benchmark λ	0.544	0.234	3.774	
No health λ ($\phi(h) = 1$)	0.453	0.335	2.831	0.972

Notes: Estimates use base year respondent analysis weights. No health measure removes health from flow utility. Spearman's rank correlation between the two welfare measures denoted by ρ.

There is substantial variation in welfare across individuals—the benchmark Gini coefficient is 0.544. Moreover, welfare at the tenth percentile of the distribution is only 23% of the median welfare while that of the ninetieth percentile is over 3.7 times higher than the median. Ignoring the utility costs of poor health largely preserves the rank ordering of welfare ($\rho = 0.972$) but significantly under-estimates the inequality—the Gini coefficient is under-estimated by about 17%. Moreover, this morbidity bias occurs at both ends of the distribution. For example, relative to our benchmark measure, welfare at the tenth percentile increases to 33% of the median while that at the ninetieth falls to under three fold.

The difference in welfare measures between the two models suggests substantial and varied individual utility costs of living with poor health and morbidities among the elderly. Figure 2 plots remaining life expectancy at age sixty against the ratio of QALE to life expectancy for each individual in the EHRS cohort.[18] The positive correlation implies those with higher life expectancy also expect better health over remaining life. For example, those with a remaining life expectancy of 30 years have a quality-adjusted life expectancy of about 24 *healthy* life years—or about 80%. In contrast, those at the bottom end of the distribution expect greater utility losses from poor health (with considerably more variability). For example, those with a remaining life expectancy of 10 years expect anywhere from about 3 to 7 quality-adjusted life years.

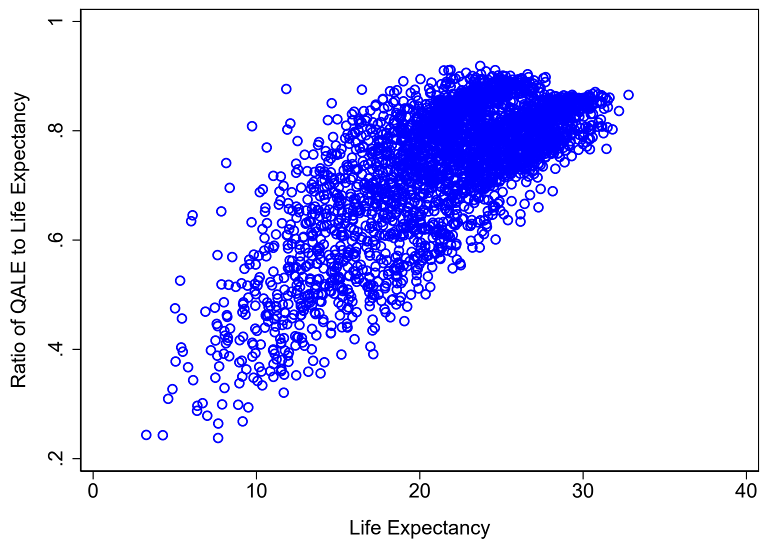

Figure 2: Life expectancy and quality-adjusted life expectancy (QALE) at age sixty

We turn now to the welfare impact of QALE relative to consumption and leisure. Table 3 shows welfare and its decomposition averaged within each decile of the welfare distribution. Consumption and QALE are the largest source of welfare loss for those in the bottom of the distribution. Low consumption and quality-adjusted life expectancy costs the bottom decile an average of 81.8 and 106.9 log points in welfare relative to the median individual. Higher leisure adds a comparatively modest 6.7 log points to welfare. In contrast, the top of the distribution experiences the highest consumption and quality-adjusted life expectancy, though marginally less leisure.

Recall that welfare for each individual is estimated using potential outcome bundles of a reference person. While the reference individual has the median expected lifetime utility by definition, their expected levels of consumption, leisure, and quality-adjusted life expectancy could be somewhat arbitrary. Comparing average log point gaps across deciles provides a more robust examination of the strength of the relative components across the welfare distribution. For example, the log point gap in consumption between

[18] Life expectancy at age j defined by $E\left[\sum_{a=j}^{J} \psi_a\right]$ and QALE as $E\left[\sum_{a=j}^{J} \psi_a \phi\left(h_a\right)\right]$.

the highest and lowest decile is $169.6 = 87.8 + 81.8$. The analogous gap for QALE is 198.6 (and -7.2 for leisure). In other words, compared to consumption, QALE differences explain about 17% more of the welfare gap between the highest and lower deciles. Overall, decomposition results suggest the strongest driver of welfare inequality are differences in health and mortality followed by consumption differences.

Table 3: Mean welfare by decile of distribution

			Decomposition		
Welfare Decile	Mean λ	Mean log λ	Consumption	Leisure	QALE
Lowest	0.171	-1.820	-0.818	0.067	-1.069
2nd	0.300	-1.227	-0.469	0.039	-0.797
3rd	0.458	-0.792	-0.310	0.027	-0.509
4th	0.647	-0.441	-0.184	0.013	-0.270
5th	0.871	-0.142	-0.060	0.005	-0.087
6th	1.158	0.143	0.053	0.001	0.089
7th	1.544	0.429	0.185	0.003	0.241
8th	2.085	0.727	0.312	-0.002	0.417
9th	3.026	1.090	0.468	-0.004	0.626
Highest	7.248	1.789	0.878	-0.005	0.917

Notes: Estimates use base year respondent analysis weights.

In order to see the pattern of health and consumption driving the welfare gaps, Figure 3 plots quality-adjusted life expectancy against annual consumption at age sixty for individuals in the top and bottom deciles of the welfare distribution. A majority of sixty year olds in the highest decile had a QALE of over 20 years. There was more substantial variation in annual consumption in the group with values ranging from around \$20,000 to more than \$100,000. This suggests health as the major driver of welfare at the top end of the distribution. A majority of individuals in the lowest decile had annual consumption under \$30,000 and a QALE of less than 10 years. However, some had relatively higher QALE but low welfare due to very low consumption.[19]

Finally, Figure 4 plots average expected life-cycle profiles for select welfare deciles to gain a sense of the differences across individuals. Consumption, leisure, and health gaps are largest at age sixty and gradually decline as individuals age. However, substantial gaps remain for consumption and health even among individuals who survive into their late nineties. Over the entire remaining life-cycle, the gaps in consumption are relatively small between the first and fifth deciles compared to the much higher average consumption in the top decile. In contrast, the health and leisure gaps are substantially larger between the bottom and middle deciles with a smaller difference between the middle and the top.

[19] To be clear, there is a strong positive correlation overall between consumption and QALE. The negative relationship *within* deciles in Figure 3 is due to restricting to a given decile of the welfare distribution. For example, one only remains in the lowest decile with high QALE if consumption is very low.

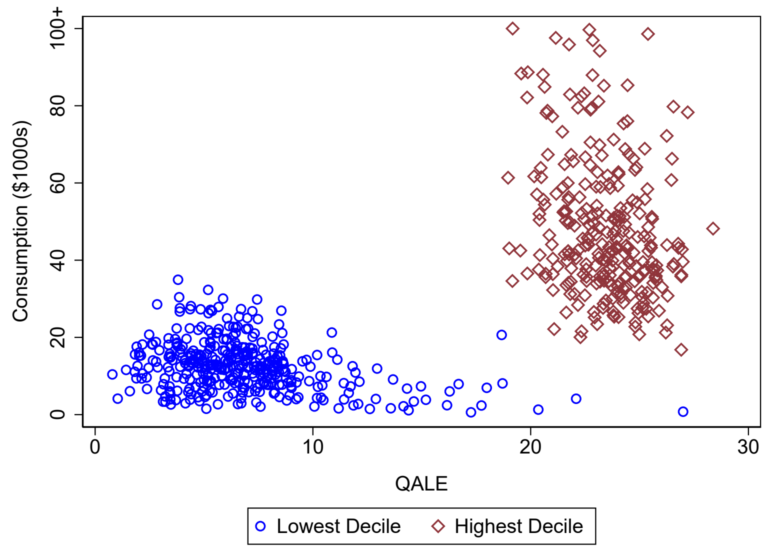

Figure 3: Age sixty consumption and QALE by decile of welfare distribution

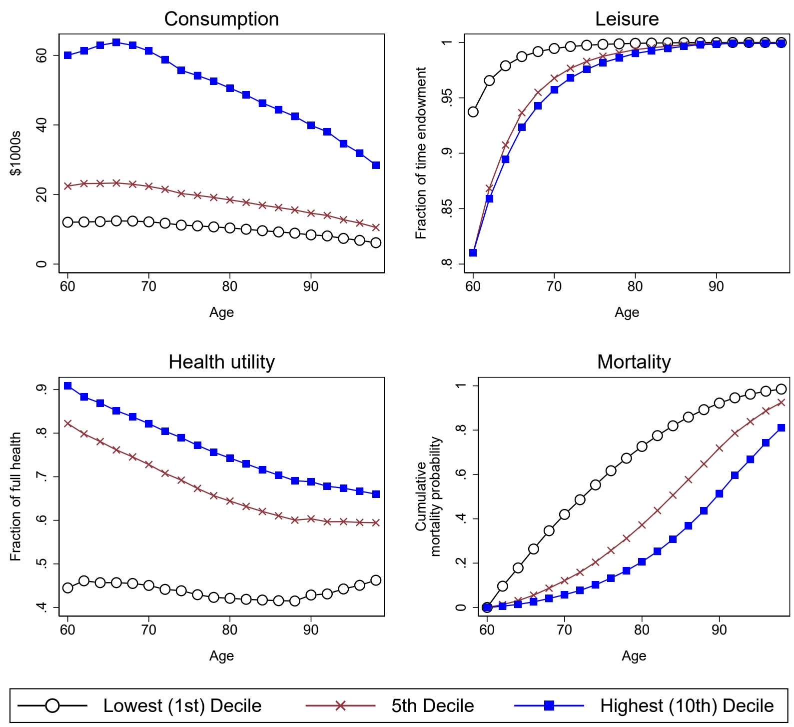

Figure 4: Average life-cycle profiles by select deciles of welfare distribution

Notes: Consumption, leisure, and health profiles are expected values conditional on survival.

4.2 Welfare over cohorts

We next examine how welfare has changed across time by comparing individuals in the EHRS cohort (reaching age sixty between 1992-1996) to the LHRS (1997-2001), War Babies (2002-2007), and Baby Boomers (2008-2013). Recall that aggregate welfare differences across cohorts stem from two sources—the distribution of age sixty initial conditions, and a time trend and cohort specific intercept in all modeled outcome forecasts.

Table 4 provides summary welfare measures for each cohort. There was an increase in both median and mean welfare over the first three cohorts. However, while mean welfare also increased between War Babies and Baby Boomers, median welfare declined, suggesting an unequal shift in the welfare distribution. The reported disparity measures provide further evidence of a sustained increase in welfare inequality over all four cohorts. Relative to the EHRS, the welfare Gini rose 11.4% for LHRS, 18.2% for War Babies, and 23.9% for Baby Boomers.[20] At the bottom end of the distribution, the welfare of the tenth percentile declined from 23.4% of the median in the EHRS to 19.6% among Baby Boomers. The welfare at the top also pulled further away from the center—the 90/50 ratio increased more than 50% between the EHRS and Baby Boomers.

Table 4: Summary welfare measures by cohort

Cohort	Median λ	Mean λ	Gini	10/50 ratio	90/50 ratio
EHRS	1.000	1.750	0.544	0.234	3.774
LHRS	1.067	2.283	0.606	0.210	4.667
War Babies	1.122	2.688	0.643	0.196	5.159
Baby Boomers	0.991	2.799	0.674	0.196	5.727

Notes: Estimates use base year respondent analysis weights.

Table 5 provides the decomposition of mean log welfare for each of the four cohorts. The gain in mean welfare for the LHRS over the EHRS cohort was driven by an average increase of 4.5 log points from consumption ($5.0 - 0.5$) and 5.0 log points from quality-adjusted life expectancy ($4.4 + 0.6$). There was also a 0.2 log point decline in average welfare due to delayed retirement and consequently less leisure ($1.4 - 1.2$). Significant improvements in health and longevity increased mean welfare for War Babies over the LHRS cohort despite a small decline from consumption. Comparing Baby Boomers to War Babies, the decomposition reveals that while health and longevity continued to improve—average QALE contribution to welfare rose one log point ($9.6 - 8.6$)—consumption declines cost average welfare 9.7 log points ($7.7 + 2.0$). The consumption decline is presumably driven by the timing of the 2008 recession, which hit when Baby Boomers were in their late fifties. Moreover, as we will show below, while the *average* log consumption declined for Baby Boomers (and to a lesser

[20]The pattern of increased inequality over time also holds when examining two-year birth cohorts (see appendix Figure 18).

extent War Babies), this was entirely due to losses at the bottom end of consumption distribution.

Table 5: Welfare decomposition by cohort

Cohort	Mean log λ	Decomposition		
		Consumption	Leisure	QALE
EHRS	-0.025	0.005	0.014	-0.044
LHRS	0.068	0.050	0.012	0.006
War Babies	0.115	0.020	0.009	0.086
Baby Boomers	0.022	-0.077	0.003	0.096

Notes: Estimates use base year respondent analysis weights.

Figure 5 provides the distribution of log welfare, expected remaining lifetime consumption, life expectancy, and QALE at age sixty across cohorts for closer examination. The welfare distribution became flatter and more skewed over time demonstrating the rise in welfare inequality. Compared to the EHRS cohort, welfare improved for the top end of the LHRS and War Babies distributions but declined slightly for the bottom end. The distribution of welfare for Baby Boomers was similar to War Babies at the top but a fatter left tail implies a more substantial welfare decline for the bottom end. Expected remaining lifetime consumption follows a similar pattern as the welfare distribution with initial gains concentrated at the top end for the LHRS and War Babies followed by some declines concentrated in the bottom end for Baby Boomers.

While life expectancy has shown broad improvements over time, the distribution has increased in skewness. This implies mortality gains have disproportionately benefited those in the top of the distribution. This is consistent with the existing evidence of increasing socioeconomic gradients in mortality. When adjusting life expectancy for quality of health, the distribution becomes more disperse. The fattening left tail in QALE relative to previous cohorts shows that life expectancy gains are even outweighed by health losses at the bottom end of the distribution. Overall, cohort results demonstrate an increase in welfare inequality driven primarily by a combination of increasing gaps in consumption and quality-adjusted life expectancy.

4.3 Comparison with other measures of well-being

Our welfare measure incorporates inequality of various components of well-being into a single metric. Moreover, our measure captures the static welfare effect of each component at age sixty as well as their expected joint dynamic influence throughout remaining life. As a comparison, Table 6 provides inequality statistics across alternate measures of well-being for the age sixty population in the EHRS cohort. The final column provides Spearman's rank correlation coefficient between our welfare measure and each alternative measure.

Cross-sectional income inequality is lower than welfare inequality, though income does well predicting welfare at the bottom end of the distribution—the 10/50 ratio is

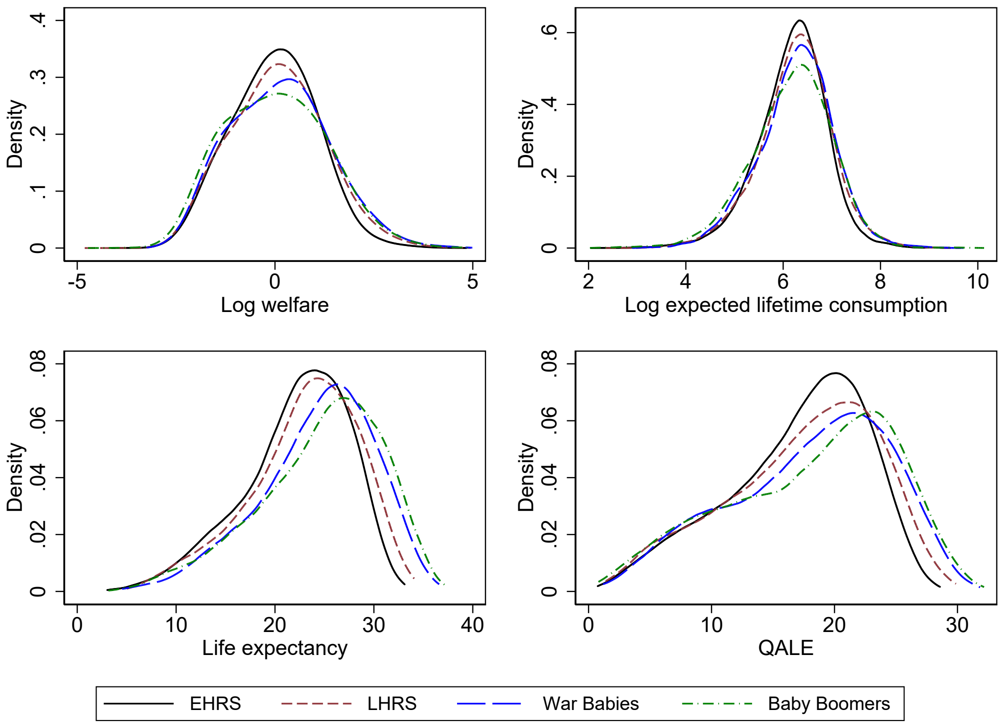

Figure 5: Distribution of welfare, consumption, and life expectancy over cohorts

Table 6: Comparing measures of inequality at age sixty

Measure	Gini	10/50 ratio	90/50 ratio	ρ
Welfare (λ)	0.544	0.234	3.774	
Income	0.492	0.225	2.788	0.508
Consumption	0.424	0.359	2.652	0.573
Health utility	0.109	0.628	1.126	0.745
Flow utility	0.235	0.433	1.533	0.767
Life expectancy	0.132	0.630	1.231	0.818
QALE	0.176	0.483	1.271	0.872
Expected lifetime consumption	0.364	0.383	2.125	0.921

Notes: Estimates for EHRS cohort using base year respondent analysis weights. Income, consumption, and health utility are cross-sectional measures at age sixty. Flow utility is calculated using cross-sectional consumption, leisure, and health along with our benchmark preferences. Spearman's rank correlation between λ and each measure denoted by ρ.

similar in the two measures. However, the rank correlation of 0.50 between welfare and income is quite low. So while income may provide a reasonable measure of aggregate inequality, relative income and welfare can be quite different at the individual level. Cross-sectional consumption at age sixty provides a somewhat better ranking of individual welfare, but under-estimates welfare inequality substantially more than income. The improved rank correlation is perhaps unsurprising as consumption directly enters preferences used in our welfare model. Age sixty health alone severely under-estimates aggregate welfare inequality, although it provides a better individual ranking than income or consumption. This speaks to the substantial influence of health and mortality in determining the distribution of our welfare measure.

The next row of Table 6 provides an estimate of welfare incorporating age sixty consumption, leisure, and health into our benchmark flow utility specification (2) but ignoring the subsequent life-cycle dynamics. Incorporating all three components in a static framework provides a better ranking of individual welfare, but ignoring the dynamics substantially under-estimates welfare inequality. For example, the Gini of age sixty flow utility is only 43% of that based on our dynamic welfare measure. The next two rows show that isolated dynamic health measures—life expectancy and QALE—continue to improve the rank correlation but under-predict welfare inequality. Finally, when combining longevity and consumption dynamics through estimating expected lifetime consumption, the rank correlation reaches 0.92, but inequality continues to be significantly lower than our more complete welfare measure.

Figure 6 further illustrates the nuanced relationship between health and economic outcomes by plotting quality-adjusted life expectancy against the ratio of consumption to welfare at age sixty. There is a clear negative correlation between the two measures with consumption over-predicting welfare for those of poor health and under-predicting for those of good health. This pattern is consistent with the relatively modest rank correlation between welfare and consumption. A key takeaway here is even in cases where economic outcomes provide a reasonable approximation to aggregate welfare inequality, they may still fail to be an adequate welfare measure at the individual level.

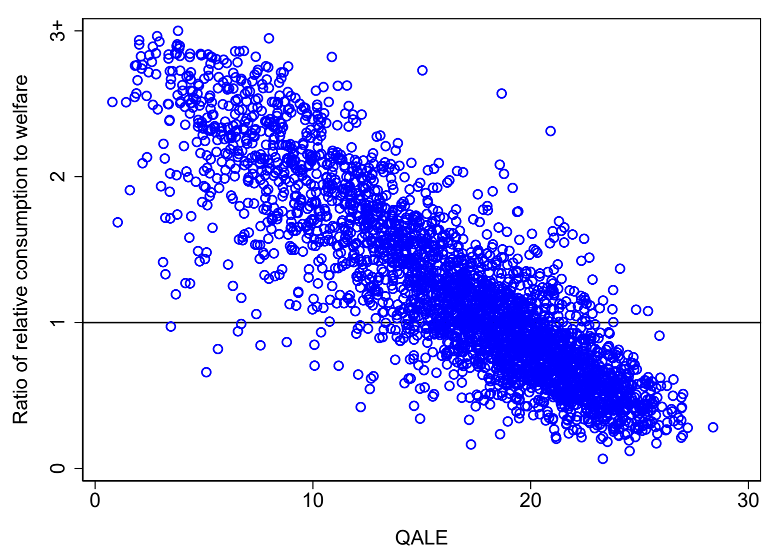

Figure 6: Variation in relationship between age sixty consumption, welfare, and QALE

Notes: Ratio of relative consumption to welfare given by $(c_i/c_m)/\lambda$ where c is age sixty consumption. Plot includes EHRS cohort only.

Finally, Table 7 provides Gini coefficients by cohort for age sixty distributions of our welfare and other measures. Welfare inequality has grown significantly more than age sixty cross-sectional income, consumption, or health inequality. This implies disparity measures based on economic outcomes such as income or consumption have become worse estimates of welfare inequality over time (as inequality has increased) at least partially due to growing gaps in health and mortality.

Table 7: Gini coefficients of welfare measures over cohorts

Cohort	Welfare (λ)	Income	Consumption	Health utility	QALE	ELC
EHRS	0.544	0.492	0.424	0.109	0.184	0.364
LHRS	0.606	0.532	0.442	0.123	0.198	0.390
War Babies	0.643	0.489	0.443	0.129	0.203	0.403
Baby Boomers	0.674	0.503	0.449	0.136	0.215	0.427

Notes: Estimates use base year respondent analysis weights. Income, consumption, and health utility are cross-sectional measures at age sixty. QALE is quality-adjusted life expectancy at age sixty. ELC is expected lifetime consumption at age sixty.

4.4 Robustness

Table 8 provides sensitivity results estimated under alternate modeling assumptions from our benchmark specification. While the magnitude of inequality measures are somewhat sensitive to underlying assumptions, the finding that welfare inequality is substantial and has grown over time is quite robust across specifications. Moreover, the rank correlations between welfare and alternate well-being measures remain relatively stable. We discuss each of these robustness checks in more detail below.

Table 8: Robustness results

			Gini by cohort				
	λ 10/50	λ 90/50	EHRS	LHRS	WB	BB	ρ
Benchmark	0.234	3.774	0.544	0.606	0.643	0.674	0.573
Compensating variation	0.059	2.856	0.505	0.533	0.546	0.566	0.556
Reference 90th %tile	0.314	2.842	0.446	0.500	0.533	0.555	0.573
$100k per QALY	0.076	6.465	0.670	0.731	0.763	0.784	0.502
$\beta = 0.90$	0.256	3.130	0.491	0.539	0.567	0.590	0.616
$\epsilon = 0.5$	0.231	3.726	0.539	0.600	0.637	0.665	0.572
$\epsilon = 2$	0.239	4.074	0.560	0.620	0.658	0.692	0.570
$\theta = 15.9$	0.258	3.539	0.525	0.584	0.621	0.652	0.571
Survival adjusted	0.177	4.015	0.568	0.618	0.648	0.674	0.573
Non-imputed data	0.242	3.543	0.522	0.568	0.591	0.627	0.603

Notes: Estimates use base year respondent analysis weights. War Babies denoted by WB and Baby Boomers by BB. Spearman's rank correlation between welfare and cross-sectional consumption at age sixty denoted by ρ.

4.4.1 Compensating variation

Our benchmark welfare measure is calculated in terms of consumption *equivalent* variation (EV). Alternatively, we could use the inverse of a *compensating* variation (CV) measure—what share of individual i's consumption would the median individual need to be ex-ante compensated to make them indifferent to receiving the current and potential future outcome bundles of individual i. Consumption-compensating variation satisfies:

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

where the only difference with the EV measure is the denominator is now the QALE for individual i as opposed to the median individual. The new measure effectively weights the benefit/loss of a consumption change by the flow utility of individual i as opposed to the median. This implies gaps in QALE are more detrimental to welfare for those below the median and less beneficial for those above.

The second row of Table 8 shows the sensitivity of inequality measures to the choice of consumption variation metric. Welfare at the tenth percentile is 23.4% of the median based on the baseline EV measure, compared to only 5.9% using CV. CV welfare estimates are also substantially lower at the top end of the distribution. On net, the decline in welfare at the very top of the distribution was strong enough to result in a lower Gini coefficient compared to the EV measure. However, the CV based measure still finds welfare inequality increased over cohorts, was higher than cross-sectional income or consumption inequality, and maintained only a modest rank correlation with consumption.

4.4.2 Reference individual

The third row of Table 8 shows sensitivity of results when using the individual at the ninetieth percentile of welfare in the EHRS cohort as the reference person instead of the median. The magnitude of inequality measurements are sensitive to the reference individual chosen—the welfare Gini falls to a lower but still substantial 0.44. This puts aggregate welfare inequality closer in line with cross-sectional income and consumption at age sixty. Nonetheless, individual rank correlation does not change (by definition) and a substantial increase in inequality across cohorts remains.

4.4.3 Preference parameters

Next we examine the impact of assuming a higher monetary value per QALY to calibrate the flow intercept $\bar{u}$. 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.[21] As a robustness check, we double our target to \$100,000 per QALY, which aligns more closely with VSL studies. The change results in a higher $\bar{u}$ placing additional weight on health and longevity differences in the welfare calculations. Inequality is substantially higher across all measurements but continues to increase across cohorts. The rank correlation with consumption also falls to 0.50.

The next four rows in Table 8 indicate sensitivity of other preference parameters. With a lower time discount rate β, our measure indicates somewhat less welfare inequality as differences in mortality and future consumption and health declines are less important. However, the pattern of main results hold. Main conclusions are also insensitive to alternate values of the Frisch elasticity of labor supply ϵ or altering the disutility weight on labor supply θ such that the first order condition holds for the sixty year old at the 75th percentile of the distribution (as opposed to our benchmark choice of the median).

4.4.4 Survival to sixty

An additional factor driving results is differential survival rates to age sixty across cohorts. For example, if medical advances are keeping sick individuals alive to age sixty in later cohorts that would be deceased in earlier cohorts, this would partially explain the growing welfare inequality. Using survival tables from the Social Security Administration (Bell and Miller, 2005), we estimate the EHRS cohort would be about 6.7% larger at age sixty if the cohort had realized the same survival rates as the Baby Boomers. In order to examine the quantitative influence of this channel, we increase the cohort size of the EHRS by 6.7% and assign these individuals a remaining lifetime utility of zero. We preform an analogous adjustment for the LHRS and War Babies, using cohort size increases of 4.0% and 1.9%. The "survival adjusted" results are provided near the bottom of Table 8. By construction, the inequality estimates increase for the first three cohorts. However, the increase in inequality over cohorts remains.

4.4.5 Imputed data

The last row in Table 8 provides results when using only raw data for estimating our forecasting model (i.e. we exclude imputed consumption and other data).[22] The cost of this estimation strategy is a loss in precision and potential bias due to any systematic pattern in missing observations. Moreover, we continue to use the imputed data for initial age sixty conditions for simulations to ensure there is enough data to conduct

[21] The VSL studies reviewed by Ryen and Svensson (2015) 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.

[22] Here we use only a single lag in the consumption equation (instead of the benchmark two) to preserve observations.

our cohort comparisons and so welfare estimates remain representative of the larger population. With only non-imputed data used in the forecasting model, estimated welfare inequality decreases somewhat in each cohort, but inequality continues to rise across cohorts. This demonstrates that overall patterns are not an artificial result of the underlying dynamics of the imputation model.

4.4.6 Form of flow utility

Finally, we examine the robustness of results to a more general form of flow utility given by:

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

which reduces to our benchmark case with $\gamma = 1$. These preferences follow those proposed by Trabandt and Uhlig (2011) and Jones and Klenow (2016) which maintain a constant Frisch elasticity of labor supply. With $\gamma > 1$ there is more curvature over consumption and the welfare cost of consumption inequality increases. However, leisure and consumption become less substitutable implying welfare inequality may be reduced if the inputs are strongly negatively correlated across individuals.

Table 9: Robustness results

	EV 10/50 ratio by cohort				CV 90/50 ratio by cohort				
	EHRS	LHRS	WB	BB	EHRS	LHRS	WB	BB	ρ
$\gamma = 1$	0.234	0.210	0.196	0.196	2.856	3.161	3.211	3.563	0.573
$\gamma = 1.5$	0.207	0.180	0.163	0.165	3.567	3.915	3.829	4.158	0.520
$\gamma = 2$	0.231	0.197	0.163	0.167	4.237	4.500	4.183	4.502	0.471

Notes: Estimates use base year respondent analysis weights. War Babies denoted by WB and Baby Boomers by BB. Spearman's rank correlation between EV measure of welfare and cross-sectional consumption at age sixty denoted by ρ.

We examine sensitivity of results to increases in curvature to $\gamma = 1.5$ and $\gamma = 2$. However, with higher curvature over consumption than the benchmark, it is no longer possible to calculate EV welfare for those at the very top of the distribution as no amount of consumption increase would provide the same expected life-time utility to the median individual. Likewise, using the CV measure is not possible for the worst off as no amount of consumption would be enough to compensate the median individual. Under these feasibility considerations, Table 9 provides the 10/50 welfare ratio based on the EV measure and the 90/50 ratio based on CV for alternate curvatures.

The 90/50 ratio monotonically increases with the curvature parameter suggesting welfare inequality may be under-predicted in the benchmark case. The change in 10/50 ratio is non-monotonic but both robustness experiments suggest increased inequality relative to the benchmark as well. Looking more closely across cohorts, the 10/50 ratio continues to decrease or be stable over time regardless of the curvature value. The

90/50 ratio increases over cohorts at higher curvatures, with the exception of moving from the LHRS cohort to War Babies. To understand this change, note that consumption inequality was relatively stable between the LHRS and War Babies, while health inequality increased (see Table 7). With low curvature in the benchmark case, consumption inequality plays a lesser role in determining welfare inequality. So the increase in health inequality dominates the benchmark and the 90/50 welfare ratio increases slightly between the cohorts. However, the welfare cost of health inequality decreases with γ, resulting in the observed decline in 90/50 ratio at higher curvatures. While these patterns are inconclusive, they suggest the increase in inequality across cohorts could be somewhat muted with higher curvature, at least between the LHRS cohort and War Babies. Lastly, there is some decline in the rank correlation with consumption as leisure and consumption are less substitutable in welfare and are negatively correlated across individuals.

5 Conclusion

We propose and estimate an individual measure of welfare incorporating heterogeneity and uncertainty in future consumption, leisure, health, and mortality at age sixty. Our measure broadly indicates that inequality is larger and has increased more rapidly than suggested by other welfare metrics such as income or consumption. We also find health and mortality gaps are more important than consumption in explaining welfare inequality among the elderly in our sample, with leisure playing a comparatively minor role. Moreover, health at age sixty is a better indicator of individual well-being rank than income or consumption.

While there are limitations to our approach, the framework provides important insight and ample opportunities for further analyses. As our framework allows estimation of welfare for each individual, it is possible to compare welfare distributions across various sub-groups of the population (see appendix C for examples of welfare breakdowns by education, region, gender, and race). While we focus on welfare at age sixty, changes in welfare can also be calculated and analyzed over the elderly life-course, for example comparing our measure with welfare at age seventy or eighty. Our measure could also be used as an outcome in designed or natural experiments, for example to examine the effect of healthcare policy on the distribution of welfare. Moreover, our framework could be extended in multiple directions to examine additional cohorts, younger ages, or welfare inequality differences across countries.

References

Attanasio, O., Hurst, E., and Pistaferri, L. (2014). The evolution of income, consumption, and leisure inequality in the United States, 1980–2010. In *Improving the Measurement of Consumer Expenditures*, pages 100–140. University of Chicago Press.

Attanasio, O. and Pistaferri, L. (2014). Consumption inequality over the last half century: Some evidence using the new PSID consumption measure. *The American Economic Review*, 104(5):122–126.

Attanasio, O. P., Battistin, E., and Padula, M. (2010). *Inequality in living standards since 1980: Income tells only a small part of the story*. Government Institutes.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2):300–323.

Barr, R., Petrie, C., Furlong, W., Rothney, M., and Feeny, D. (1997). Health-related quality of life during post-induction chemotherapy in children with acute lymphoblastic leukemia in remission. *International Journal of Oncology*, 11(2):333–339.

Barro, R. J. and Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics*, 126(1):51–102.

Becker, G. S., Philipson, T. J., and Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *The American Economic Review*, 95(1):277–291.

Bell, F. C. and Miller, M. L. (2005). *Life tables for the United States social security area, 1900-2100: Actuarial Study No. 120*. Number 11-11536. Social Security Administration, Office of the Chief Actuary.

Blanchard, C., Feeny, D., Mahon, J. L., Bourne, R., Rorabeck, C., Stitt, L., and Webster-Bogaert, S. (2003). Is the health utilities index responsive in total hip arthroplasty patients? *Journal of Clinical Epidemiology*, 56(11):1046–1054.

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

Bosworth, B., Burtless, G., and Zhang, K. (2016). Later retirement, inequality in old age, and the growing gap in longevity between rich and poor. *Economic Studies at Brookings*.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA*, 315(16):1750–1766.

Colby, S. L., Ortman, J. M., et al. (2015). Projections of the size and composition of the US population: 2014 to 2060. *US Census Bureau*, 9.

Crystal, S. and Shea, D. (1990). The economic well-being of the elderly. *Review of Income and Wealth*, 36(3):227–247.

Crystal, S. and Waehrer, K. (1996). Later-life economic inequality in longitudinal perspective. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 51(6):S307–S318.

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

Currie, J. and Schwandt, H. (2016). Inequality in mortality decreased among the young while increasing for older adults, 1990–2010. *Science*, 352(6286):708–712.

Cutler, D. M., Ghosh, K., and Landrum, M. B. (2013). Evidence for significant compression of morbidity in the elderly US population. *NBER Working Paper No. 19268*.

Cutler, D. M. and Katz, L. F. (1992). Rising inequality? Changes in the distribution of income and consumption in the 1980s. *NBER Working Paper No. 3964*.

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).

Deaton, A. S. and Paxson, C. H. (1998). Aging and inequality in income and health. *The American Economic Review*, 88(2):248–253.

Dworkin, R. (1985). *A matter of principle*. OUP Oxford.

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

Fanshel, S. and Bush, J. W. (1970). A health-status index and its application to health-services outcomes. *Operations research*, 18(6):1021–1066.

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.

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, 47(4):1029–1075.

Fleurbaey, M. and Gaulier, G. (2009). International comparisons of living standards by equivalent incomes. *The Scandinavian Journal of Economics*, 111(3):597–624.

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.

Furlong, W. J., Feeny, D. H., Torrance, G. W., and Barr, R. D. (2001). The health utilities index (hui®) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, 33(5):375–384.

Gottschalk, P., Moffitt, R., Katz, L. F., and Dickens, W. T. (1994). The growth of earnings instability in the US labor market. *Brookings Papers on Economic Activity*, 1994(2):217–272.

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.

Heathcote, J., Perri, F., and Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics*, 13(1):15–51.

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.

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

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.

Hurd, M. D. and Rohwedder, S. (2007). Economic well-being at older ages: Income- and consumption-based poverty measures in the HRS. *RAND Corporation Working Paper WR-410*.

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.

Katz, L. F. et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of Labor Economics*, 3:1463–1555.

Kenworthy, L. (2008). *Jobs with Equality*. Oxford University Press.

Kitagawa, E. M. and Hauser, P. M. (1973). *Differential mortality in the United States: A study in socioeconomic epidemiology.* Harvard University Press, Cambridge, MA.

Klarman, H. E. and Rosenthal, G. D. (1968). Cost effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care*, 6(1):48–54.

Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27(2):350–360.

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.*

Minkler, M., Fuller-Thomson, E., and Guralnik, J. M. (2006). Gradient of disability across the socioeconomic spectrum in the United States. *New England Journal of Medicine*, 355(7):695–703.

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

National Academies of Sciences, Engineering, and Medicine (2015). *The growing gap in life expectancy by income: Implications for federal programs and policy responses.* National Academies Press, Washington, DC.

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

Pijoan-Mas, J. and Ríos-Rull, J.-V. (2014). Heterogeneity in expected longevities. *Demography*, 51(6):2075–2102.

Piketty, T. and Saez, E. (2014). Inequality in the long run. *Science*, 344(6186):838–843.

Rawls, J. (1971). *A theory of justice.* Harvard university press.

Roemer, J. E. (1998). *Equality of opportunity.* Harvard University Press.

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

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.

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

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.

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.

Torrance, G. W., Thomas, W. H., and Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research*, 7(2):118.

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

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

Zeckhauser, R. and Shepard, D. (1976). Where now for saving lives? *Law and Contemporary Problems*, 40(4):5–45.

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 multiply impute their 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.[23] We set $m = 12$ but test the sensitivity of results to higher values of m.

There are two primary assumptions underlying the proposed imputation method. First, the complete data is assumed 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

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

(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 2014). 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 data file (v.P) included in our imputation model are age (AGEY_E), aged squared, number of household members (HHRES), total wealth (ATOTA), wealth squared, log household income (ITOT), log income squared, and dummy indicators for 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), 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.[24]

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 same, the comparison gives a sense of the plausibility of imputations (Honaker et al., 2011). Figure 7 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.

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 8 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

[24]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).

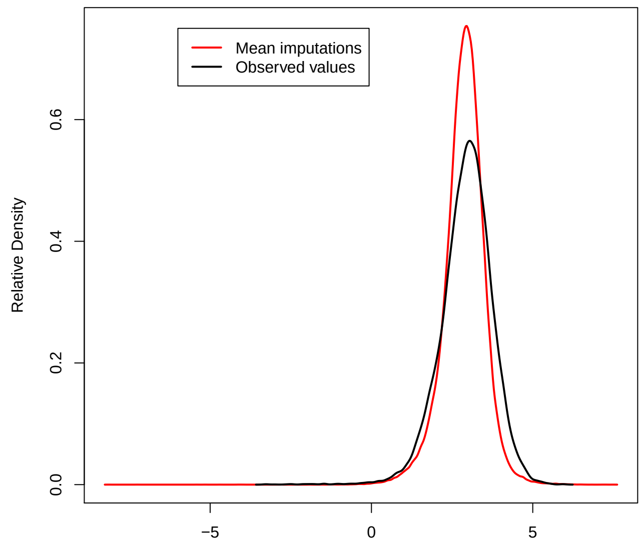

Figure 7: Distributions of observed and imputed values of consumption

confidence intervals should cross the $y = x$ line). In our case, 94% of the observed values are within the confidence bounds.

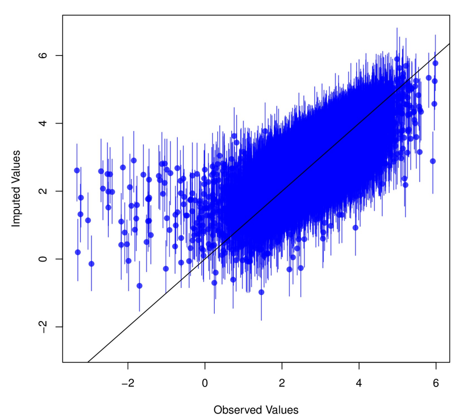

Figure 8: Overimputed values of consumption

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 9 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 twelve (2014) all values are imputed as these waves are outside of our CAMS data window.

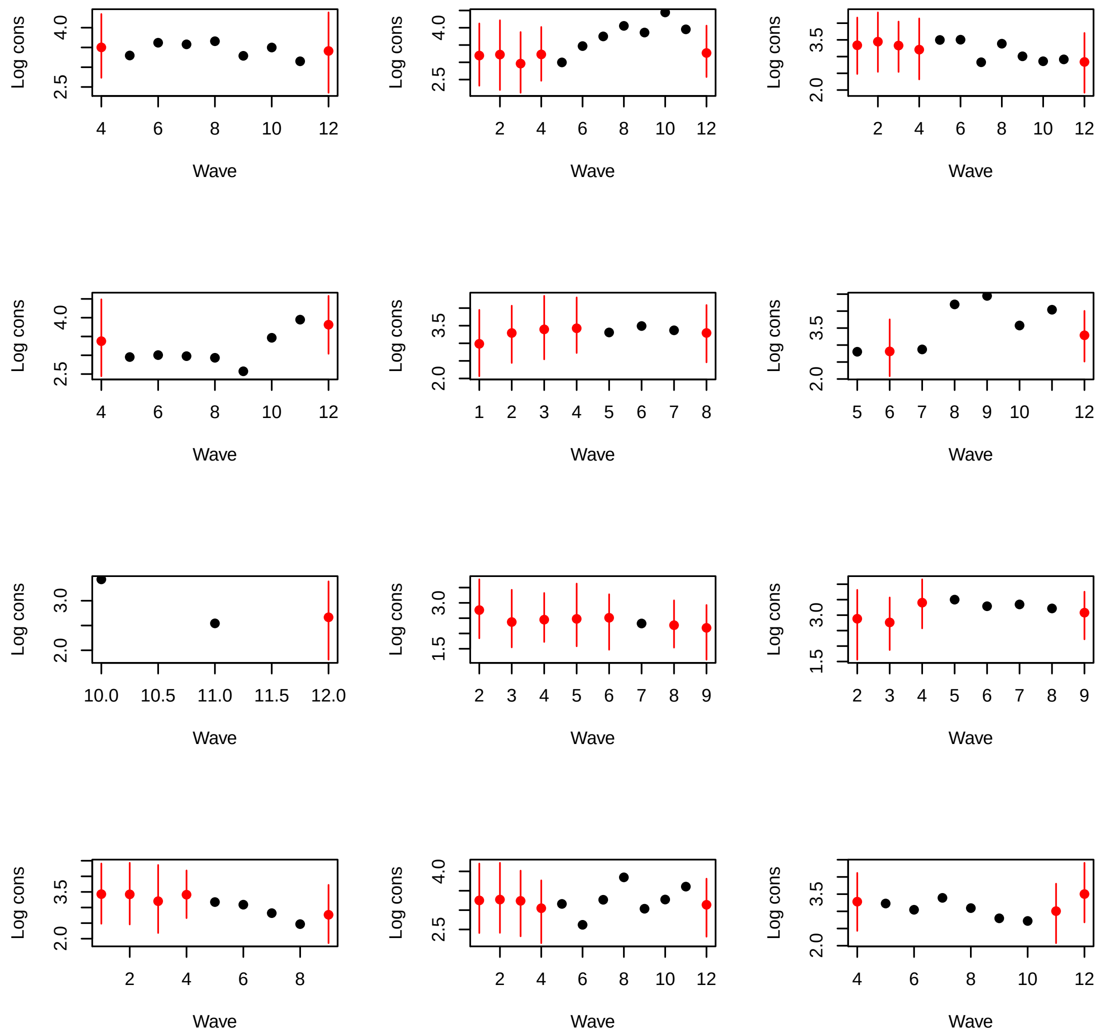

Figure 9: Observed and imputed consumption over time for a random sub-sample

B Forecasting model

In this appendix we lay out our forecasting model and underlying assumptions. We also detail our estimation and simulation procedures. For additional applications of this framework see Miller et al. (2019), which uses a simplified version of the model to look more closely at the relationship between health and elderly life-cycle consumption.

B.1 Panel VAR representation

While we allow for multiple lags in estimation of the forecasting model, the following VAR(1) demonstrates the key features of the framework. Let Y_{it} be a vector of outcomes for individual i at time t that includes log consumption c, retirement indicator r, self-rated health s, and our $n = 9$ morbidity states given by $n \times 1$ vector M. Conditional on survival, the outcomes evolve according to the structural VAR(1) model:

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

where ϵ is a vector of normally distributed shocks with mean zero. 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 in our benchmark model that all parameters are homogeneous across individuals and time (e.g. $A_{it} = A \quad \forall i, t$).

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

$$
\begin{array}{cc}
 & \begin{array}{cc} \overbrace{\hphantom{-A_{11}}}^{n} & \overbrace{\hphantom{-A_{12}\;\; -A_{13}\;\; -A_{14}}}^{3} \end{array} \\
\begin{array}{r} n\left\{\vphantom{-A_{11}}\right. \\ 3\left\{\vphantom{\begin{array}{c} a \\ a \\ a \end{array}}\right. \end{array} &
\left[\begin{array}{c:ccc}
-A_{11} & -A_{12} & -A_{13} & -A_{14} \\ \hdashline
-A_{21} & 1 & -a_{23} & -a_{24} \\
-A_{31} & -a_{32} & 1 & -a_{34} \\
-A_{41} & -a_{42} & -a_{43} & 1
\end{array}\right]
\end{array}
\left[\begin{array}{c}
M_{it} \\ \hdashline
s_{it} \\ r_{it} \\ c_{it}
\end{array}\right]
=
\begin{array}{c}
\begin{array}{cc} \overbrace{\hphantom{B_{11}}}^{n} & \overbrace{\hphantom{B_{12}\;\; B_{13}\;\; B_{14}}}^{3} \end{array} \\
\left[\begin{array}{c:ccc}
B_{11} & B_{12} & B_{13} & B_{14} \\ \hdashline
B_{21} & b_{22} & b_{23} & b_{24} \\
B_{31} & b_{32} & b_{33} & b_{34} \\
B_{41} & b_{42} & b_{43} & b_{44}
\end{array}\right]
\end{array}
\left[\begin{array}{c}
M_{it-1} \\ \hdashline
s_{it-1} \\ r_{it-1} \\ c_{it-1}
\end{array}\right]
+
\left[\begin{array}{c}
\epsilon_{1,it} \\ \hdashline
\epsilon_{2,it} \\ \epsilon_{3,it} \\ \epsilon_{4,it}
\end{array}\right],
$$

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

As illustrated in Figure 1, the causal pathways we propose suggest a block recursive system. Specifically, we assume that $A_{12} = A_{13} = A_{14} = 0$ in the morbidity block, $a_{23} = a_{24} = 0$ in the self-rated health block, and $a_{34} = 0$ in the retirement block. In other words, we assume the contemporaneous causal pathway runs from morbidities to self-rated health to retirement to consumption. However, we allow health and retirement to affect future outcomes through lagged effects (though we assume there is no such feedback from consumption and set $B_{14} = b_{24} = b_{34} = 0$). Block triangulation of the system eliminates simultaneity across blocks and allows for block-by-block estimation.[25]

B.1.1 Exogenous characteristics

We also include a $k \times 1$ vector of exogenous individual characteristics X_{it} as predictors in our model. The VAR(1) model with exogenous regressors takes the following form:

$$AY_{it} = BY_{it-1} + CX_{it} + \epsilon_{it}. \tag{7}$$

Exogenous characteristics include dummies for age, education, gender, race, census division, census occupation code, and birth cohort. We also include a linear calendar year trend and post-2008 indicator to help account for the influence of the great recession on outcomes.[26] Lastly, we include a time invariant individual unobserved endowment π in the consumption equation. The endowment π is modeled as a fixed effect with no restriction on the correlation with other model regressors. The unobserved individual effect helps maintain the appropriate variance in consumption across time by effectively acting as a person specific drift in the autoregressive process. We exclude time invariant exogenous regressors (education, gender, race, census division, occupation code, birth cohort) from the consumption equation due to colinearity with the fixed effect. However, we include socioeconomic characteristics instead of additional individual fixed effects in the health and retirement equations because 1) morbidities and retirement are absorbing states and 2) self-rated health is ordinal, each of which poses difficulties in estimating dynamic panel models with fixed effects.[27] The resulting exogenous effects then take the following form:

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

[26] Note 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).

[27] For example, as many individuals never enter a given absorbing state in the data, it is not possible to estimate unique unobserved fixed effects for each individual for each absorbing state. As these estimates would be required for our simulations, we do not include additional unobserved fixed effects in absorbing state models.

$$CX_{it} = \begin{array}{c} n\left\{ \right. \\ \\ \\ \end{array} \underbrace{\left[\begin{array}{cccccccccc} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} & C_{19} & 0 \\ \hdashline c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} & c_{29} & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} & c_{37} & c_{38} & c_{39} & 0 \\ c_{41} & 0 & 0 & 0 & 0 & 0 & 0 & c_{48} & c_{49} & c_{410} \end{array} \right]}_{(n+3)\times k} \underbrace{\left[\begin{array}{c} Age_{it} \\ Education_i \\ Gender_i \\ Race_i \\ Division_i \\ Occupation_i \\ Cohort_i \\ Year_t \\ Post_t \\ \pi_i \end{array} \right]}_{k \times 1}.$$

Lastly we normalize $c_{410} = 1$ to allow identification of the unobserved fixed effects in the consumption block.

B.1.2 Consumption

The resulting consumption forecasting equation given in system (7) can be explicitly written as:

$$\begin{aligned} c_{it} = A_{41}M_{it} + a_{42}s_{it} + a_{43}r_{it} + B_{41}M_{it-1} + b_{42}s_{it-1} + b_{43}r_{it-1} \\ + b_{44}c_{it-1} + c_{41}Age_{it} + c_{48}Year_t + c_{49}Post_t + \pi_i + \epsilon_{4,it}. \end{aligned} \quad (8)$$

This is a standard linear dynamic panel data model with lagged dependent variable and individual level fixed effects (π). Given our block recursive system, this equation may be estimated independently of other blocks with all structural parameters identified including the variance of ϵ_4.

B.1.3 Retirement

As retirement is a binary outcome, forecasting of the measure is not a true linear VAR process. In contrast, we assume a continuous latent variable $r^\star$ underlies the observed outcome such that:

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

Conditional on working the previous period (and setting $b_{33} = 0$), the retirement model as defined in system (7) is then given by:

$$r^\star_{it} \quad = \quad A_{31}M_{it} + a_{32}s_{it} + B_{31}M_{it-1} + b_{32}s_{it-1} + [c_{31}, \ldots, c_{39}]\, X_{it} + \epsilon_{3,it}. \quad (9)$$

Similar to consumption, current and lagged values of health (both self-rated and specific morbidities) are allowed to influence the probability of retirement. We assume ϵ_3 is an *iid* shock with standard normal distribution implying the retirement model has a standard probit structure.

B.1.4 Self-rated health

As self-rated health is measured on a five point scale, we again assume a continuous latent variable $s^\star$ underlies the observed outcome. The self-rated health model as defined in system (7) is then given by:

$$s_{it}^\star \;=\; A_{21}M_{it} + B_{21}M_{it-1} + b_{22}s_{it-1} + b_{23}r_{it-1} + [c_{21}, \ldots, c_{29}]\, X_{it} + \epsilon_{2,it}, \qquad (10)$$

with the observed health state defined as:

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

for cut-points $(\kappa_0, \ldots, \kappa_5)$ with $\delta = 1$ representing the worst health state (poor) and $\delta = 5$ the best health state (excellent). Note that latent self-rated health is assumed to depend on the lagged value of the *observed* self-rated health category to incorporate the persistence in general health shocks over the life-course. We assume ϵ_2 is an *iid* shock with standard normal distribution. Thus the evolution of self-rated health follows an ordered probit structure.

B.1.5 Morbidities

Unlike other outcomes, block triangulation of the system does not allow direct identification of the structural parameters in the morbidity block as there are $n = 9$ separate outcomes. Instead the morbidity block is estimated as a reduced form VAR. The reduced form system is obtained by premultiplying the structural system block by the inverse of matrix $-A_{11}$:

$$M_{it} = -A_{11}^{-1}B_{11}M_{it-1} - A_{11}^{-1}B_{12}s_{it-1} - A_{11}^{-1}B_{13}r_{it-1} - A_{11}^{-1}\,[C_{11}, \ldots, C_{19}]\, X_{it} - A_{11}^{-1}\epsilon_{1,it}.$$

Denoting $-A_{11}^{-1}B_{1j} = \hat{B}_j$, $-A_{11}^{-1}\,[C_{11}, \ldots, C_{17}] = \hat{C}$ and $-A_{11}^{-1}\epsilon_{1,t} = e_t$ yields the following reduced form system:

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

In the reduced form VAR 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. $cov\,(e_{it}, e'_{it}) \neq 0$). 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.

Contemporaneous morbidity shocks are assumed to follow a standard multivariate normal distribution with an $n \times n$ covariance matrix given by Σ. 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.

As morbidity outcomes are binary, we again assume a continuous latent variable $m^\star$ underlies each observed morbidity state 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 $j = 1 \ldots n$. We then have the following model:

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

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

B.1.6 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:

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

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

$$\left[\begin{array}{c:ccc} D_{11} & D_{12} & D_{13} & D_{14} \\ \hdashline D_{21} & d_{22} & d_{23} & d_{24} \\ D_{31} & d_{32} & d_{33} & d_{34} \\ D_{41} & d_{42} & d_{43} & d_{44} \end{array}\right] \left[\begin{array}{c} M_{it-2} \\ \hdashline s_{it-2} \\ r_{it-2} \\ c_{it-2} \end{array}\right].$$

Here, for example, coefficient vector D_{41} allows the second lag of the morbidity state vector to directly affect current consumption. 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 $d_{42} = 0$.

B.1.7 Mortality

The final process to be modeled is survival from one period of life to the next. As all other outcomes are conditional on survival, mortality probabilities are estimated independently of the VAR system above. Conditional on being alive at time $t-1$, survival to the following period of life is given by:

$$\psi_{it} = I\left(\sum_{k=1}^{K}\left[\gamma_k^M M_{it-k} + \gamma_k^s s_{it-k} + \gamma_k^r r_{it-k}\right] + \delta X_{it} + u_{it} > 0\right), \tag{12}$$

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 specification allows K lags of morbidity states, self-rated health, and retirement to influence survival probability.

B.2 Estimation

The pooled sample used to estimate the forecasting model includes all individuals born prior to 1960 and aged fifty and over at the time of the survey. This consists of 35,889 unique individuals and 216,626 total individual-year observations. Table 10 shows descriptive statistics for modeled outcomes for each cohort in the HRS. Incidence of each morbidity state was substantial among respondents, allowing for relatively precise estimates of their effects on dynamic processes. However, there was still significant variation in incidence rates across morbidity states. For example, in the EHRS cohort, over 57% of observations reported arthritis while only 7% reported having suffered a stroke. In terms of labor supply, the share of retired individuals ranged from 37% in the most recent Baby Boomer cohort to 95% in the (much older) AHEAD cohort. Annual real consumption averaged between $20-$28,000 across cohorts.

The dependency structure of our forecasting model is further motivated by the observed correlations in consumption, health, and labor supply found in the HRS data (see Figure 10). There are positive associations of varying strength across morbidities highlighting the importance of modeling their evolution jointly. Morbidity states have a strong negative correlation with self-rated health and a more modest positive association with retirement. Consumption is positively associated with self-rated health and negatively associated with retirement and all morbidities except cancer. Cancer is the clear outlier with the weakest co-morbidity correlations and a small positive correlation with consumption.

The positive relationship between self-rated health and consumption—and negative relationship with morbidities—suggests using consumption as the sole basis of a well-being metric could understate the inequality among the elderly. On the other hand, the negative association between consumption and retirement suggests a possible overstatement of welfare inequality as those with low consumption may enjoy more leisure. Our welfare model allows us to gauge the relative strength of these channels.

Table 10: Estimation sample descriptive statistics by cohort

	AHEAD	CODA	EHRS	LHRS	WB	BB	MBB
Individuals	7,758	4,231	5,365	5,133	3,615	4,759	5,028
Observations	36,896	27,522	44,237	44,213	26,805	22,251	14,702
Age (mean)	81.65	74.78	67.03	62.01	59.63	57.49	54.38
Hypertension (%)	54.69	56.98	52.89	49.77	48.25	48.46	46.21
Diabetes (%)	15.40	18.88	18.99	17.50	17.72	19.10	18.47
Cancer (%)	16.79	17.50	13.49	10.61	10.03	7.95	7.12
Lung disease (%)	9.46	10.16	9.54	8.33	7.04	6.73	7.29
Heart disease (%)	35.26	30.45	22.67	18.61	16.12	14.04	12.03
Stroke (%)	15.22	11.94	7.15	5.73	5.41	4.77	3.97
Psyche problem (%)	11.81	11.63	11.03	12.79	16.88	19.34	18.94
Arthritis (%)	55.99	59.96	57.15	51.63	50.54	44.53	37.92
Difficulty with ADLs (%)	40.26	27.96	23.12	20.80	21.41	20.04	17.89
Self-rated health (%)							
Poor	14.26	10.30	9.20	7.80	6.67	7.88	7.41
Fair	25.74	21.48	19.12	18.49	16.55	19.30	20.74
Good	30.86	32.29	31.50	30.62	30.39	29.68	30.86
Very good	21.33	26.54	28.30	29.13	32.06	30.40	29.49
Excellent	7.80	9.39	11.89	13.95	14.33	12.73	11.50
Retired (%)	95.32	90.45	74.70	60.57	54.12	43.50	37.00
Annual consumption ($1000s, mean)	22.45	25.13	25.13	27.25	27.58	24.23	20.60
Male (%)	37.67	47.16	45.18	45.41	37.38	42.40	42.18
Education (%)							
<HS	41.78	32.68	31.38	28.37	21.48	19.98	21.77
HS	29.60	31.38	32.72	32.92	30.96	24.76	25.16
Some College	16.30	17.75	18.50	20.42	24.42	28.24	29.21
College	12.32	18.19	17.40	18.29	23.14	27.02	23.85
Race (%)							
White	84.93	86.87	80.22	79.96	80.23	68.31	60.43
Black	12.97	9.73	16.37	15.97	14.87	20.80	25.96
Other	2.11	3.40	3.41	4.07	4.90	10.89	13.61

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.

B.2.1 Methods

As there is no simultaneity across blocks in the system, we estimate the model block-by-block. The consumption block is comprised only of equation (8), 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.[28] Including a single

[28]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. However, results are insensitive to the use of the burn-in initialization which assumes that initial conditions are in the infinite past and are drawn using the same bootstrap procedure used for bias

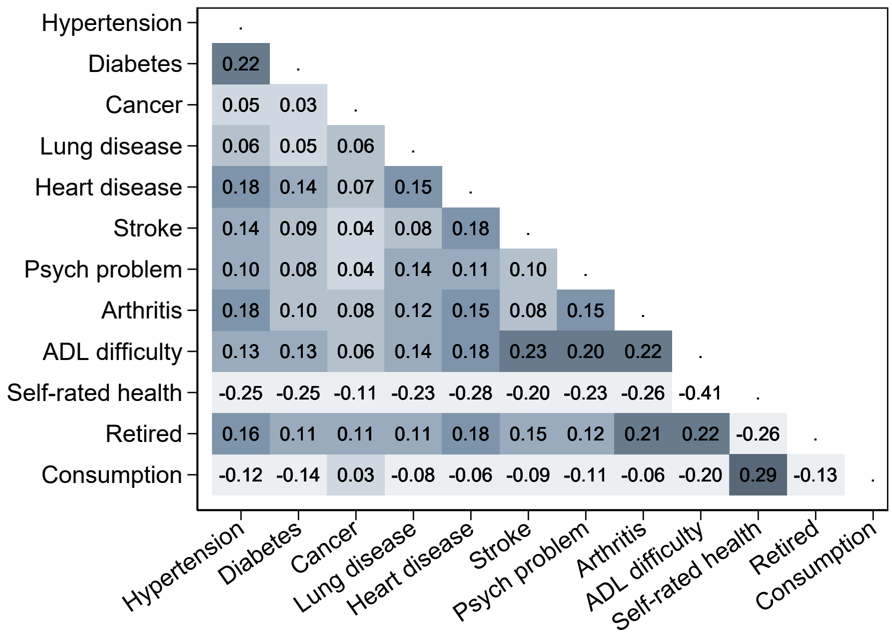

Figure 10: Outcome correlations

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.[29]

For consistency with the model of consumption, we use two lags of outcomes in all retirement, health, and survival equations (i.e. we estimate a VAR(2) system and set $K = 2$ in the survival model). The ordered probit model of self-rated health (10) is estimated independently of other VAR blocks using maximum likelihood.[30] The retirement equation (9) and mortality equation (12) are estimated independently using standard probit regressions.

This leaves the morbidity block. The morbidity model (11) is structured as a multivariate probit with correlated shocks. We 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

correction (see De Vos et al. (2015) for details).

[29]Second lags of retirement and health outcomes were insignificant and noisy so we opted for parsimony by excluding them. This is equivalent to estimating the VAR(2) system with $D_{41} = d_{42} = d_{43} = 0$. 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 0.63 regardless of imputed dataset used for the test.

[30]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 retirement and morbidity) model. The standard (ordered) probit estimator is consistent and provides asymptotically valid test statistics and standard errors.

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 (11)). 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 we estimate the model as is, without conditioning on time t morbidity status. While inclusion of all observations in the bivariate probit does not effect the likelihood or estimates of the remaining (non-infinite) coefficients, it is possible there could be numerical convergence problems. However, inclusion of all observations does not result in numerical instability in our case and likelihoods converge without issue. Moreover, conditioning the bivariate probit on, for example, 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.

B.2.2 Estimation results

Select estimation results for the forecasting model are provided in Figure 11 while the full set of results are shown in Tables 13-15 in appendix C. The first panel of Figure 11 shows the average marginal effects of morbidities on the contemporaneous probability of reporting poor self-rated health. For example, a recent stroke increases the probability of reporting poor health by 5.9 percentage points. Similarly, the second panel shows that a stroke directly increases the probability of retirement by 14 percentage points. However, a stroke can additionally influence retirement indirectly by lowering self-rated health. Continuing the example, a stroke directly decreases consumption by about 0.08 log points and decreases the probability of survival to the next model period by 2.6 percentage points. Again there are additional indirect effects operating on consumption and mortality due to any changes in self-rated health and retirement induced by the stroke. The final panel in Figure 11 further illustrates the dynamics of the system by showing the average marginal effects of current health states and retirement on the probability of having a stroke the following model period. For example, heart disease increases the probability of a stroke the following period by 0.8 percentage points. Moreover, these relationships continue to evolve dynamically throughout the system over time (see impulse response Figure 12 in appendix C which demonstrates the dynamic relationships for heart disease).

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

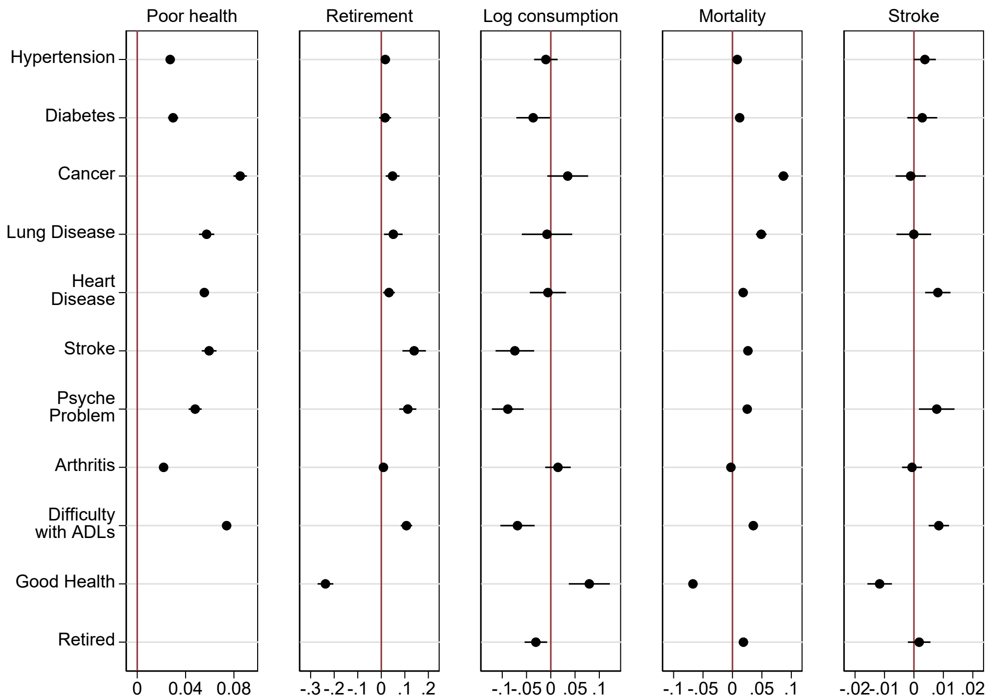

Figure 11: Select estimation results

Notes: Dependent variables across columns. Average marginal effects on the probability of an outcome reported for probit results—poor health, retirement, mortality, and stroke. Contemporaneous associations reported for poor health, retirement, and consumption as dependent variables. Lagged associations reported for mortality and stroke. Good health coefficients use poor health state as reference group. Spikes indicate 95% confidence intervals.

sixty data is not available for the older AHEAD or CODA cohorts, so these are excluded from our welfare analysis. Moreover, the mid-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.[31] 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 11 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 move in same directions for all cohorts.

Table 11: 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	2,735
Male (%)	47.20	46.36	46.82	46.61	47.89	47.92	48.25	47.53
Education (%)								
<HS	29.07	28.88	25.32	25.43	18.73	18.47	14.88	14.87
HS	33.61	33.80	32.08	32.33	30.45	30.27	24.80	24.94
Some College	19.28	19.25	21.52	21.37	24.35	24.45	29.19	28.92
College	18.04	18.08	21.09	20.87	26.46	26.81	31.13	31.27
Race (%)								
White	86.30	86.55	86.15	86.54	85.48	86.00	81.57	81.79
Black	10.38	10.24	10.00	9.97	9.73	9.18	10.81	10.58
Other	3.31	3.20	3.85	3.49	4.79	4.82	7.63	7.64

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 12 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. 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 savings motive) could also potentially explain some of the decline in flow consumption for later cohorts. Supportive of these hypotheses, our simulations suggest the consumption of younger cohorts to some extent caught up to older cohorts later in the life-cycle (see Figure 19 in appendix C).

Using age sixty data as initial ($t = 0$) conditions[32], we simulate the remaining life outcomes for each individual (i) as follows:

[31]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.

[32]Initial conditions also include unobserved endowments $\hat{\pi}$ estimated from model (8) using the prediction method of De Vos et al. (2015).

Table 12: Simulation sample initial conditions by cohort

	EHRS	LHRS	WB	BB
Age (mean)	60	60	60	60
Hypertension (%)	38.10	41.93	47.60	51.23
Diabetes (%)	11.81	12.77	16.45	20.13
Cancer (%)	6.84	8.25	10.82	9.48
Lung disease (%)	7.11	6.78	7.37	8.15
Heart disease (%)	13.85	14.75	16.11	16.25
Stroke (%)	2.90	3.88	5.22	4.66
Psyche problem (%)	7.44	11.85	17.32	21.85
Arthritis (%)	44.79	48.12	51.62	52.53
Difficulty with ADLs (%)	11.75	19.35	22.40	22.42
Self-rated health (%)				
Poor	7.31	6.68	6.61	7.26
Fair	15.20	16.71	16.60	17.15
Good	28.32	30.12	31.08	29.34
Very good	31.66	30.80	31.72	34.19
Excellent	17.51	15.70	13.98	12.06
Retired (%)	48.66	50.46	48.07	47.47
Annual consumption ($1000s, mean)	27.59	30.29	29.43	26.41

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.

1. Survival shock u_{i1} is drawn and survival to time $t = 1$ (age 62) is determined according to equation (12). 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 15). This shock vector along with the model outlined in equation (11) is used to compute simulated age 62 morbidity vector M_{i1}.

3. Given age 62 morbidities (M_{i1}), general health shock $\epsilon_{2,i1}$ is drawn and age 62 self-rated health (s_{i1}) is computed using equation (10).

4. Given age 62 self-rated health (s_{i1}) and morbidities (M_{i1}), retirement shock $\epsilon_{3,i1}$ is drawn to determine age 62 retirement (r_{i1}) using equation (9).

5. Given all other age 62 outcomes (r_{i1}, s_{i1}, M_{i1}), consumption shock $\epsilon_{4,i1}$ is drawn to determine age 62 consumption (c_{i1}) using equation (8).

6. Steps 1-5 are repeated for $t = 2, 3, \ldots$ until death or $t = 30$ (age 120).

7. Steps 1-6 are repeated 5,000 for each individual.

A comparison between the average simulated life-cycle profiles and those based on available data is shown by cohort in Figures 13-17 in appendix C. 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. 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 18 years later (when the EHRS cohort is age 78). As opposed to average profiles, the simulations match somewhat less well for the standard deviation of consumption and health (Figure 17). The spikes in standard deviation of consumption in the data is driven by sensitivity to high consumption outliers. There are also relatively few data observations for Baby Boomers at age 66, resulting in the sharp declines in standard deviations shown in the figure.

C Additional Figures and Tables

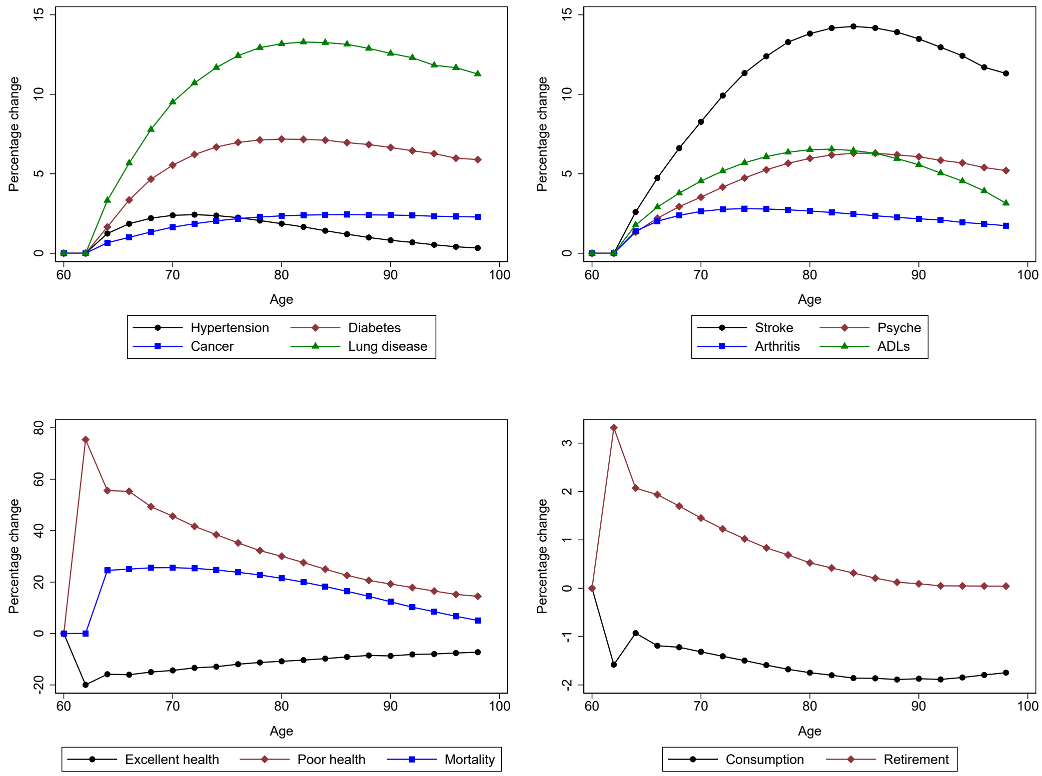

Figure 12: Impulse response to incidence of heart disease at age 62

Notes: Results plot percentage difference in expected outcomes with the exogenous onset of heart disease at age sixty-two relative to remaining without heart disease at sixty-two. Sample includes all individuals in the simulation sample without heart disease at age sixty. Expected outcomes are conditional on survival.

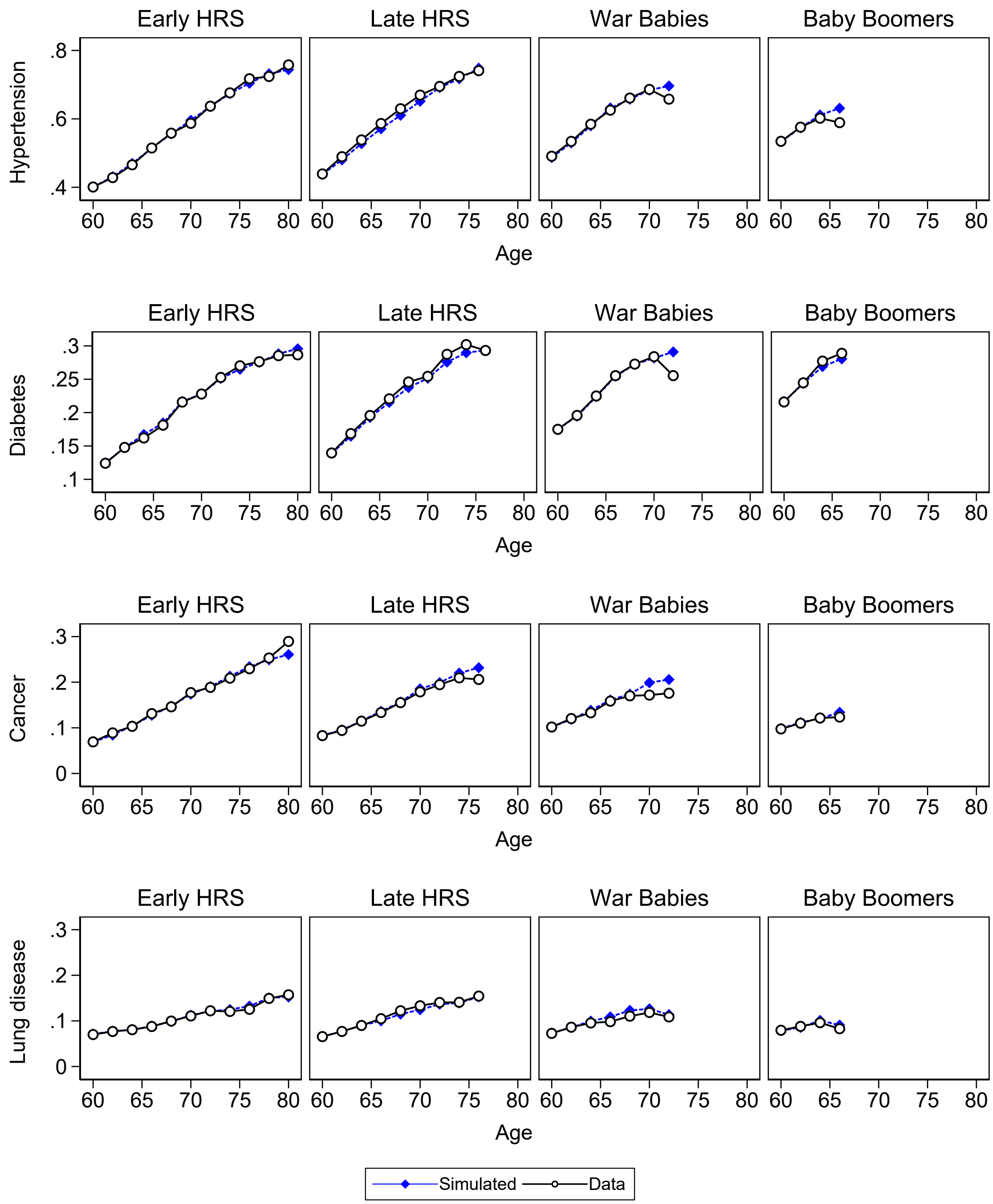

Figure 13: Mean of life-cycle morbidity profiles by cohort

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).

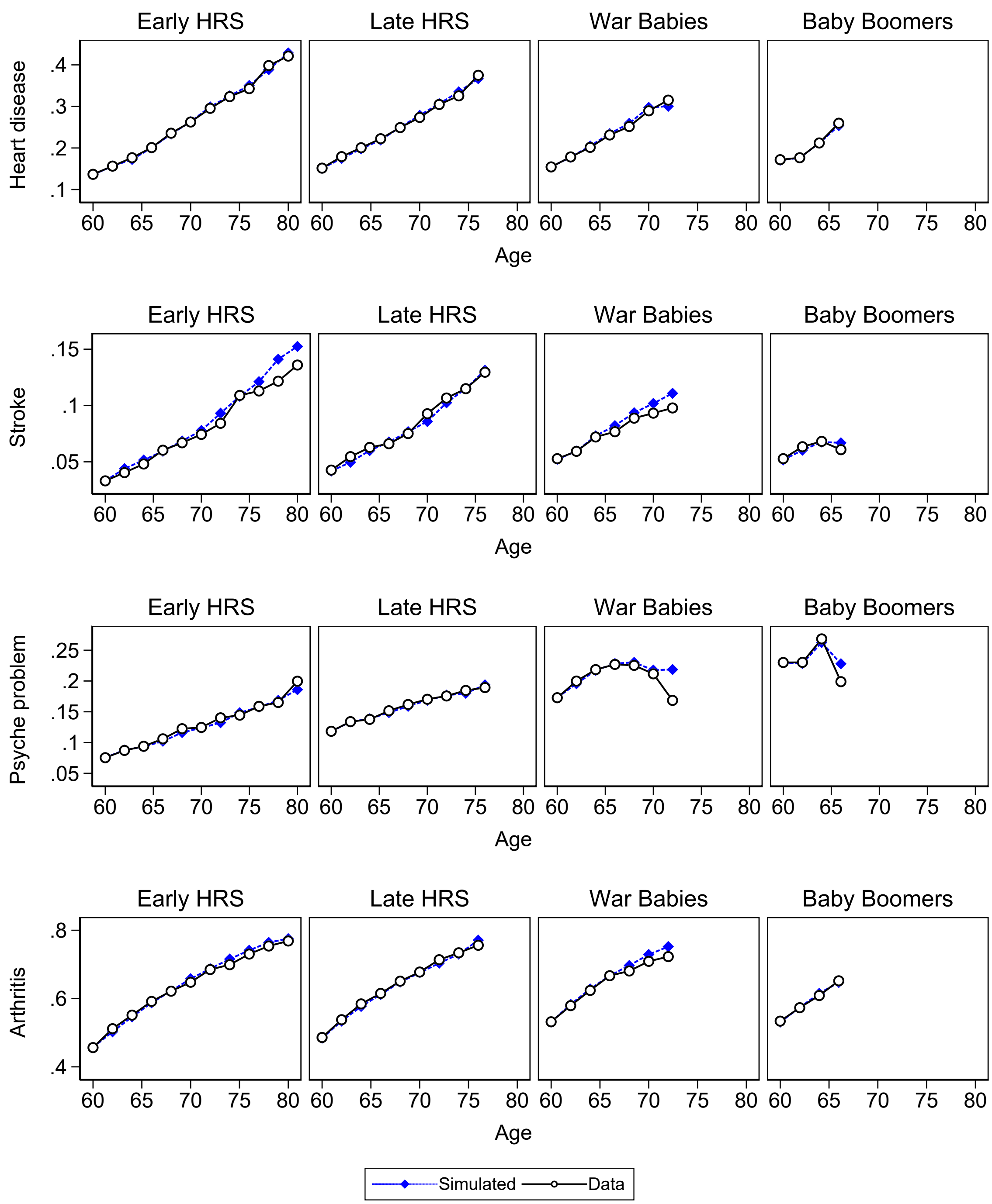

Figure 14: Mean of life-cycle morbidity profiles by cohort

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).

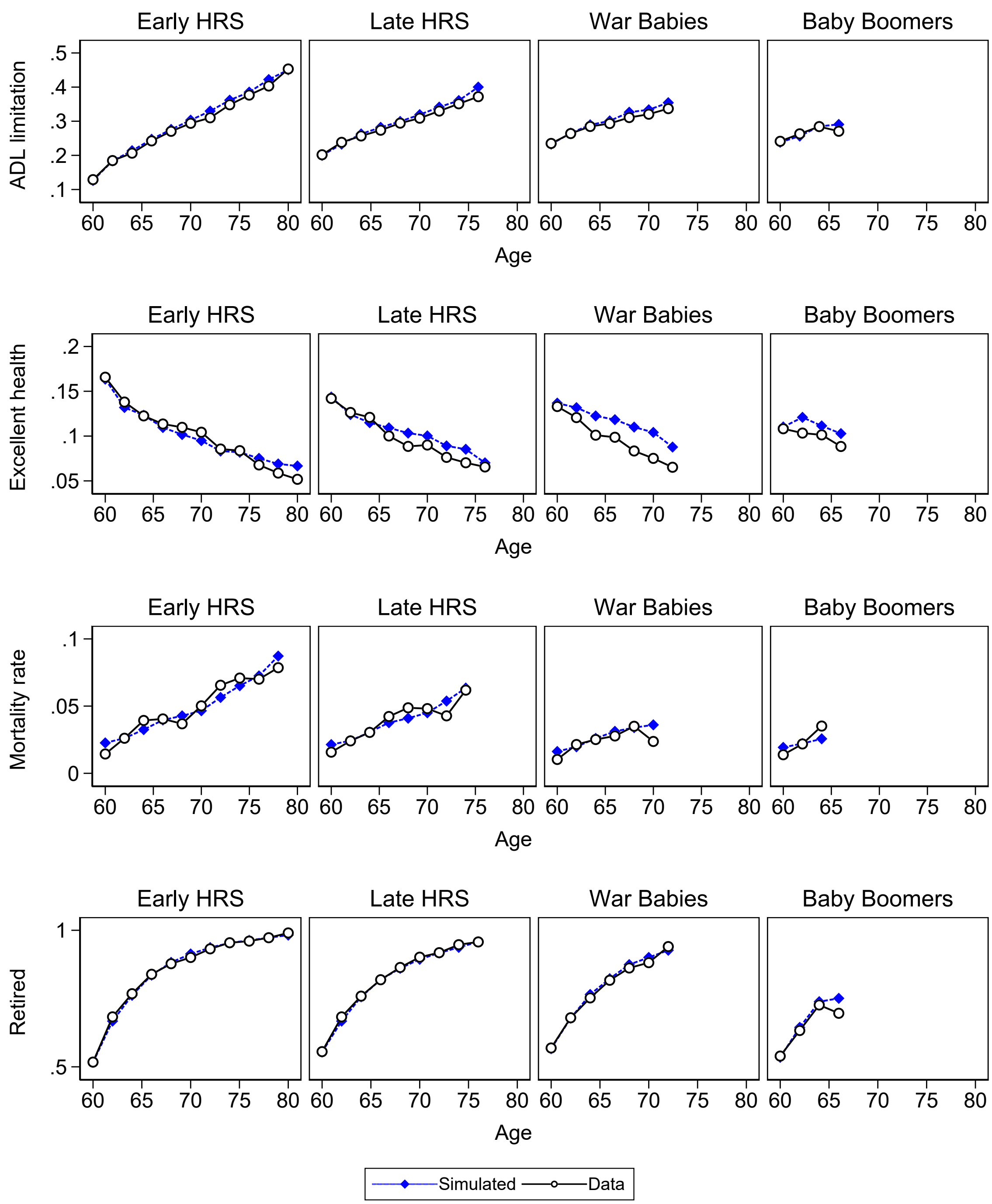

Figure 15: Mean of life-cycle health, mortality, and retirement profiles by cohort

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).

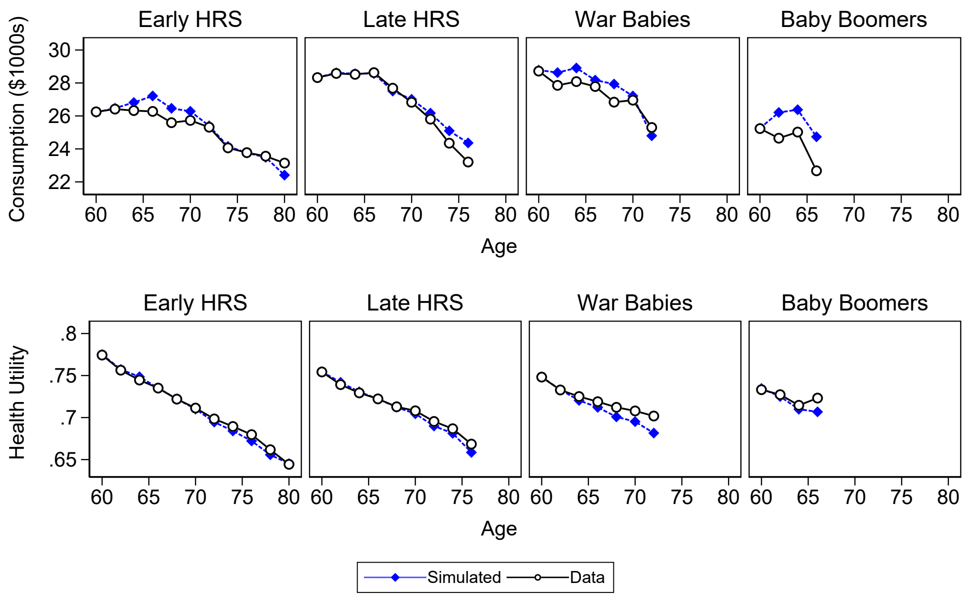

Figure 16: Mean of life-cycle consumption and health utility profiles by cohort

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).

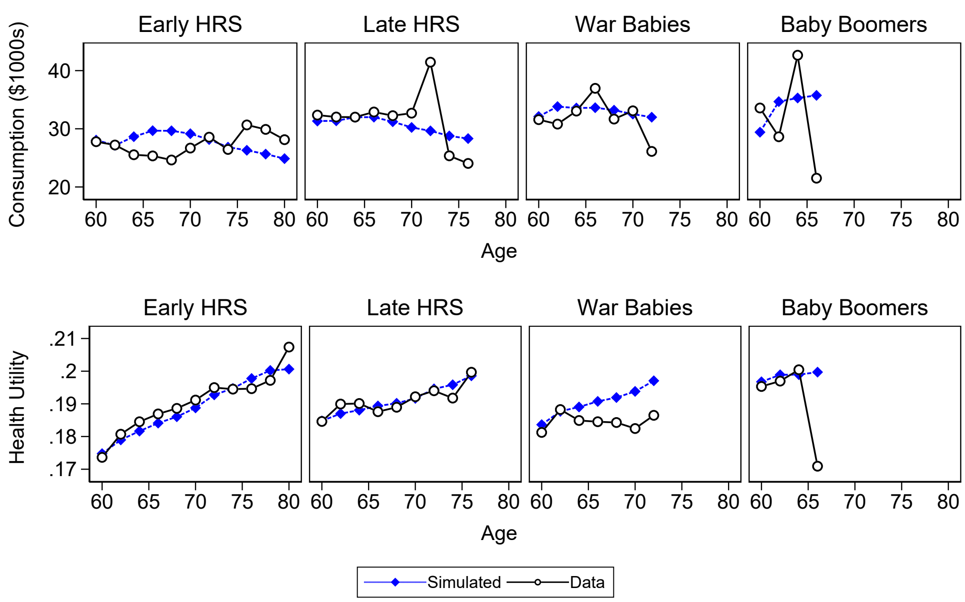

Figure 17: Standard deviation of consumption and health utility life-cycle profiles by cohort

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).

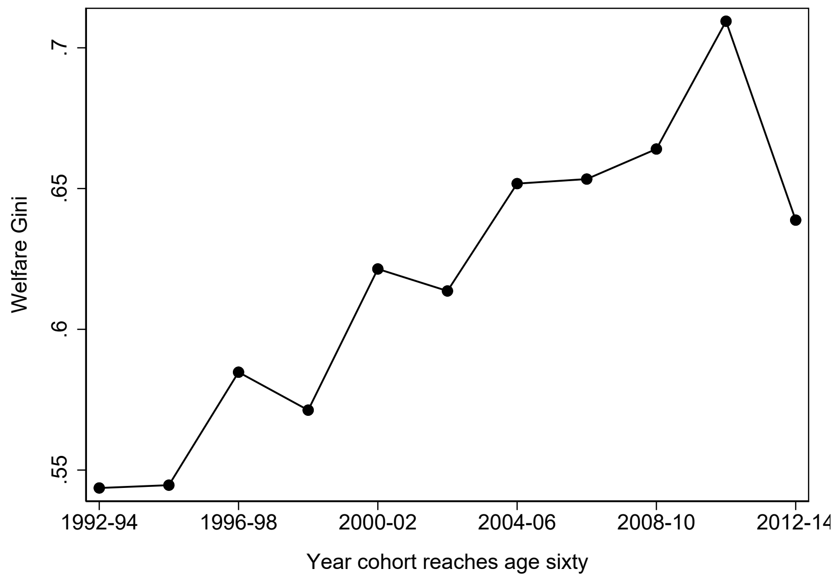

Figure 18: Welfare (λ) Gini by two-year birth cohort

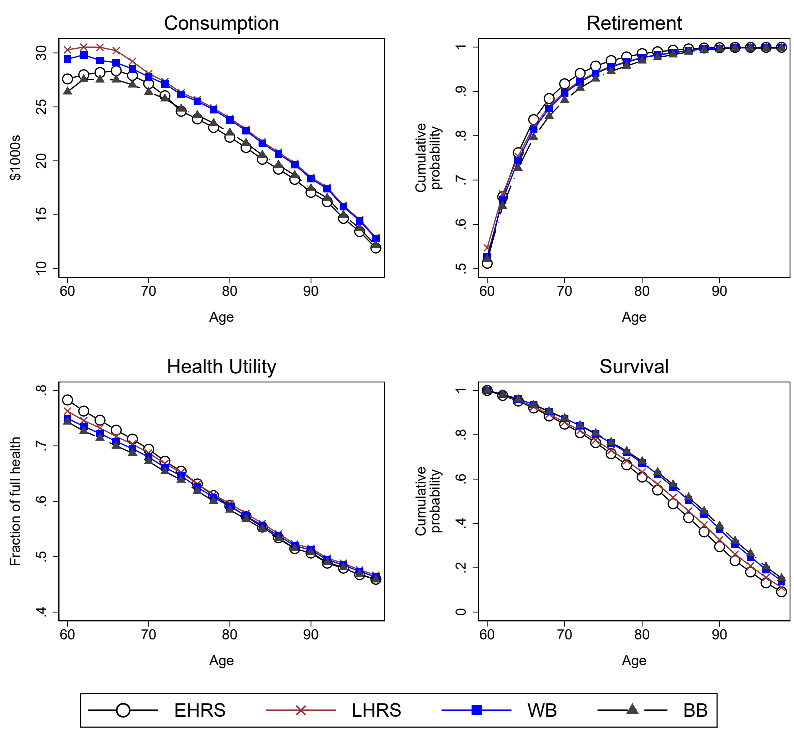

Figure 19: Average life-cycle profiles by cohort

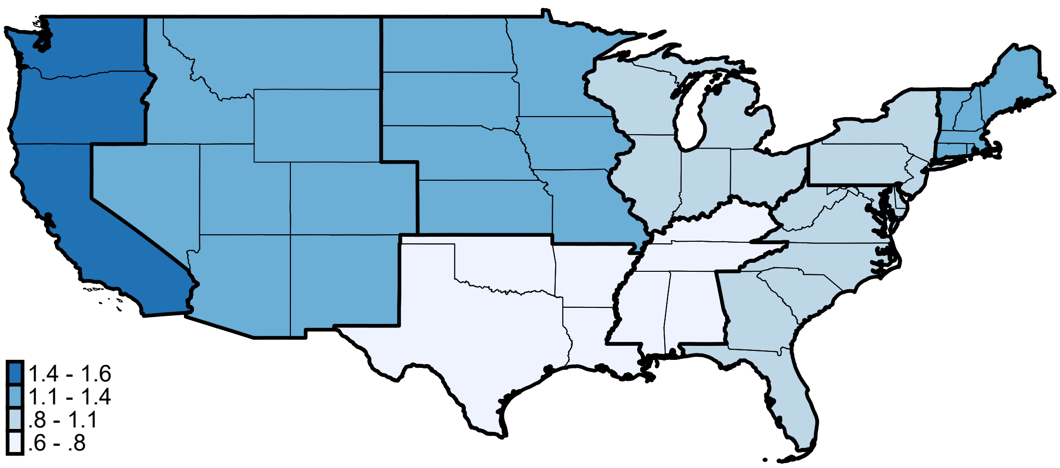

Figure 20: Median welfare by census division for HRS cohort

Table 13: Model estimates for morbidities

	Hypertension		Diabetes		Cancer		Lung disease		Heart disease		Stroke		Psych		Arthritis	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Lag Hyper			0.269	0.034	-0.046	0.040	0.063	0.041	0.116	0.033	0.084	0.041	0.159	0.037	0.084	0.033
Lag Diab	0.262	0.051			0.074	0.047	0.001	0.052	0.033	0.044	0.059	0.053	-0.015	0.052	0.047	0.046
Lag Cancer	-0.029	0.052	0.002	0.054			0.041	0.060	-0.123	0.052	-0.024	0.059	-0.049	0.060	0.024	0.052
Lag Lung	0.062	0.060	0.044	0.059	0.103	0.059			0.243	0.052	-0.000	0.064	0.128	0.062	0.146	0.065
Lag Heart	0.085	0.044	0.075	0.041	0.027	0.042	0.189	0.042			0.165	0.041	0.084	0.042	0.109	0.042
Lag Stroke	0.071	0.068	-0.023	0.064	-0.027	0.061	0.011	0.063	0.052	0.055			0.258	0.053	-0.046	0.063
Lag Psych	0.097	0.054	0.064	0.053	-0.076	0.060	0.084	0.057	0.107	0.048	0.148	0.055			0.249	0.055
Lag Arthritis	0.078	0.030	-0.013	0.034	-0.013	0.035	0.123	0.038	0.070	0.031	-0.011	0.037	0.104	0.036		
Lag ADL	0.067	0.033	0.021	0.036	-0.005	0.036	0.102	0.036	0.056	0.031	0.171	0.034	0.209	0.033	0.157	0.037
Lag Health 2	0.015	0.035	-0.022	0.033	-0.035	0.035	-0.076	0.033	-0.102	0.030	-0.121	0.033	-0.194	0.031	-0.088	0.038
Lag Health 3	0.012	0.036	-0.028	0.035	-0.051	0.037	-0.136	0.036	-0.158	0.032	-0.214	0.035	-0.266	0.034	-0.112	0.039
Lag Health 4	-0.025	0.038	-0.107	0.038	-0.079	0.039	-0.293	0.041	-0.245	0.034	-0.241	0.039	-0.344	0.039	-0.154	0.041
Lag Health 5 (best)	-0.107	0.042	-0.242	0.047	-0.117	0.046	-0.448	0.057	-0.290	0.041	-0.358	0.051	-0.455	0.052	-0.250	0.045
Lag2 Hyper			0.025	0.034	0.064	0.040	-0.077	0.041	0.047	0.033	0.060	0.041	-0.101	0.037	0.018	0.033
Lag2 Diab	-0.079	0.055			-0.083	0.051	-0.092	0.055	0.114	0.046	0.082	0.055	0.034	0.055	-0.045	0.049
Lag2 Cancer	0.008	0.056	0.004	0.058			0.057	0.064	0.139	0.056	0.006	0.063	0.073	0.064	0.039	0.056
Lag2 Lung	-0.134	0.066	-0.057	0.064	0.027	0.064			-0.087	0.058	0.056	0.069	-0.006	0.067	-0.050	0.072
Lag2 Heart	-0.028	0.047	0.008	0.043	-0.006	0.044	-0.062	0.044			-0.013	0.043	-0.053	0.045	-0.041	0.045
Lag2 Stroke	-0.031	0.077	0.026	0.070	0.023	0.067	-0.029	0.069	0.081	0.061			-0.188	0.059	0.047	0.070
Lag2 Psych	-0.045	0.058	-0.068	0.057	0.067	0.063	0.048	0.060	-0.039	0.052	-0.041	0.058			-0.114	0.060
Lag2 Arthre	-0.035	0.030	0.015	0.034	0.061	0.035	-0.012	0.036	0.031	0.030	-0.000	0.037	-0.011	0.035		
Lag2 ADL	-0.068	0.036	0.043	0.039	0.007	0.038	-0.015	0.037	0.023	0.032	-0.091	0.035	-0.062	0.034	-0.053	0.042
Lag2 Health 2	-0.034	0.037	-0.058	0.034	-0.047	0.037	-0.082	0.035	0.001	0.033	-0.065	0.035	-0.077	0.033	0.049	0.042
Lag2 Health 3	-0.036	0.038	-0.080	0.036	-0.015	0.038	-0.130	0.037	-0.010	0.034	-0.058	0.037	-0.137	0.036	0.077	0.043
Lag2 Health 4	-0.058	0.040	-0.126	0.039	-0.016	0.040	-0.192	0.041	-0.029	0.037	-0.044	0.041	-0.227	0.040	0.050	0.044
Lag2 Health 5	-0.069	0.043	-0.154	0.046	-0.003	0.046	-0.293	0.053	-0.098	0.042	-0.080	0.050	-0.285	0.050	-0.030	0.048
Time	0.051	0.008	0.017	0.009	-0.001	0.008	0.007	0.010	-0.011	0.007	-0.027	0.009	0.003	0.009	-0.035	0.008
2008+	-0.072	0.027	-0.041	0.030	0.022	0.030	0.009	0.035	-0.033	0.027	0.021	0.033	-0.102	0.033	0.047	0.028
CODA	-0.040	0.039	-0.021	0.043	-0.008	0.040	0.006	0.045	-0.009	0.036	0.021	0.039	0.079	0.042	-0.107	0.039
Early HRS	-0.084	0.054	-0.025	0.060	-0.062	0.056	-0.052	0.064	0.026	0.050	0.008	0.056	0.095	0.060	-0.109	0.055
Late HRS	-0.084	0.069	-0.019	0.076	-0.067	0.072	0.017	0.082	0.049	0.064	0.025	0.072	0.148	0.077	-0.031	0.069
War Babies	-0.099	0.084	0.047	0.093	-0.043	0.089	0.007	0.102	0.100	0.079	0.115	0.091	0.285	0.094	0.088	0.085
Boomers	-0.172	0.103	0.064	0.115	-0.054	0.109	-0.000	0.126	0.138	0.097	0.114	0.111	0.399	0.114	0.121	0.104
Mid Boomers	-0.327	0.123	0.004	0.138	-0.009	0.134	0.027	0.153	0.189	0.118	0.152	0.142	0.391	0.137	0.166	0.123
Black	0.190	0.022	0.076	0.022	-0.029	0.023	-0.169	0.027	-0.129	0.021	0.045	0.024	-0.193	0.026	-0.007	0.021
Other race	0.032	0.032	0.216	0.035	-0.193	0.045	-0.113	0.046	-0.105	0.037	-0.144	0.050	-0.039	0.041	-0.047	0.034
Female	0.031	0.016	-0.121	0.018	-0.204	0.018	-0.057	0.021	-0.181	0.016	-0.069	0.020	0.115	0.019	0.161	0.016
HS grad	-0.038	0.019	-0.084	0.020	0.005	0.021	-0.065	0.023	0.019	0.018	0.043	0.022	-0.066	0.022	-0.027	0.020
Some college	-0.072	0.022	-0.092	0.024	0.047	0.024	-0.036	0.027	0.037	0.022	0.057	0.026	-0.014	0.026	0.006	0.022
College grad	-0.109	0.025	-0.165	0.028	0.049	0.028	-0.173	0.034	-0.025	0.025	0.030	0.031	-0.047	0.031	-0.037	0.025
Lag Retired	-0.008	0.030	0.061	0.032	0.029	0.035	0.049	0.042	0.002	0.032	0.043	0.044	0.073	0.040	0.006	0.029
Lag2 Retired	-0.010	0.029	-0.061	0.031	-0.003	0.033	0.002	0.039	0.004	0.031	0.040	0.040	-0.026	0.037	-0.040	0.029
Constant	-1.626	0.088	-1.937	0.097	-1.943	0.096	-2.016	0.110	-1.749	0.087	-2.510	0.116	-1.862	0.100	-1.277	0.090

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

Table 14: Model estimates for ADLs, self-rated health, retirement, consumption, and mortality

	ADLs		Self-rated health		Retirement		Consumption		Mortality	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Hyper			-0.271	0.015	0.067	0.037	-0.010	0.012	0.083	0.027
Diab			-0.272	0.019	0.063	0.050	-0.036	0.018	0.118	0.033
Cancer			-0.690	0.020	0.177	0.053	0.035	0.021	0.658	0.027
Lung			-0.481	0.024	0.188	0.070	-0.008	0.026	0.406	0.033
Heart			-0.496	0.017	0.123	0.046	-0.006	0.019	0.175	0.025
Stroke			-0.496	0.023	0.471	0.076	-0.074	0.020	0.235	0.030
Psych			-0.416	0.022	0.395	0.059	-0.089	0.017	0.228	0.030
Arthritis			-0.218	0.015	0.034	0.035	0.015	0.013	-0.026	0.024
ADL			-0.669	0.014	0.377	0.040	-0.069	0.018	0.333	0.020
Health 2					-0.603	0.047	0.060	0.019	-0.329	0.017
Health 3					-0.769	0.048	0.079	0.022	-0.534	0.019
Health 4					-0.779	0.050	0.113	0.022	-0.651	0.023
Health 5 (best)					-0.760	0.055	0.142	0.026	-0.642	0.033
Lag Hyper	0.041	0.031	0.143	0.020	-0.021	0.050	-0.003	0.012	-0.036	0.027
Lag Diab	0.095	0.040	0.096	0.026	-0.034	0.072	0.011	0.016	0.059	0.035
Lag Cancer	0.028	0.045	0.544	0.029	-0.094	0.081	-0.009	0.023	-0.451	0.030
Lag Lung	0.183	0.049	0.216	0.034	-0.024	0.104	-0.001	0.021	-0.103	0.036
Lag Heart	0.061	0.034	0.293	0.023	-0.154	0.069	0.008	0.018	-0.015	0.026
Lag Stroke	0.379	0.047	0.358	0.033	-0.213	0.128	0.004	0.019	-0.046	0.033
Lag Psych	0.335	0.044	0.230	0.031	-0.126	0.088	0.036	0.019	-0.135	0.033
Lag Arthritis	0.198	0.026	0.107	0.018	0.068	0.046	-0.009	0.013	-0.078	0.024
Lag ADL			0.326	0.018	-0.168	0.059	0.015	0.018	-0.112	0.020
Lag Health 2	-0.228	0.031	0.623	0.014	0.017	0.063	0.015	0.011	-0.031	0.019
Lag Health 3	-0.468	0.032	1.123	0.015	-0.017	0.064	0.013	0.012	-0.056	0.021
Lag Health 4	-0.634	0.034	1.643	0.016	-0.047	0.066	0.012	0.011	-0.095	0.024
Lag Health 5	-0.719	0.041	2.259	0.019	-0.062	0.069	0.018	0.015	-0.127	0.032
Time	-0.057	0.007	0.022	0.003	-0.021	0.010	0.001	0.009	-0.025	0.005
2008+	0.050	0.026	0.008	0.012	-0.004	0.033	-0.037	0.011	0.045	0.021
CODA	0.094	0.032	0.028	0.016	0.079	0.080			-0.020	0.025
Early HRS	0.119	0.046	0.023	0.023	0.070	0.096			-0.063	0.035
Late HRS	0.125	0.059	0.005	0.029	0.000	0.111			-0.075	0.046
War Babies	0.178	0.073	-0.011	0.036	0.053	0.128			-0.142	0.058
Boomers	0.248	0.090	-0.100	0.043	0.047	0.150			-0.120	0.072
Mid Boomers	0.367	0.107	-0.168	0.052	0.034	0.172			-0.255	0.095
Black	0.082	0.019	-0.056	0.009	0.051	0.024			0.062	0.017
Other race	0.026	0.034	-0.101	0.015	-0.052	0.038			-0.066	0.032
Female	-0.001	0.015	0.035	0.007	0.112	0.018			-0.223	0.013
HS grad	-0.089	0.017	0.085	0.008	-0.018	0.024			0.033	0.015
Some college	-0.045	0.020	0.121	0.009	-0.039	0.026			0.021	0.018
College grad	-0.087	0.023	0.200	0.011	-0.038	0.029			0.008	0.021
Retired							-0.031	0.012	0.203	0.033
Lag Retired	0.085	0.029	-0.025	0.013			-0.038	0.012	-0.019	0.028
Lag2 Retired	0.019	0.027	-0.017	0.013						
Lag Con							0.173	0.005		
Lag2 Con							0.084	0.005		
Constant	-0.900	0.080			-0.853	0.206			-1.723	0.303

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

Table 15: Morbidity shock covariance matrix (Σ)

	Hyper	Diabetes	Cancer	Lung	Heart	Stroke	Psych	Arthritis	ADLs
Hyper	1.00	0.26	0.04	0.08	0.28	0.29	0.14	0.09	0.09
Diabetes	0.26	1.00	0.07	0.05	0.10	0.14	0.06	0.03	0.08
Cancer	0.04	0.07	1.00	0.12	0.03	0.06	0.12	0.05	0.13
Lung	0.08	0.05	0.12	1.00	0.22	0.10	0.17	0.09	0.19
Heart	0.28	0.10	0.03	0.22	1.00	0.28	0.16	0.10	0.14
Stroke	0.29	0.14	0.06	0.10	0.28	1.00	0.21	0.10	0.40
Psych	0.14	0.06	0.12	0.17	0.16	0.21	1.00	0.15	0.29
Arthritis	0.09	0.03	0.05	0.09	0.10	0.10	0.15	1.00	0.26
ADLs	0.09	0.08	0.13	0.19	0.14	0.40	0.29	0.26	1.00

Table 16: Welfare decomposition in HRS cohort and welfare Gini for each cohort by select characteristics

			Decomposition			Welfare Gini by cohort			
	Median λ	Mean $\log \lambda$	Cons.	Leisure	QALY	EHRS	LHRS	War	Boomers
Education									
<HS	0.444	-0.802	-0.393	0.031	-0.440	0.434	0.478	0.506	0.523
HS grad	1.058	-0.020	-0.015	0.019	-0.025	0.430	0.493	0.542	0.536
Some college	1.402	0.271	0.196	0.006	0.069	0.477	0.526	0.578	0.565
College grad	2.536	0.893	0.476	-0.012	0.429	0.479	0.539	0.573	0.625
Gender									
Male	0.862	-0.150	0.045	-0.005	-0.190	0.525	0.564	0.605	0.624
Female	1.182	0.083	-0.030	0.031	0.081	0.548	0.624	0.660	0.693
Race									
White	1.112	0.070	0.063	0.013	-0.005	0.534	0.594	0.633	0.665
Black	0.457	-0.742	-0.404	0.028	-0.366	0.498	0.537	0.539	0.587
Other	0.771	-0.304	-0.245	0.011	-0.070	0.533	0.677	0.606	0.683

Notes: Estimates use base year respondent analysis weights.