Endogenous Health Groups and Heterogeneous Dynamics of the Elderly*

Dante Amengual
CEMFI
<amengual@cemfi.es>

Jesús Bueren
European University Institute
<jesus.bueren@eui.eu>

Julio A. Crego
Tilburg University
<jacrego@uvt.nl>

*We would like to thank Nezih Guner, Barton Hamilton, Pedro Mira, Josep Pijoan-Mas, and Silvia Tiezzi for their comments and suggestions. We have also benefited from the feedback provided by seminar participants at CEMFI, as well as audiences at the VII Workshop in Time Series Econometrics (Zaragoza), the 26th European Workshop on Econometrics and Health Economics (Prague), Annual Health Econometric Workshop 2017 (Washington University in St. Louis), Universidad Carlos III de Madrid, and SED meetings 2019 (St. Louis). Financial support from the Spanish Ministry of Economy & Competitiveness through grant 2014-59262 (Amengual) and Santander Research chair at CEMFI (Crego) is gratefully acknowledged.

Abstract

We propose a novel methodology to classify individuals into groups of health and characterize their transition across these groups. We estimate a panel Markov switching model that exploits information from the cross-sectional and time series dimensions. Using the HRS, we identify four persistent health groups, depending on individual's physical and mental disabilities, with heterogeneous transitions across gender and education. Our classification outperforms existing health measures at explaining entry in nursing homes, home health care, out-of-pocket medical expenses and mortality. Through a workhorse model of savings, we recover an asset cost of bad health twice as big as using self-reported health.

Keywords: Latent groups, Frailty, Long-Term Care, Medical expenses.

JEL: C23, C38, E21, I14.

1 Introduction

Health is an important input in structural models because of its role as a determinant of survival, savings, and insurance choices. In spite of the rich information available from retirement surveys describing individual's health state, researchers need to rely on only one variable to avoid the curse of dimensionality. The choice varies depending on the question at hand, for example self-reported health (SRH) to predict mortality; mobility or cognition problems to capture needs for long-term care (LTC) or a combination of medical expenses and physical/cognitive limitations to capture the correlation between mortality and medical costs. Different choices hampers comparisons across papers and thus our understanding on how health affects welfare overall. Therefore, a discrete health measure based on objective and widely-available variables that captures both mortality and long-term care risks would contribute to the comparability of future studies.

In this paper, we produce such a measure by estimating the health status of individuals using a novel dynamic latent variable model that exploits panel-data information on objective health measures: twelve dummy variables on reported difficulty with Activities of Daily Living ADLs and Instrumental Activities of Daily Living (IADLs). The econometric model assumes that individuals belong to one out of a prespecified finite number of latent health groups. In the cross-section, individuals belonging to different health groups differ in the probability of reporting difficulties with each I-ADLs. In the time-series, individuals belonging to different groups differ in health transition and survival probabilities. The estimated model parameters define health groups and provide information to classify individuals over time in these groups.

Using the Health and Retirement Study, we find that all I-ADLs can be parsimoniously represented by four health groups, which divide individuals into *physically frail*, *mentally frail*, *impaired*, and *healthy* represent health suitably. The *impaired* have both types of limitations, physical and cognitive, while the *healthy* have no or light difficulties with I-ADLs.[1] In turn, the *physically frail* have limited mobility, while the *mentally frail* have difficulties with more cognitive tasks such as managing money. Importantly, and in line with gerontology literature (e.g. Morris et al., 2013), not all the I-ADLs are equally informative for classifying individuals in health groups. For example, if a person has difficulties with getting in or out of bed, she belongs to the *physically frail* group with a probability higher than one third but to the

[1] Along the paper, we use italics to refer to our states, hence a *healthy* individual is a member of the group we label *healthy*.

mentally frail with a probability lower than 5%. In contrast, an individual incapable of taking medications is much more likely to belong to the *mentally* rather than the *physically frail* group.

To classify each individual at a given point time to one of the groups, the model provides two different sets of probabilities: filtered and smoothed. The filtered probability contains all the health information from the present and past health of the individual; hence, it belongs surely to the information set of the individual. On the contrary, the smoothed probability includes as well information about the future health and survival of the individual. Hence, it provides a more accurate description of the individuals' health but it might not belong to the individual information set. To ensure that our results only rely on information known by the individual, we classify observations according to the filtered probabilities.

At the age of 60, around 80% of individuals are *healthy*, 10% are *physically frail* and the remaining includes almost as many *impaired* as *mentally frail*. As individual age, we observe the same stylized facts previously documented in the literature of aging (Manton and Soldo, 1985): older individuals have relatively worse health, health deteriorates with age, individuals in worse health have larger chances of dying, and females live longer than males. Furthermore, in line with Brown (2002) and Meara et al. (2008), we find a large educational gradient in life expectancy. Nonetheless, despite living longer, educated individuals spend, on average, less time *impaired*, consistent with Pijoan-Mas and Ríos-Rull (2014).

We then compare access to medical and care services across health groups. On average, *impaired* individuals spend $12,475 per year in out-of-pocket (OOP) medical spending while *healthy* ones spend $2,990. Likewise, *mentally frail* individuals spend $1,070 more than *physically frail* ones, who employ $4,602. The use of long-term care (LTC) services also presents large differences across groups. While 7% of the individuals *mentally frail* live in a nursing home at the time of the interview, only 1.3% of the *physically frail* do so. This disparity widens between members of the *healthy* group, who avoid the nursing home almost surely, and those of the *impaired*, out of which 31.8% reside in these facilities. A similar pattern arises if we compare the received professional care of these two extreme groups. Nonetheless, *mentally* and *physically frail* individuals need a medical-trained person to look after them at home with a similar probability.

To assess the predictive power of our classification, we contrast it with other commonly used health classifications, namely, five different levels of self-reported health, whether the individual reports difficulty with any ADL, and the division of a frailty index into five equally sized groups.[2] To do so, we consider three health-related spending, out-of-pocket medical expenditures, and indicators of residing in a nursing home and receiving care, which the macro literature has identified as crucial drivers of savings (De Nardi et al. 2010; Barczyk and Kredler 2018; Ameriks et al. 2020). Our classification generates more differentiated groups in all three variables. This leads to our measure being able to explain a higher proportion of the variance of this variables. For instance, conditional on age, education, and gender, self-reported health can only explain 1.2 percentage points of the variance of OOP medical expenses and 2.8 of the variance of residing in a nursing home. Our measure explains 3.2 and 21.9, respectively. Besides health measures, we analyze the ability of the different classifications to predict mortality and find that our measure dominates the alternative ones.

We explore the consequences of using our measure in a life-cycle model by solving the model proposed by De Nardi et al. (2010).Although the unconditional dissaving pattern of the elderly remains unchanged, they largely differ conditional on health. While De Nardi et al. (2010) predicts a similar dissaving pattern by healthy and unhealthy individuals, our estimation indicates that unhealthy individuals are responsible for most of the dissaving during retirement. This result arises from a tighter correlation between health and medical expenses and larger differences in survival conditional on health inherent in our classification.

Using the same model, we analyze the effect of uncertainty about individual's health. As we do not observe the health status but, instead, we rely on probabilities from our econometric model, we consider the probabilities themselves as the observed health state. The dissaving pattern remains identical to the one with certainty. This result is not surprising given that the average filtered probability of belonging to the modal health group is 94%.

Last but not least, we provide in Bueren's webpage the probabilities of belonging to each estimated health group for the most commonly used retirement surveys: HRS, ELSA, and SHARE; and a different number of health groups, from two to four. In order to ease the comparison across samples, we fix the definition of health groups by keeping the HRS probabilities of reporting difficulties with I-ADLs and reestimate country-specific health transitions.

Literature– Our paper complements the literature analyzing the effect of health on eco-

[2] De Nardi et al. (2010), Kopecky and Koreshkova (2014), Pijoan-Mas and Ríos-Rull (2014), Dobrescu (2015), and De Nardi et al. (2016) rely on self-reported health, Bohacek et al. (2015) on ADLs, and Braun et al. (2017) on a frailty index.

nomic decisions. This literature relies on dynamic structural models to quantify the importance of mechanisms or to derive implications for policymaking. Due to the curse of dimensionality, researchers undertake an ad-hoc decision over which of all the possible health variables from the available surveys to use as a state variable. For instance, De Nardi et al. (2010) and O'Donnell et al. (2015) use self-reported health which provides the best mortality forecast (Idler and Benyamini, 1997); hence, it is ideal to assess the risk of living longer. However, this measure does not capture long-term care needs, which constitute an important component of medical expenses; thus, Ameriks et al. (2016) and Ko (2016) rely on individuals' difficulties with mobility or cognition. Likewise, the subjective nature of self-reported health makes it unfeasible for some applications such as analyzing health and mortality insurance choices for which Koijen et al. (2016) relies on medical expenses and morbidities to construct a measure of health and Braun et al. (2017) constructs a continuous frailty index and divides it into five quintiles. Finally, facing the lack of an optimal measure, some researchers directly use, as the health state, an individual choice such as receiving home care or residing in a nursing home (Barczyk and Kredler, 2018; Kopecky and Koreshkova, 2014).

This paper relates to an extensive literature which proposes econometric methods to analyze different issues in health economics (see Jones, 2000, for a survey). Closely related to our paper is Deb and Trivedi (1997) who show that a finite mixture of negative binomials, characterizing "healthy" and "ill" individuals, explains counts of medical care utilization by the elderly in the U.S. better than previously proposed specifications. They, however, do not classify individuals into the aforementioned categories. Moreover, they disregard health dynamics which is of first-order relevance. Contoyannis et al. (2004) stress the importance of health persistence using a dynamic panel ordered probit model for self-reported health.

Likewise, we contribute to a growing literature that summarizes health variables into a single index that explains most of the variation related to health (see Searle et al., 2008). Regarding HRS, Yang and Lee (2009) compute a frailty index based on chronic conditions, ADLs, IADLs, depressing symptoms, self-reported health, and obesity. Nonetheless, its continuous nature prevents researchers to include it in structural models. One exception is Bound et al. (2010) who considers health as a continuous latent variable and include it into a structural model to analyze retirement. To be able to solve the model, they assume that individuals are completely unable to self-insure against medical expenses.

On the econometric modeling side, our paper closely relates to Todd and Zhang (2020) who also consider discrete latent states across which individuals can move over time. They use the latent states to represent types of individuals according to personality and unobserved educational and professional characteristics in order to understand educational and occupational choices. Besides the context, our latent model differs from theirs on the transitions across states. Even if both are first-order markovian conditional on the latent state, our model allows for any transition probability matrix while they assume that once an individual changes state, her probability of belonging to any state is independent of the past state. In our context, this assumption would imply that the probability of recovering is the same as the probability of getting worse. Likewise, they assume that the probability of staying in the same state is the same across states. In terms of health, this assumption would imply that a healthy individual maintains his good health with the same probability as an unhealthy individual remains in poor health. Both assumptions do not hold in the data for our latent groups.

Our paper contributes to the recent panel data literature on groups clustering such as Wang et al. (2018) who allow for heterogeneous-across but homogeneous-within groups slope coefficients. While we also have unobserved individual's health status classified within groups, we explicitly identify their changes across time. Similarly, Bonhomme and Manresa (2015) restrict individuals to belong to the same group forever but they allow the group characteristics to change over time. In our context, this feature forces every individual in a group to have the same dynamics. Using physically impaired as an example, their model would imply that either everyone that is physically impaired remains physically impaired, or everyone recovers, or everyone gets worse. While this characteristic makes a lot of sense in their application, it is unrealistic in our setup. With a different objective, Cunha et al. (2010) propose a structural model for skill formation. Our paper relate to theirs in the non-lineal filtering of unobserved variables, though ours are discrete in nature, and, since our model is not structural, there is no need for identifying unobserved optimal decisions.

The rest of the paper is structured as follows. We briefly describe the HRS data in Section 2. Then, the econometric model and the estimation strategy are presented in Section 3. Next, we present the main results in Section 4 and we compare our proposed classification with alternative ones in Section 5. Finally, Section 6 concludes.

2 HRS and I-ADLs

Our data comes from the RAND HRS dataset which comprises a cleaned version of the Health and Retirement Study conducted by the University of Michigan.[3] It contains subjective and objective indicators of health, as well as demographic and economic characteristics, of a representative panel of US households surveyed biannually from 1992 to 2014. In addition, the HRS exit interview records the death of the individual and includes the answers from a proxy informant. The completeness of this data source has led to its omnipresence in the recent literature.

Since not all the variables used in the estimation are available for early waves, we restrict the sample from 1996 until 2014, which includes ten waves. Moreover, to focus on health needs, we select individuals over 60 years old. The final sample, after excluding individuals whose education, gender or age are missing ($<$0.1% of observations), consists of 159,025 interviews (including exit waves), which corresponds to 27,369 individuals followed on average six waves (12 years). The composition of the sample reflects the survival probabilities. While the median age is 72 years, the share of individuals is decreasing in age as they die. Likewise, females account for 58% of the sample as their life expectancy is higher than the males' one. In terms of education, 72% of individuals completed high school which constitutes 74% of the sample due to its superior life expectancy.

The HRS provides dozens of health-related variables, but we restrict to individual's ability to perform ADLs and IADLs to infer the health status. ADLs were proposed by Katz et al. (1963) as a measure of how independent a patient is, and consequently, they include very basic activities such as if they can walk or dress. IADLs, in contrast, consist of activities more closely related with cognition as the possibility of using a phone or controlling her medication. Accordingly, these variables relate to the need for LTC which is the dimension of health we aim to identify. Although our model could incorporate more information, reducing the set of variables eases the interpretation of the groups. Besides, by excluding other variables, we can use them to compare the performance of our classification against other alternatives.

Precisely, we utilize twelve binary variables, denoted as I-ADLs, which include six ADLs and six IADLs that describe whether individuals have any difficulty to perform these types

[3]Version P. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (August 2016). The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

of basic tasks. We extract this information from the HRS questionnaire to which respondents select one out of six possible answers: *Yes* and *Can't Do* that we label as 1, *No* to which we assign a value of 0, and *Don't Do*, *Don't Know*, and *Refuse to answer*, which are recorded as missing. Table 1 defines the activities included in the HRS and provides the proportion of observations in which an individual declares to have difficulties realizing each of them. The most common ADL is not being capable of dressing (12%) whereas eating is the ADL that present fewer difficulties (5%). Likewise, the frequency of IADLs differs across activities from 5% of respondents who claim to face problems when taking medications to 15% that struggle reading a map. Table 1 also indicates that 21% of individuals report difficulties with at least one ADL; meanwhile, 23% of them encounter problems when they carry out one or more IADLs. Altogether, 30% of respondents battle with at least one I-ADL.

[Table 1 about here.]

The HRS also includes a question to qualify respondent's self-reported health (SRH). Since another strand of the literature hinges on subjective measures of health to classify individuals, in the last five columns of Table 1 we compare this measure with the answers related to ADLs and IADLs. Not surprisingly, we observe that as people report worse health, they are more likely to present problems with I-ADLs, nonetheless, the importance of each activity differs. In particular, individuals reporting poor health are not able to walk, dress, or bath with probabilities around 40%, while for the remaining three ADLs the corresponding figures barely surpass 30%. Similarly, difficulties with IADLs are also diverse within the worst self-reported health groups since 50% of individuals endeavor to shop but only 20% encounter complications to take their medications.

3 Econometric model

We have an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t_i = 0, \ldots, T_i$ periods which correspond from age a_0^i to age $a_{T_i}^i$. For each individual, we observe K dummy variables corresponding to each I-ADL across time $(x_{1,i,t}, x_{2,i,t}, \ldots, x_{K,i,t})$, provided the individual is alive and interviewed. All or some of the variables for a given individual who is alive can also be missing for some period t_i. We take missing observations into account under the assumption that they occur completely at random, but we abstract from them in the model description to simplify the exposition.

We assume that the main source of heterogeneity in the population is represented by a finite number of possible health groups or clusters which are not observed by the researcher. Individuals belong to each cluster with a probability which depends on education, e; age, a; gender, s; and the current health cluster but it is independent of the previous individual's health (Markov first-order property). Besides transiting across health groups, individuals may also die, which we represent as an observable and absorbing state, D.

Specifically, we consider that individual i at time t belongs to a health group $h_{i,t}$ out of H possible ones. Given her group is g, the probability of facing difficulties with the k'th I-ADL, say $x_{i,k,t} = 1$, is $\mu_{k,g}$. Under the assumption that I-ADLs are independently distributed conditional on the health status, the joint distribution of $\mathbf{x}_{i,t} = (x_{1,i,t}, x_{2,i,t}, \ldots, x_{K,i,t})'$ is characterized by

$$p(\mathbf{x}_{i,t}|\mu_g, h_{i,t} = g) = \prod_{k=1}^{K} \mu_{k,g}^{x_{k,i,t}} (1 - \mu_{k,g})^{1-x_{k,i,t}}, \tag{1}$$

where $\mu_g = (\mu_{1,g}, \mu_{2,g}, \ldots, \mu_{K,g})'$. Therefore, individuals within the same health group have the same probabilities of experiencing problems with an I-ADL whereas these probabilities might vary if individuals do not belong to the same group. Similarly, the same individual might face a different likelihood regarding I-ADLs if she changes groups during her life.

In favor of parsimony, we model health outcomes as independent across time and individuals *conditional* on the health group. In the case of I-ADLs, it seems plausible that their persistent component is only due to health, nonetheless, the model can accommodate other types of persistence if the researcher wants to extend the set of conditioning variables. We take into account health dynamics by explicitly modeling the transition probabilities across groups. In particular, an individual i at time t who belongs to group g transits to group c with probability

$$Pr[h_{i,t+1} = c|a_{it}, s_i, e_i, h_{i,t} = g] = \frac{\exp[f_{g,c}(a_{it}, s_i, e_i)]}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_i, e_i)]} \tag{2}$$

where $\mathcal{H}$ is the set that contains the H health groups. The remaining possible event is that the individual dies, which is an observable state that occurs with probability

$$Pr[h_{i,t+1} = D|a_{it}, s_i, e_i, h_{i,t} = g] = \frac{1}{1 + \sum_{c \in \mathcal{H}} \exp[f_{g,c}(a_{it}, s_i, e_i)]}.$$

This specification allows health groups to own distinct dynamics as parameters differ according to the current health group. Moreover, to capture within-group heterogeneity, transition probabilities can depend on age, gender and education level through the function $f_{g,c}(a, s, e)$

whose parametric specification is given by

$$f_{g,c}(a, s, e) = \beta_{1,g,c} + \beta_{2,g,c}a + \beta_{3,g,c}s + \beta_{4,g,c}e + \beta_{5,g,c}(a \times s) + \beta_{6,g,c}(a \times e).$$

3.1 Posterior simulation

We aim to recover the posterior of all the parameters and the latent variables that classify the health group to which each individual belongs at each point in time. To do so, we use a Gibbs sampling procedure to estimate the models for different choices of the number of health groups H. In essence, this amounts to reducing a complex problem, that is, sampling from the joint posterior distribution of both parameters and state variables, into a sequence of tractable ones, i.e., sampling from conditional distributions for a subset of the parameters conditional on all the other parameters, for which the literature already provides a solution.

We define $\mathbf{H} = \{\mathbf{h}_i\}_{i=1}^{N}$, where $\mathbf{h}_i = \{h_{i,t}\}_{t=1}^{T_i}$, as the collection of health groups after the first time period, and $\mathbf{H}_0 = \{h_{i,0}\}_{i=1}^{N}$ the collection of health states at the first time period. Further, we denote the vectors stacking the parameters of the I-ADLs process and the transition probabilities as μ and β. In addition, we include in $\mathbf{X}$ the data we observe; that is, age, gender, education, if the individual is death or alive, and her situation in terms of ADLs and IADLs. The Metropolis-within-Gibbs algorithm involves sampling sequentially from several blocks. Specifically, iteration m involves:

1. $p(h_{i,0}^{(m)}|\beta^{(m-1)}, \mu^{(m-1)}, \mathbf{X})$: sampling the initial condition using Hamilton (1989)'s smoother.

2. $p(\mathbf{h}_i^{(m)}|\beta^{(m-1)}, \mu^{(m-1)}, \mathbf{X}, \mathbf{H}_0^{(m-1)})$: sampling the latent health indicator for each $i = 1, ..., N$ and $t > 0$ using the Kim (1994)'s smoother.

3. $p(\beta^{(m)}|\mu^{(m-1)}, \mathbf{H}^{(m)}, \mathbf{H}_0^{(m)}, \mathbf{X})$: sampling the transition parameters (Metropolis).

4. $p(\mu^{(m)}|\beta^{(m)}, \mathbf{H}^{(m)}, \mathbf{H}_0^{(m)}, \mathbf{X})$: sampling the Bernoulli mixture parameters (Metropolis).

The empirical results shown in the next sections are based on 40,000 draws. The first 2,000,000 draws are disregarded as burn-in and, of the remaining 4,000,000, one every 100 draws is retained.

3.1.1 Sampling the states: Kim's Smoother

We sample the first health state of each individual directly from Hamilton's smoother. To sample the remaining states, we apply the methodology developed by Kim (1994):

1. Using the filter proposed in Hamilton (1989) we obtain $p(h_{i,T} = g|\beta, \mu, \mathbf{X})$ for all $g \in \mathcal{H}$.

2. We sample $h_{i,T}$ from $p(h_{i,T}|\beta, \mu, \mathbf{X})$.

3. Similarly, we sample $h_{i,t}$ conditional on β,μ, $\mathbf{X}$ and $h_{i,t+1}$, using the following result: $p(h_{i,t} = g|\beta, \mu, \mathbf{X}, h_{i,t+1} = c) = \dfrac{p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)}{\sum_{g\in\mathcal{H}} p(h_{i,t+1} = c|\beta, h_{i,t} = g) \cdot p(\mathbf{x}_{i,t}|\mu, h_{i,t} = g)}$ for all $g, c \in \mathcal{H}$

As a result, each individual has a different probability of belonging to a given group depending on her past, current, and future answers regarding I-ADLs. Moreover, this probability also incorporates information about the individuals' death wave, as well as her age, gender, and education.

3.1.2 Sampling the transition probabilities and the Bernoulli parameters

In this step, we sample from the posterior of the parameters of the Bernoulli distributions and the ones governing the health dynamics (μ, β) conditional on the health groups, $\mathbf{H}$, and the data, $\mathbf{X}$.

Regarding priors, we consider a uniform on $[0, 1]$ for the elements of μ and a diffuse Gaussian prior centered at $\mathbf{0}$ and covariance matrix $100 \cdot \mathbf{I}$ for β. Hence, the posterior of the parameters governing the health dynamics and the one driving the Bernoulli distributions are independent conditional on the latent health group. Precisely, their posterior distributions are given by

$$p(\mu|\mathbf{X}, \mathbf{H}) = \prod_{i=1}^{N}\prod_{t=1}^{T_i} p(\mathbf{x}_{i,t}|h_{i,t}, \mu) \cdot p(\mu)$$

and

$$p(\beta|\mathbf{X}, \mathbf{H}) = \prod_{i=1}^{N}\prod_{t=2}^{T_i} p(h_{i,t}|\beta, h_{i,t-1}) \cdot p(h_{i,1}|\beta) \cdot p(\beta).$$

3.1.3 Starting the algorithm

To obtain the starting set of parameters μ^0 and β^0 for the algorithm, we sample from an approximate model in two steps. First, we obtain μ^0 as the mode of the posterior described in equation (1) under the assumption that $h_{i,t}$ are independent across both dimensions.[4] Second, we use the same model to simulate $h_{i,t}$ from the posterior probability $p(h_{i,t}|\mu, \mathbf{x}_{i,t})$. Given a sample of health groups, we get the mode of the posterior of β, β^0, under the assumption that groups follow the same multinomial logit specification as in the baseline model.

3.1.4 Selecting the number of groups

Two main aspects play a role when selecting the number of groups. On the one hand, more groups might lead to a better model fit. Indeed, Bayes odd ratios indicate that more groups is the best choice. On the other hand, each additional group creates an extra burden for structural models, which rarely use more than four. Therefore, the optimal number of groups depends on the application. We focus on four groups as they can be implemented in many models, and provide a significant improvement against two and three groups. Nonetheless, we also provide results for the two-groups case for more complex structural models. Considering more than four groups would improve, mildly, some results; hence, the paper provides a conservative lower bound of our model performance.

4 Results

In this section we first describe the estimated health groups, then we explain how health evolves as individuals age taking into account differences in education and gender. Finally, we show how we can use our econometric model to produce a new health classification.

In what follows, we report the median of the posterior distribution of the parameters -or relevant functions of them.

4.1 Health Groups

Figure 1 reports the probability of reporting difficulties with each I-ADL conditional on being in each cluster, that is $\mu_{k,g}$ in equation (1). Each panel corresponds to a different number of clusters H. Meanwhile, each marker symbol represents a cluster and each tick in the horizontal axis refers to an ADL (the first six) or an IADL (the remaining ones). The

[4] This model is also known as latent class analysis (Lazarsfeld, 1950; McLachlan and Peel, 2004).

higher the marker is, the more likely is that an individual in that specific group struggles with the corresponding I-ADL.

[Figure 1 about here.]

If we set $H = 2$, the algorithm divides individuals into one group whose probability of declaring problems with an I-ADL is close to 0 for every I-ADL and another one which owns a higher likelihood of facing problems with every I-ADLs. We label the former group as *healthy* (circumferences) and the latter as *impaired* (triangles).[5] We also find large differences in the probabilities across I-ADLs within the *impaired* group which suggests that activities differ in their importance for categorizing individuals. For example, in the *impaired* group, these probabilities range from 31% in the case of eating to 77% in the case of shopping.

The upper right panel of Figure 1 presents the same graph but with $H = 3$. There is still one group with almost zero probability to face difficulties with any I-ADL and another with again the highest probabilities of struggling with all I-ADL. Nevertheless, the probabilities of this group are slightly higher than when we consider only two groups as some individuals previously classified as *impaired* belong to the new group whose probabilities lie between the other two.

When we allow for four groups, the *impaired* and the *healthy* groups become more distant. In addition, the middle group splits into two very different ones. One group with moderate probabilities to suffer difficulties with an ADL but low probabilities to have problems with IADLs, reflecting that those individuals are *physically frail*; and another one which consists of *mentally frail* elderly in the sense that they are mostly dependent in terms of IADLs but not as much in terms of ADLs.

Lastly, we consider $H = 5$ in the lower right panel. In that case, the previous groups remain almost unchanged and the new group that emerges is extremely similar to the *healthy* one, with the exception that individuals struggle reading a map. As one adds more groups, their connection to health is even weaker; therefore, in the remaining of the paper, we focus on the case of four groups.

While Figure 1 characterizes individual's health in each cluster, it is silent about the meaningfulness of each I-ADLs for classifying individuals. For instance, in the case of $H = 2$,

[5] As in most latent variable models, the labeling of the groups is somewhat arbitrary. Nonetheless, it eases the presentation without altering the main conclusions.

the elderly in the *impaired* group present a much higher probability of facing difficulties reading a map than eating. This comparison, however, disregards that unconditionally only 5% of individuals struggle to eat but 16% are not able to read a map.

To overcome this issue, Figure 2 plots the probability of belonging to group g given that the individual faces difficulties with I-ADL k, that is,

$$\Pr(h = g|x_k = 1) = \Pr(x_k = 1|h = g)\frac{\Pr(h = g)}{\Pr(x_k = 1)};$$

where the relative size of the bars indicates which I-ADL is more informative.

[Figure 2 about here.]

Following the same example, if a person has difficulties to eat, she belongs to the *impaired* group with probability 90%, according to the upper left panel. Meanwhile, individuals incapable of reading a map have almost the same likelihood to be part of the *impaired* or *healthy* group; thus, MAP is uninformative. The pattern of these two I-ADLs remains unchanged when $H = 3$ and $H = 4$; MAP is never informative while EAT is the best indicator to classify individuals into the *impaired* group. This evidence is in line with previous evidence in the medical literature (see Morris et al., 2013, and references therein) which argues that difficulties with eating are the best predictor of full dependence.

Figure 2 characterizes the importance of each I-ADL separately for descriptive purposes; however, the joint structure of these variables also contributes significantly to identification. To see this, in the third and fourth columns in Table 2 we provide the proportion of respondents who report difficulties with at least one ADL or IADL. Consistent with the previous discussion, individuals in the *impaired* group are the ones more likely to present difficulties with an I-ADL; actually, they face problems with one I-ADL almost surely. On the other side of the coin lies the individual of the *healthy* group whose probability of reporting troubles with ADLs varies between 1% and 9% depending on the number of groups. In the third panel (four groups), the distinction between *physically frail* and the *mentally frail* becomes salient. While in the former 87% of respondents struggle with ADLs, the latter faces more problems with IADLs (99.4%) and less with ADLs (51.3%).

[Table 2 about here.]

Groups are not only different in terms of I-ADLs but also in terms of demographics. For instance, if our classification correctly identifies the health status of individuals we expect members of the *impaired* group to be older than those of the other groups. In that regard, Table 2 shows they are indeed on average nine years older than the ones in the *healthy* cluster and six years older than those *physically frail*. Additionally, the difference between *mentally frail* and *impaired* is smaller which is consistent with mental conditions caused by aging. In terms of education, high school graduates are overrepresented in the *healthy* group in line with previous literature on health inequality such as Mackenbach et al. (2008). Another interesting pattern is that worse health groups contain a significantly higher proportion of women. These differences lead us to study pattern of heterogeneity of health dynamics across gender and education groups.

4.2 Heterogeneous health dynamics

The distribution of elderly into health groups changes with age, gender and education. Figure 3 plots the probability of being in each group through age. The left panels correspond to dropouts whereas the right ones present the results for high-school graduates; meanwhile, the upper graphs refer to males and the lower ones to females. The most common health status at early ages is *healthy* but starting around the age 90, *impaired* becomes the predominant group. Further, the *physically* and *mentally frail* have very different dynamics. The former is stable throughout life while the latter increases steeply as elderly age. These patterns are very similar across education and gender, although the initial composition of individuals varies with demographic characteristics.

[Figure 3 about here.]

Different transitions translate into different risks, which we summarize as expected time an individual at age 60 lives in each health group in Table 3. Even if the more educated elderly live longer, they spend fewer years as *impaired* and *frail*, which suggests that richer individuals face lower health risks. For instance, in the case of males, dropouts stay 80% more time (or 0.4 extra years) in the *impaired* state. The difference between males and females indicates that females expect a longer life but during those extra years, they do not expect to be healthy.

[Table 3 about here.]

4.3 Health Classification

The estimated econometric model exploits past, present, and future information on I-ADLs and survival to construct a parsimonious model for health groups and health transitions. Based on the estimated parameters, we construct a health classification at the individual level that only contains past and present information on survival and I-ADLs.

To do so, we use the mean of the posterior of μ and β $(\bar{\mu}, \bar{\beta})$ to recursively update each individual's probability of belonging to each health group. At each age t, we obtain $p(h_{i,t}|X_i^{t-1}, \bar{\mu}, \bar{\beta})$ which represents the probability of belonging to each health group using I-ADLs information up to time $t-1$. Then, we update this probability with the I-ADL information at time t to obtain the probability of each state conditional on present and past information, $p(h_{i,t}|X_i^t; \bar{\mu}, \bar{\beta})$:

$$p(h_{i,t}^{(g)}|X_i^t, \bar{\mu}, \bar{\beta}) \propto p(\mathbf{x}_{i,t}|X_i^{t-1}, h_{i,t}^{(g)}, \bar{\mu}, \bar{\beta}) \cdot p(h_{i,t}^{(g)}|X_i^{t-1}, \bar{\mu}, \bar{\beta}) = p(X_{i,t}|h_{i,t}^{(g)}, \bar{\mu}) \cdot p(h_{i,t}^{(g)}|X^{t-1}, \bar{\mu}, \bar{\beta}),$$

where $p(X_t|h_{i,t}^{(g)}, \bar{\mu})$ is the likelihood of current I-ADLs conditional on being in state g given by (1). Then, we obtain the forecast of health groups given information up to t, $p(h_{i,t+1}^{(g)}|X^t, \bar{\mu}, \bar{\beta})$, using the transition probabilities $p(h_{t+1}|h_t, \bar{\beta})$:

$$p(h_{i,t+1}^{(g)}|X_i^t, \bar{\mu}, \bar{\beta}) = \sum_{h_{i,t}^{(c)} \in \mathcal{H}} p(h_{t+1}^{(g)}|h_t^{(c)}, \bar{\beta}) \cdot p(h_{i,t}^{(c)}|X_i^t, \bar{\mu}, \bar{\beta}) \qquad t \geq 0$$

The recursion starts with $p(h_{i,0}|\bar{\mu}, \bar{\beta})$. To obtain this probability, we first assume that, at the age of 60, individuals' probability of belonging to each health group equals the average probability at that age: $p(h_{i,60-a_0^i}|\bar{\mu}, \bar{\beta}) \equiv \frac{1}{N}\sum_{i=1}^{N} p(h_{i,60-a_0^i}|\mathbf{X}, a = 60, \bar{\mu}, \bar{\beta})$. Then, since most individuals are observed for the first time older than 60, we update this probability until $t = 0$ using the transitions:

$$p(h_{i,t+1}^{(g)}|\bar{\mu}, \bar{\beta}) = \sum_{h_{i,t}^{(c)} \in \mathcal{H}} p(h_{i,t+1}^{(g)}|h_{i,t}^{(c)}, \bar{\beta}) \cdot p(h_{i,t}^{(c)}|\bar{\mu}, \bar{\beta}) \qquad 60 - a_0^i \leq t < 0$$

Intuitively, to ensure that we do not use information outside the information set of each individual, we assume they are homogeneous at the age of 60 and transition across states as the average individual until we observe their health data.

After updating until the last period, we obtain $p(h_{i,t}|X_i^t, \bar{\mu}, \bar{\beta})$ per individual and time period. The inverse of this probability for health group g acts as the probability weighting of that observation when we estimate moments conditional on being in group g. However, in our context, the average probability of the most likely group is 0.98 ($H = 2$), 0.95 ($H = 3$), and 0.94

($H = 4$); hence, these weights play a minor role. Since alternative health classifications do not provide weights, in the rest of the text, we disregard the weights and allocate each observation to its most likely group to make the comparison more transparent.

5 Comparison with alternative indices

The need for a discrete measure of health has led researchers to use ad-hoc classifications. In this section, we compare our proposed classification with the main three alternatives: self-reported health, if the individual struggles with an ADL, and the quintiles of a frailty index. In addition, we also propose an alternative classifications: the Cartesian product of whether the individuals report difficulty with i) at least one ADL and ii) IADLs (excluding MAP) as an unsophisticated proxy of our classification. This classification does not require any estimation and can be implemented with many health surveys (e.g. SHARE or ELSA); nonetheless, its similarity to our measure suggests that it might perform better than current classifications.

To perform the comparison, we focus on mortality and three variables related to the financial risk due to health: OOP medical expenditures, and indicators of receiving home-care and residing in a nursing home. OOP medical spending is a direct measure of the economic consequences of health. It includes the costs -in constant 2000 US dollars- of hospital and nursing home stays, doctor visits, dental treatments, outpatient surgery, prescription drugs, home health care, and special facilities. Received home care equals 1 if a medically-trained person has come to the respondent's home to help her, and nursing home resident takes value 1 for those individuals who live in a nursing home at the time of the interview.

[Table 4 about here.]

The health classification most widely used in the literature relies on an individuals self-assessment on their health status which can take 5 different values between *excellent* and *poor*. The self-reporting nature of the answer induces two opposing effects. On the one hand, individuals might know more about their health than researchers can ever measure. On the other hand, respondents might misjudge their health condition, incorporate other information as mood or consider different benchmarks of being *good*. Previous literature has analyzed the net effect of these two channels and establishes that the disadvantages often offset any benefit (see Currie and Madrian, 1999, for a survey). Nonetheless, the first panel of Table 4 confirms that self-reported health has information about the financial risks. Those respondents reporting worse health spend more on medical consumption and care, and are more likely to reside in a nursing home than those who claim to be healthy. The difference between the five groups varies though. In particular, answering *excellent*, *very good*, and *good* relates to almost the same risk, whilst *fair* and *poor* correspond to much more spending. Previous literature, thus, merges the three healthiest and the two worst groups. We denote this latter classification as *self-reported health (2 groups)*.

Grouping individuals according to if they have an ADL or not is similar to our approach, specifically to identify the *healthy* respondents; hence the proportions of *healthy* and No-ADL almost coincide. This classification, however, considers every ADL equally important and disregards the number of ADLs, as well as difficulties with IADLs.

Braun et al. (2017) construct a frailty index based on Searle et al. (2008) by merging information on I-ADLs, chronic conditions, cognitive impairment, and information about smoking and alcohol consumption to create a frailty index. Although the inclusion of more information improves the measure of health and allows to create more groups, the relevance of each variable is still assumed to be the same. Additionally, the resulting index is continuous which forces them to allocate individuals into five equally sized groups according to the quintiles of the index. As a result, the healthiest groups are very similar among themselves and the worst group present the same features as those who have an ADL in the Yes/No classification.

Finally, classifying individuals regarding whether they struggle with at least one ADL, IADL, both or none, which we denote as 4-I-ADL, can be understood as a simple approximation to our four groups. In contrast to the frailty index by Braun et al. (2017) who effectively separates individuals without problems with any ADL in four groups, this method divides respondents who recognize problems to perform an ADL into three groups. Since these individuals are more heterogeneous, the resulting groups become more differentiated in all the variables considered.

Even if the four aforementioned alternative classifications are highly correlated with the health outcomes that we use, our estimated groups seem to be more differentiated across them. For instance, using our methodology, the average difference in terms of OOP between *healthy* and *impaired* elderly is \$9,079. According to self-reported health, however, an individual belonging to the worst group only expends \$3,333 more than one in the best group. Similarly, the fact that you report an ADL implies that your average OOP medical spending is \$2,648 higher; meanwhile, being a part of the worst, rather than the best, frailty quintile costs \$3,305. Not surprisingly, 4-I-ADL is the closest to our classification but the distance between the best and worst groups hardly surpasses \$4,000. As for the intermediate groups, they are again less distinct in the case of the alternative classifications as their increment in spending is below \$1,000, the minimum difference between our groups, except from the two worst groups.

Regarding the probability of residing in a nursing home, a similar pattern arises and the difference between the best and worst of our health groups at least duplicates the same difference using the alternative methods. The same holds true for home care when we look at self-reported health or struggling with at least one ADL but, in this case, our four groups outranks 4-I-ADL just mildly.

In line with the previous discussion, our classification also identifies future death events more accurately. In particular, an *impaired* individual dies with 33% probability whereas only 3.6% of *healthy* ones do not survive to the next wave. Instead, the difference between the healthiest and unhealthiest groups does not reach 25 percentage points with alternative classifications.

5.1 A horse race

Most of the time, the researcher's concern might not be to classify individuals into distant groups but to create a categorical index that captures most of the variation coming from health conditional on age, gender, and education. To assess the performance of the grouping methods in that context, Table 5 displays the R^2 of the following regression:

$$y_{i,t} = c + \mathbf{d}'_{i,t}\beta + \mathbf{z}'_{i,t}\gamma + (\mathbf{d} \otimes \mathbf{z})'\,\theta + age_{i,t}\left(\mathbf{d}'_{i,t}\beta_1 + \mathbf{z}'_{i,t}\gamma_1\right) + \varepsilon_{i,t}$$

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}_{i,t}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs. As a comparison we also add the R^2 of the regression including, in the $\mathbf{d}_{i,t}$ vector, dummies for each I-ADL.

[Table 5 about here.]

Even though self-reported health only explains 1.9% of the variation of out-of-pocket medical spending, it doubles the variance explained solely by age, education, gender and their interactions. Similarly, we can explain up to 2.2% by dividing individuals according to whether they report problems with at least an ADL. If we also include IADLs, the fit improves by 1 percentage point. Altogether, our classification explains 3.8% of the medical spending variance which exceeds every alternative. The same conclusion arises from considering the spending reported in the following wave. Interestingly, our classification explains just 0.2% less than including all the ADLs and IADLs.

Nursing home residency, by virtue of being binary, contains a lower measurement error; nevertheless, the same ranking persists. Any measure that includes ADLs beats self-reported health by at least five percentage points, which doubles if we consider our unsophisticated method. Further, weighting each I-ADL, our health groups enhance the naive 4-I-ADL by 62% because it identifies the extreme dependent individuals better. Likewise, the naive classification performs poorly compared with including each I-ADL separately because the importance of each I-ADL is key to predict nursing home residency. Our proposed classification explains almost 4 times more variance than self-reported health and 2.5 times more than *ADL:Yes/No*.

In contrast to nursing home residents, most elderly who need home care preserve a high degree of independence. As a consequence, the weighting of I-ADLs loses importance and our measure, although remains to be the optimal, barely improves classifications based on ADLs. Nonetheless, it explains 50% more variance than self-reported health and reaches the variance explained by including all the I-ADLs.

Regarding mortality, we have constructed a division that performs better than self-reported health. This contribution is relevant because most of the literature (see Idler and Benyamini, 1997, for a survey) shows that subjective measures of health usually predict mortality beyond objective indicators. Notably, the R^2 using 4-I-ADL exceeds by 0.7 percentage points that of self-reported health which indicates that part of the improvement on the mortality prediction relies on the incorporation of I-ADLs, while the remaining improvement is due to the dynamics and the weighting of each I-ADL.

5.2 Dynamics: self-reported health versus endogenous classification

The comparison regarding groups' dynamics generates new insights about the differences between grouping methods. To obtain smooth dynamics, we assume that the transition probabilities of self-reported health follow a logistic specification as described by Equation (2). Furthermore, to ease the comparison we focus on the best and worse groups of each method, that is, we compare *healthy* according to our method with *excellent* as reported by individuals and *impaired* with *poor*. For completeness, we also include the two-group classification based on self-reported health in the comparison.

The upper plots in Figure 4 report the median probability of dying. The left one corresponds to the healthiest groups, whereas the right one presents the results for the most unhealthy ones. Up to the age of 80, individuals who report an *excellent* health, as well as those classified as *healthy* own very small probabilities of dying. After this age, elderly with a low survival probability still assess their health as *excellent*. On the other hand, age is not as important for the *healthy* group as mortality less than doubles between age 80 and 98. One possible explanation is that individuals compare themselves with relatives and friends of the same age to assess their health status; thus, respondents of age 65 and 90 have a different benchmark. Furthermore, while the difference between the mortality rates of *healthy* and *impaired* are sizable, this is not the case for the groups based on self-reported health, which suggests that this method does not predict mortality at older ages. In addition, *impaired* individuals feature a higher death probability than those who assess themselves as in *poor* health at any age.

[Figure 4 about here.]

The second relevant element of health risk is persistence. If the process is not persistent, health today would contain relatively little information on tomorrows health and survival probabilities thus affecting individuals' saving behavior. Additionally, the persistence of each classification sheds some light on the type of health process. In particular, we aim to create an indicator of LTC needs which is by definition persistent in contrast to others such as the flu or a sprained ankle. The lower plots in Figure 4 depict the probability of remaining in the same group conditional on the group you are at a given age. We find that individuals that report excellent in one wave have less than 40% of probability to provide the same answer in the following wave, whereas respondents classified as *healthy* are extremely likely to remain in that state. This fact indicates that some non-persistent factors might drive self-reported health. If we focus on individuals in bad health, our classification displays a larger persistence as individuals age which is in line with the idea that as you become older the harder it is to recover. In contrast, for fair and/or poor self-reported health, individuals are more likely to report improvements in their health status as people age which points towards changes in their health benchmark.

Lower persistence and a worse ability to predict mortality indicate that self-reported health overestimates the uncertainty faced by individuals. The effect of this bias on individuals' decisions depends on its severity across socio-economic groups and the specific structural model. To shed some light on the former, Figure 5 plots the additional percentage of time than a high-school graduate spends in the healthiest state (left-hand panel) and the unhealthiest state (right-hand side) in expectation. While our classification indicates that high school graduates spend around 40% more time in the *healthy* state and 30% less in the *impaired* state, using self-reported, these differences at least double. Given that our classification was able to explain a larger fraction of the variance of different health outcomes, these results suggest that self-reported health contains a measurement error correlated with education. More precisely, low educated individuals tend to report worse health status or high-school graduates overestimate their wellness, or both.

[Figure 5 about here.]

5.3 The asset cost of bad health

In order to show how to implement our health classification in a structural model and its implications, we replicate De Nardi et al. (2010) and use the model to quantify the asset cost of being in bad health across classifications. For this purpose, we solve and simulate the baseline model of De Nardi et al. (2010) using two levels of self-reported health (as the original paper) and three different versions of our proposed classification. First, we consider our two-group classification to compare the effects of our method without increasing the state space. Then, we consider the same two groups but we define the health state based on the probability of being *impaired*, using four equally space groups. This measure assesses the empirical relevance of probabilities and sets up a benchmark for a four-group classification. Finally, we implement our four-group classification.

Besides the health classification, we follow as much as possible the original paper to maintain the comparison. De Nardi et al. (2010) estimate the model on a sample of single retirees using a two-step procedure. In the first step, the authors estimate the process for survival, health transitions, and medical expenses risk. All these sources of uncertainty are functions of health; hence we reestimate them for every classification. In the second step, they estimate preference parameters and the generosity of Medicaid using the method of simulated moments. To ease the comparison, and due to the good aggregate fit, we maintain the same parameters as in the original model. Therefore, differences in policy functions come from differences in the health classification and not preference parameters. [6].

To analyze how health affects dissavings, we simulate 500,000 female individuals in the top quintile of the income distribution who are endowed with $175,000 (the median assets of the top income quintile in the HRS for individuals ages between 70 and 75). We fix half of the simulated sample to the good health state forever while the other half stays always in bad health ruling-out attrition in order to ease the interpretation of the results. We focus on rich female individuals because they face larger risk of outliving their assets.

[Figure 6 about here.]

The upper left panel in figure 6 shows the median asset holdings for individuals in each of the two subsamples when using self-reported health as in De Nardi et al. (2010). We see that after fifteen years in bad health a representative individual holds around 15% less wealth than an individual who remains healthy. Under any of our proposed classification the difference in the dissaving profile across health groups is much more salient. When using our classification with two and four health groups, the representative individual who remains impaired for fifteen years holds 30% and 64% lower wealth, respectively. Differences between using probabilities or the modal health state are relatively small because most individuals in the sample are classified almost surely in each group.

Differences in the asset cost of poor health become important when analyzing the optimal design of insurance products (Koijen et al., 2016), reverse mortgages (Nakajima and Telyukova, 2017), or the saving impact of LTC expenses in retirement (Ameriks et al., 2020). Importantly, in contrast to self-reported health, since we use an objective measure of health, these contracts can be conditional on the heath status.

6 Conclusion

As retirees age, they face large risks of requiring persistent and expensive care. The macroeconomic literature underlines the importance of this uncertainty to explain the dissaving pattern of the elderly and the labor supply decisions of the individuals close to retirement.

[6]The Online Appendix gathers the parameters of the first-step estimation and results of the original paper to assess the accuracy of the replication.

They face, however, an important empirical challenge: summarizing the information content of several health variables into a few groups, which is a requirement for quantitative models to be computationally feasible.

This paper develops a methodology to classify individuals into a reduced number of categories, exploiting the richness of the health information available in panel surveys. In addition, by profiting from the panel dimension of the data we estimate transitions across groups conditioning on current health, age, education, and gender, which are of paramount importance when calibrating macroeconomic models.

Individuals LTC needs can be parsimoniously represented with four different groups, namely, *healthy*, *impaired*, *physically* and *mentally frail*. While *healthy* and *impaired* have the usual extreme interpretation, the distinction between *physically* and *mentally frail* arises from the different pattern of respondents struggling with ADLs and IADLs. Moreover, and in line with the previous literature, health status is highly persistent over time, but with significant differences in the dynamics of health across demographic groups.

We then assess our proposed classification against other commonly used measures. Our comparison exercises show that previous health indices are weakly related to health outcomes and medical utilization rates. In contrast, our health groups explain a significant fraction of the variance in the use of nursing homes, home health care, out of pocket medical expenses, and mortality. Moreover, we show that using a more accurate health measure changes significantly the saving patterns predicted by life-cycle models. In particular, self-reported health predicts a slower dissaving of individuals in bad health compared to our classification.

Finally, we make available the health classification for two, three, and four groups for future research exploiting the main retirement surveys: HRS, ELSA, and SHARE. Its discreteness and good fit makes the classification valid for most applications; hence, it constitute a good candidate as a unified health measure.

References

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2016). The long-term-care insurance puzzle: Modeling and measurement. Technical report, National Bureau of Economic Research.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, 128(6):2375–2451.

Barczyk, D. and Kredler, M. (2018). Evaluating long-term care policy options, taking the family seriously. *The Review of Economic Studies*, 85(2):766–809.

Bohacek, R., Crespo, L., Mira, P., and Pijoan-Mas, J. (2015). The educational gradient in life expectancy in europe: Preliminary evidence from share. *Ageing in Europe-Supporting Policies for an Inclusive Society*, page 321.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bound, J., Stinebrickner, T., and Waidmann, T. (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics*, 156(1):106–129.

Braun, R. A., Kopecky, K. A., and Koreshkova, T. (2017). Old, frail, and uninsured: Accounting for puzzles in the us long-term care insurance market. Technical report, FRB Atlanta Working Paper No. 2017-3.

Brown, J. (2002). Differential mortality and the value of individual account retirement annuities. In *The distributional aspects of social security and social security reform*, pages 401–446. University of Chicago Press.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the British household panel survey. *Journal of Applied Econometrics*, 19(4):473–503.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

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

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

De Nardi, M., French, E., and Jones, J. B. (2016). Medicaid insurance in old age. *The American Economic Review*, 106(11):3480–3520.

Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics*, 12(3):313–336.

Dobrescu, L. I. (2015). To love or to pay savings and health care in older age. *Journal of Human Resources*, 50(1):254–299.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

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, A. M. (2000). Health econometrics. *Handbook of Health Economics*, 1:265 – 344.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of illness in the aged: the index of adl: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185(12):914–919.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Ko, A. (2016). An equilibrium analysis of the long-term care insurance market. *Draft, University of Pennsylvania*.

Koijen, R. S., Van Nieuwerburgh, S., and Yogo, M. (2016). Health and mortality delta: Assessing the welfare cost of household insurance choice. *The Journal of Finance*, 71(2):957–1010.

Kopecky, K. A. and Koreshkova, T. (2014). The impact of medical and nursing home expenses on savings. *American Economic Journal: Macroeconomics*, 6(3):29–72.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, pages 362–412.

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., and Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 european countries. *New England Journal of Medicine*, 358(23):2468–2481.

Manton, K. G. and Soldo, B. J. (1985). Dynamics of health changes in the oldest old: New perspectives and evidence. *The Milbank Memorial Fund Quarterly. Health and Society*, pages 206–285.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

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.

Morris, J. N., Berg, K., Fries, B. E., Steel, K., and Howard, E. P. (2013). Scaling functional status within the interrai suite of assessment instruments. *BMC Geriatrics*, 13(1):128.

Nakajima, M. and Telyukova, I. A. (2017). Reverse mortgage loans: A quantitative analysis. *The Journal of Finance*, 72(2):911–950.

O'Donnell, O., Van Doorslaer, E., and Van Ourti, T. (2015). Health and inequality. In *Handbook of income distribution*, volume 2, pages 1419–1533. Elsevier.

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

Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M., and Rockwood, K. (2008). A standard procedure for creating a frailty index. *BMC Geriatrics*, 8(1):24.

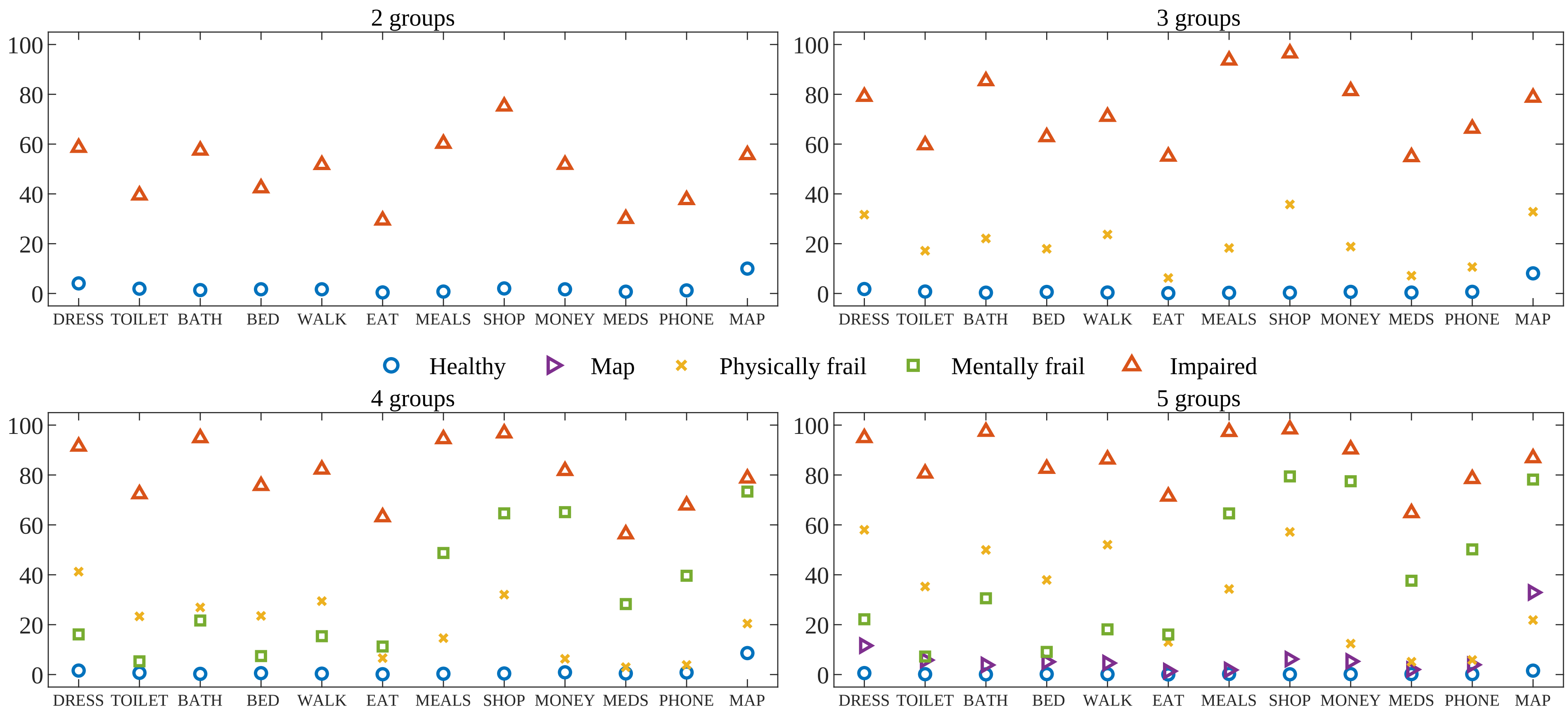
Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative Economics*, 11(1):231–275.

Wang, W., Phillips, P. C., and Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, 33(6):797–815.

Yang, Y. and Lee, L. C. (2009). Dynamics and heterogeneity in the process of human frailty and aging: evidence from the us older adult population. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, pages 246–255.

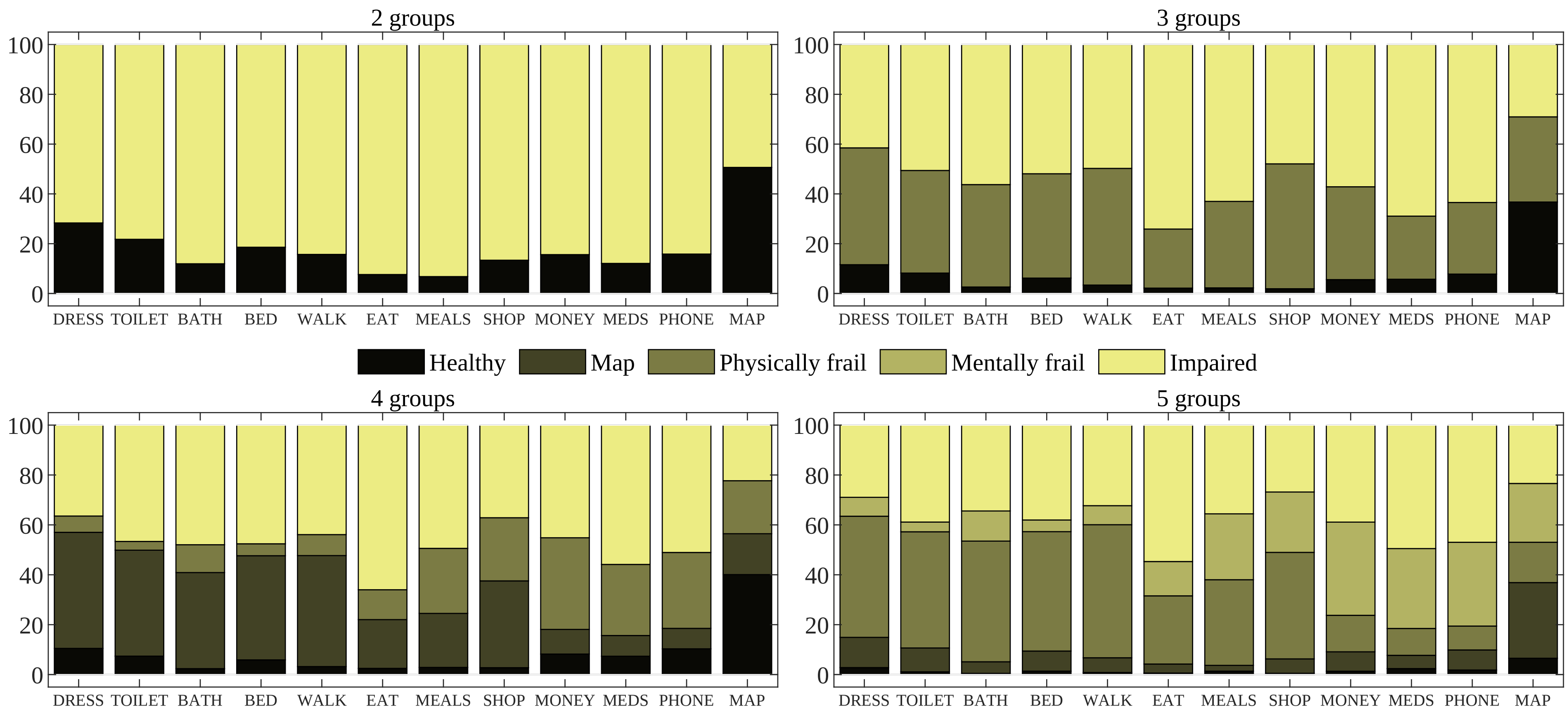
Figures

Figure 1: Probability of reporting a difficulty with a given I-ADL by health group



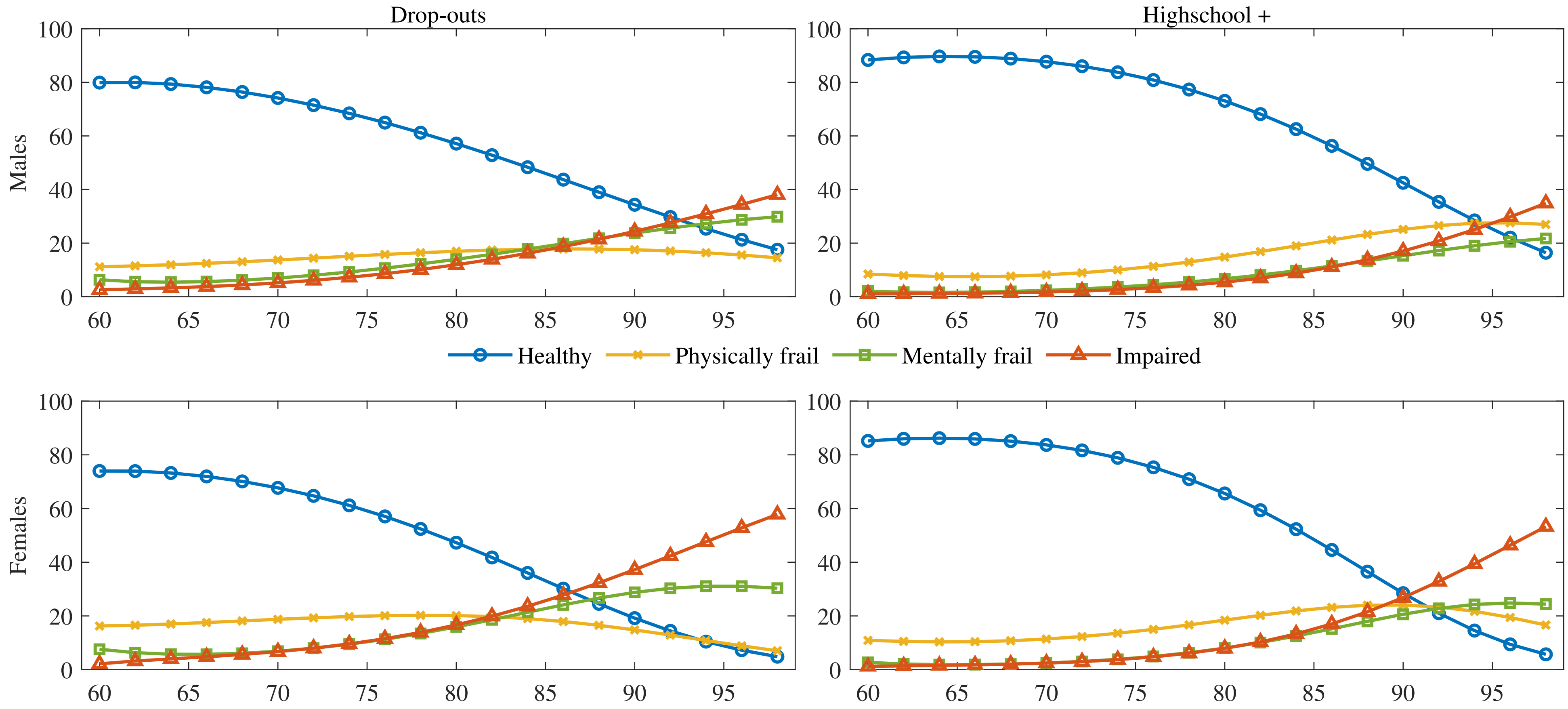
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications (MEDS), using a phone (PHONE), and using a map (MAP). The units of the y-axis are percentage points.

Figure 2: Probability of belonging to a given health group by I-ADL



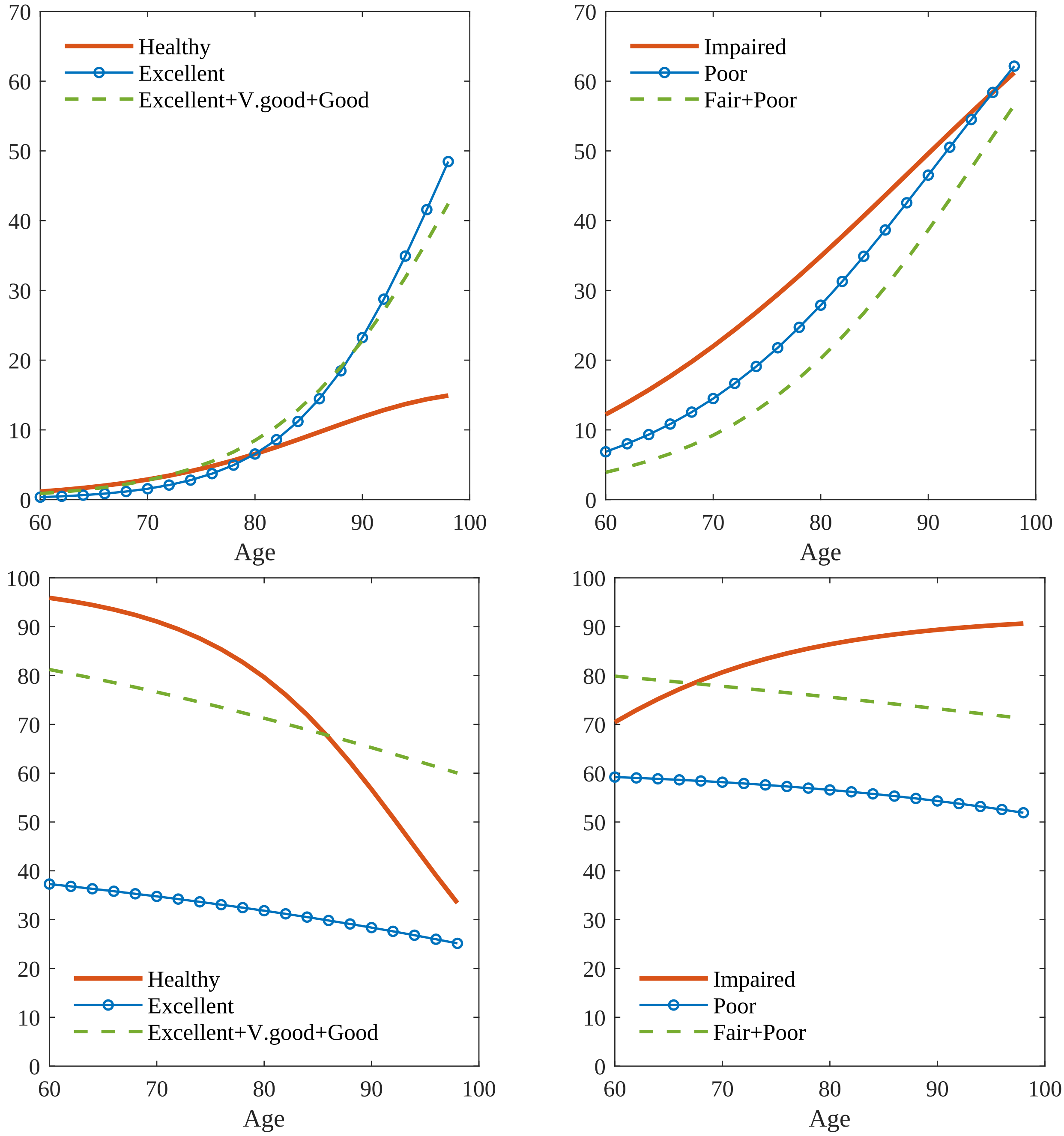
Notes: ADLs: Some difficulty with dressing (DRESS), using the toilet (TOILET), bathing (shower, BATH), getting in or out of bed (BED), to walk across a room (WALK) and eating (EAT). IADLs: Some difficulty with preparing hot meal (MEALS), shopping for groceries (SHOP), managing money (MONEY), taking medications (MEDS), using a phone (PHONE), and using a map (MAP). The units of the y-axis are percentage points.

Figure 3: Share of individuals in each group conditional on being alive by education and gender as individuals age



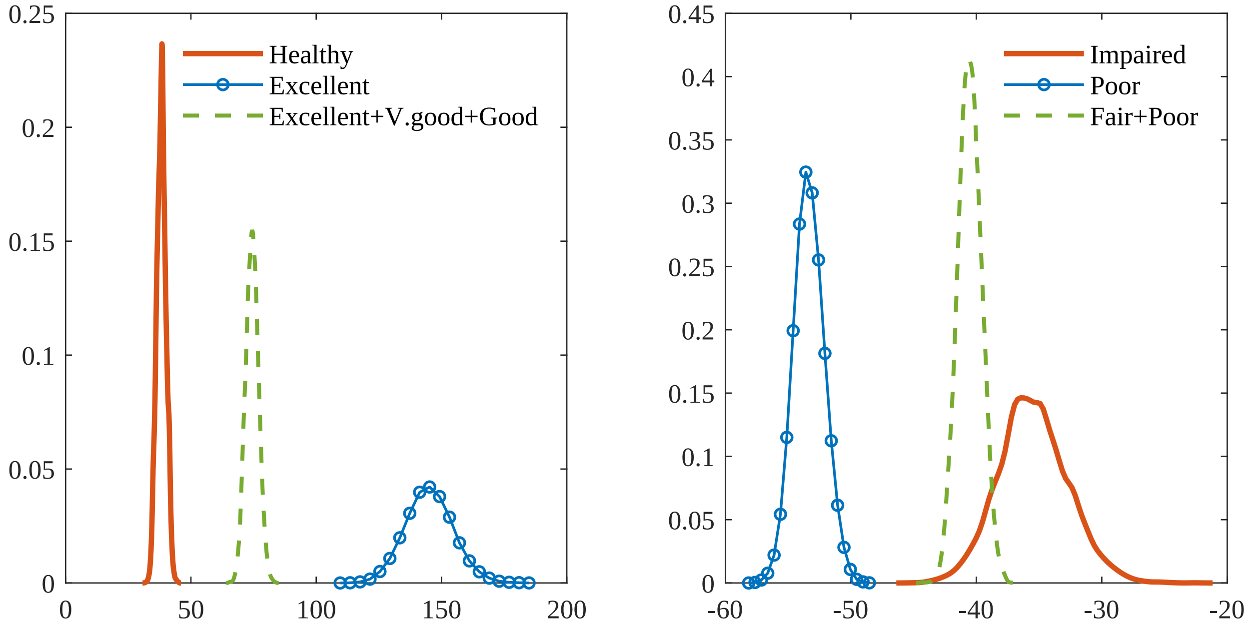
Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. See Section 3 for details about the econometric model and the estimation procedure. The units of the y-axis are percentage points and those of the x-axis are years.

Figure 4: Transition to death and persistence



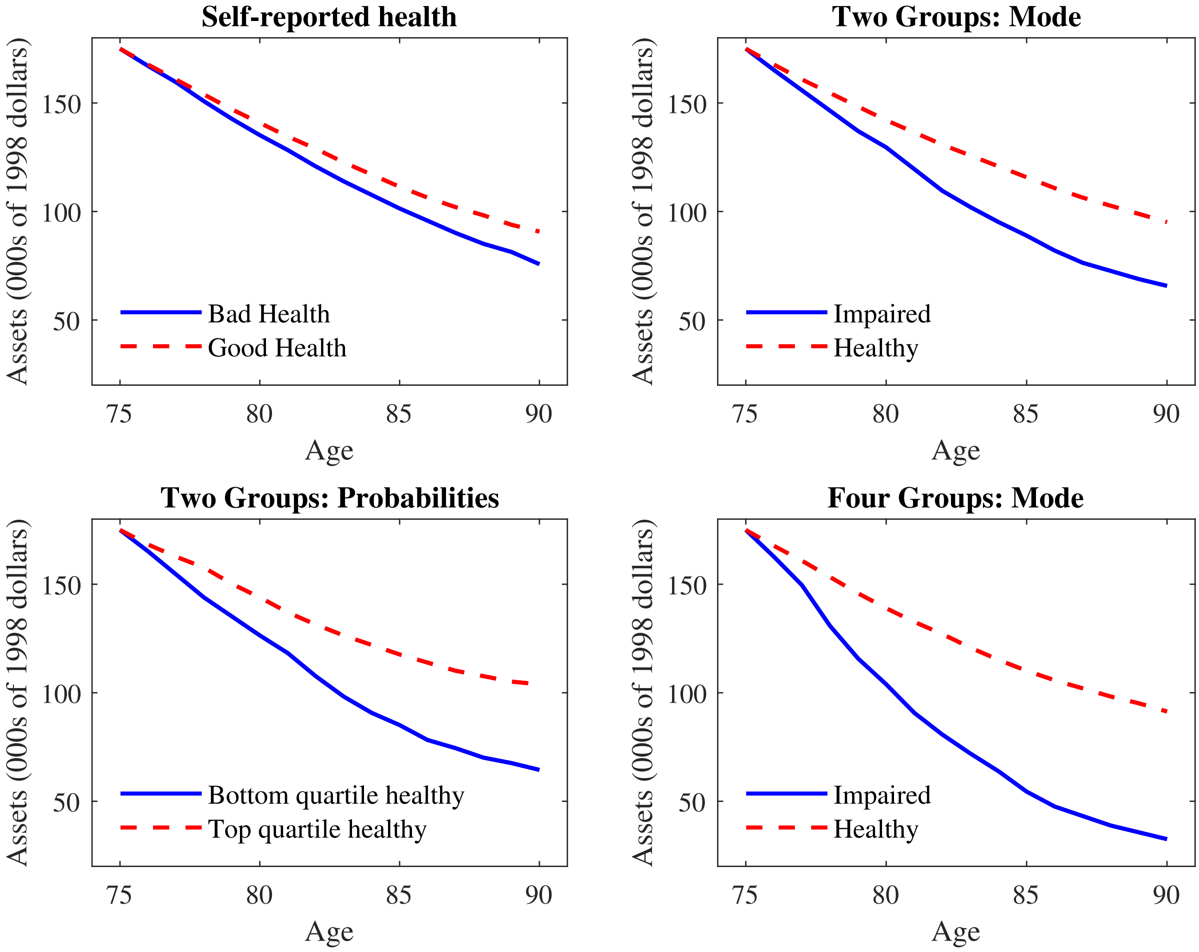
Notes: Upper plots: probability of dying per health group. Lower plots: Probability of maintaining the same health state. RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The units of the y-axis are percentage points and those of the x-axis are years. This graph corresponds to female dropouts but it is similar if we look at other socio-economic groups.

Figure 5: Expected educational gradient across health classification



Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. The units of the x-axis are percentage points.

Figure 6: Comparison of dissaving across health states



Notes: This figure represents the dissaving pattern in the model by De Nardi et al. (2010) of the median female rich individual who stays in the healthy state from the age 75 to age 90 (dashed) and the one who has bad health from age 75 to age 90. The upper left-hand plot uses self-reported as the measure of healthy/unhealthy; the upper right-hand plot uses our two-group classification; the lower left-hand plot uses discretize the prbability of being healthy in our two group classification into quartiles and uses the top/bottom quartile as the measure of healthy/unhealthy;the lower right-hand plot uses our four-group classification;

Tables

Table 1: Fraction of individuals reporting difficulties with I-ADLs by self-reported health

Variable	Definition	# Obs	All	Self-reported health Exc.	Very	Good	Fair	Poor
	Activities of daily living (ADLs): Some difficulty...							
DRESS	Dressing	134,980	12.4	2.2	3.5	8.1	20.2	44.1
TOILET	Using the toilet	134,785	7.6	1.0	2.1	4.8	12.1	29.2
BATH	Bathing (shower)	134,949	10.0	1.6	2.3	5.7	16.0	40.3
BED	Getting in or out of bed	134,900	7.9	1.0	1.4	4.3	13.0	33.2
WALK	To walk across a room	134,913	9.4	1.1	1.9	5.2	14.8	39.4
EAT	Eating	134,908	4.9	0.8	1.0	2.5	7.4	21.5
	Instrumental activities of daily living (IADLs): Some difficulty...							
MEALS	Preparing hot meal	127,840	9.6	1.8	2.4	5.6	14.7	39.3
SHOP	Shopping for groceries	130,313	12.8	2.2	3.1	7.7	21.0	50.2
MONEY	Managing money	130,013	9.2	2.5	3.1	6.2	14.1	32.2
MEDS	Taking medications	131,264	5.3	1.2	1.5	3.1	7.9	20.4
PHONE	Using a phone	134,259	6.8	1.6	2.2	4.4	10.2	24.7
MAP	Using a map	117,200	15.7	6.5	8.7	13.6	23.8	39.3
	Some difficulties with...							
ADL	At least one ADL	134,366	21.1	4.0	6.9	15.6	35.6	66.0
IADL	At least one IADL	103,910	23.2	10.8	14.2	24.5	47.0	74.3
I-ADL	At least one I-ADL	103,663	29.6	10.8	16.1	28.2	51.3	78.5

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (< 0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed on average 6 waves (12 years). The column *All* indicate the percentage of observations who have problems with a given I-ADL. The last five columns present the same percentage by group of self-reported health (excellent (Exc.), very good (Very), good, fair and poor).

Table 2: Summary statistics for estimated health clusters

Group	Share	ADL	IADL	Age	Female	Dropout
		Average				
	100	21.3	33.2	72.7	57.4	27.2
		2 groups				
Healthy	85.6	9.7	21.7	71.1	56.0	23.0
Impaired	14.4	89.0	99.9	78.0	68.0	46.4
		3 groups				
Healthy	77.2	4.2	13.7	70.7	55.1	21.4
Physically frail	16.5	70.9	97.0	75.7	65.6	41.3
Impaired	6.3	97.0	99.9	80.6	67.0	49.2
		4 groups				
Healthy	78.4	4.1	14.5	70.8	55.2	21.8
Physically frail	11.8	87.4	98.4	74.6	67.1	36.6
Mentally frail	4.9	51.3	99.4	79.5	62.9	52.7
Impaired	5.0	100.0	99.9	80.5	69.5	48.8
		5 groups				
Healthy	61.1	0.8	2.1	70.1	51.0	17.0
Map	24.3	31.8	70.4	73.8	68.7	39.0
Physically frail	6.8	97.0	99.9	74.6	68.7	38.4
Mentally frail	3.9	62.1	99.7	80.4	64.2	51.3
Impaired	3.8	100.0	99.9	81.3	69.0	49.2

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Results reported in percentage points. See Section 3 for details about the econometric model and the estimation procedure.

Table 3: Expected forthcoming time in each health group by education and gender at age 60.

Education	Healthy	+	Physically frail	+	Mentally frail	+	Impaired	=	Life Expectancy
				Females					
Dropouts	12.7		3.6		1.9		1.7		19.9
	(0.2)		(0.1)		(0.1)		(0.1)		(0.2)
Highschool	17.6		3.1		1.1		1.1		23.0
	(0.1)		(0.1)		(0.0)		(0.0)		(0.1)
				Males					
Dropouts	12.4		2.2		1.3		0.9		16.7
	(0.2)		(0.1)		(0.1)		(0.0)		(0.2)
Highschool	16.7		1.9		0.7		0.5		19.8
	(0.1)		(0.1)		(0.0)		(0.0)		(0.1)

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Results reported in years. In parentheses we report the standard deviation of the posterior distribution. See Section 3 for details about the econometric model and the estimation procedure.

Table 4: Long-term care needs by health classification

	Share	OOP med spending	Nurs-h resident	Received h-care	Dead next wave	ADL>0	IADL>0	IADL>0 w/o MAP
Self-reported health								
Excellent	9.2	1,805	0.6	2.5	2.4	3.5	9.4	4.5
Very good	28.0	2,129	0.7	3.9	3.2	6.4	13.1	6.2
Good	32.2	2,764	1.3	7.4	5.5	14.9	22.5	12.8
Fair	21.0	3,594	3.0	14.2	11.2	34.6	44.1	30.1
Poor	9.4	5,138	7.9	28.1	24.6	65.2	71.2	60.1
ADL: Yes/No								
No	79.6	2,357	0.3	5.3	4.7	0.0	16.7	7.8
Yes	20.4	5,005	8.8	25.7	19.0	100.0	69.5	59.2
Frailty Index Quintiles								
Lowest quintile	19.6	1,743	0.1	1.5	1.5	0.1	2.8	0.7
2	19.8	2,062	0.1	3.2	2.8	0.9	8.4	2.4
3	20.8	2,524	0.2	5.6	4.9	4.7	15.5	5.4
4	19.0	3,017	0.7	10.3	8.4	21.9	31.8	16.7
Highest quintile	20.8	5,048	8.9	25.9	20.6	72.5	77.1	64.6
4-I-ADL (i, j): ADL=i & IADL=j, IADL without MAP								
(0,0)	73.4	2,274	0.1	4.5	3.9	0.0	9.0	0.0
(1,0)	8.3	3,023	0.6	13.6	8.7	100.0	17.2	0.0
(0,1)	6.2	3,337	2.6	14.0	13.7	0.0	100.0	100.0
(1,1)	12.1	6,371	14.5	34.7	26.2	100.0	100.0	100.0
2 groups (mode)								
Healthy	86.1	3,122	0.2	5.7	4.1	9.5	14.6	6.6
Impaired	13.9	7,778	13.5	30.6	21.6	87.8	93.2	90.6
4 groups (mode)								
Healthy	77.6	3,330	0.1	4.8	3.6	3.9	11.0	3.3
Physically Frail	13.0	4,524	1.5	19.2	10.9	80.0	59.2	50.0
Mentally frail	5.1	5,780	7.3	22.7	18.0	51.3	99.0	96.8
Impaired	4.3	12,411	31.8	41.4	33.1	100.0	99.9	99.9

Notes: Results reported in percentage points, except for OOP med spending which is reported in 2000 US dollars. See Section 2 for details about the data and Section 3 for details about the econometric model and the estimation procedure.

Table 5: Fraction of explained variance by health classification

	OOP medical spending		Nursing home resident		Received home care		Mortality
Wave	Current	Next	Current	Next	Current	Next	Next
No health	0.7	0.8	4.3	5.1	3.6	3.5	5.9
SRH (2 groups)	1.5	1.3	6.0	6.2	7.3	6.2	9.3
SRH (5 groups)	1.9	1.5	7.1	6.8	9.0	7.3	11.2
ADL: Yes/No	2.2	1.7	11.4	9.8	9.9	7.5	9.8
Frailty index	2.7	2.3	12.6	11.5	11.6	9.6	12.1
All I-ADL dummies	4.0	2.7	31.5	20.2	13.1	8.7	13.1
4-I-ADL	3.2	2.5	16.2	13.8	12.3	9.1	11.9
2 groups (mode)	2.6	2.2	15.8	13.7	11.3	8.2	11.4
4 groups(mode)	3.8	2.5	26.3	18.2	12.8	9.1	12.8
Observations	118,706	94,544	118,706	94,544	117.408	93.268	102.292

Notes: RAND HRS Data; sample from 1996 to 2014 (10 waves). We select individuals over 60 years old and we drop individuals whose education, gender or age are missing (<0.1% of observations). The final sample consists of 159,025 interviews (including exit waves) which correspond to 27,369 individuals followed 6 waves (12 years) on average. Then, we restrict the sample to those observations that can be classified according to all criteria. Results reported in percentage points. Numbers correspond to the R^2 of the following regression:

$$y_{i,t} = c + \mathbf{d}'_{i,t}\beta + \mathbf{z}'_{i,t}\gamma + (\mathbf{d} \otimes \mathbf{z})'\,\theta + age_{i,t}\left(\mathbf{d}'_{i,t}\beta_1 + \mathbf{z}'_{i,t}\gamma_1\right) + \varepsilon_{i,t}$$

where $y_{i,t}$ is the variable used as a reference, $\mathbf{z}_{i,t}$ includes gender and education, and $\mathbf{d}_{i,t}$ is a vector of dummy variables indicating to which group the individual belongs.