Discussion Paper Series – CRC TR 224

Discussion Paper No. 660
Project C 01

Mens Sana in Corpore Sano?

Janoś Gabler[1]
Hans-Martin von Gaudecker[2]
Jürgen Maurer[3]
Mariam Petrosyan[4]

March 2025

[1]AppliedAI institute for europe
[2]University of Bonn and Institute for the Study of Labor (IZA), Email: hmgaudecker@uni-bonn.de
[3]HEC Lausanne, Email: jurgen.maurer@unil.ch
[4]Private Sector

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Mens Sana in Corpore Sano?*

Janoś Gabler[a]
Hans-Martin von Gaudecker[b, c]
Jürgen Maurer**[d]
Mariam Petrosyan[e]

[a] *appliedAI institute for europe*
[b] *Department of Economics, University of Bonn*
[c] *Institute for the Study of Labor (IZA)*
[d] *Department of Economics, HEC Lausanne*
[e] *Private Sector*

February 28, 2025

Almost by definition, maintaining high levels of physical and cognitive functioning into old age are key factors for successful ageing. Numerous studies have studied the factors influencing either physical health or cognition. However, the bio-medical literature shows both processes to be closely intertwined. On the one hand, high levels of cognitive functioning allows planning of health-related activities, gauging the consequences of actions, and adhering to medication plans. On the other hand, shocks to physical health have shown to predate declines in cognitive health. We take a broad and systematic approach to model the interdependency of physical health and cognitive functioning over the last third of the life-cycle. To do so, we adapt the model by Cunha, Heckman, and Schennach (2010), originally developed for the development of human capital during childhood and adolescence. In our case, the approach combines factor models for physical health, cognition, and investments into both of these with a nonlinear framework to describe their evolution over time. We use the HRS data from 2002–2016, which gives us individual trajectories of physical and cognitive capacity, and investments over ages 68 to 93 years. Our key results indicate 1) the measurement system and nonlinear dynamics are important modelling components, 2) the rank order of latent factors is remarkably stable, and 3) physical and cognitive capacity can be influenced by investments until very high ages.

*We would like to thank Johannes Ewald for preparatory work in his M.Sc. thesis at the University of Bonn (March 2020). The authors are grateful for support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1–390838866 – and through CRC-TR 224 (Project C01).
**Corresponding author (jurgen.maurer@unil.ch).

1 Introduction

Development and maintenance of human capital throughout the life-course enables individuals to lead longer, more productive and more satisfactory lives. The notion of human capital thereby comprises a broad range of abilities including skills, knowledge, health, and functioning, which, in turn, may influence each other and shape individuals' capabilities, behaviors and experiences throughout their life-courses (World Bank, 2018). While there is a large economic literature on early-life human capital development and its effects on adult outcomes (Heckman and Mosso, 2014), fewer studies in economics have analyzed the roles of context, individual investments and corresponding technologies for the maintenance and depreciation of human capital during later life within an integrated framework to model later-life human capital dynamics (McFadden, 2008). Reflecting economists' larger focus on early-life human capital accumulation relative to human capital maintenance and depreciation, economic studies on the later-life dynamic interplay of key forms of human capital and the role of investments in these processes remain relatively scarce to date.

Physical and cognitive capacity represent two key forms of human capital during adulthood and are perhaps the most important forms of human capital at older ages, especially after retirement. Physical and cognitive capacity are key determinants of many important outcomes in health economics and beyond such as mortality, healthcare use and healthcare cost and spending, falls and disability, long-term care needs and nursing home use, economic and social participation and subjective wellbeing to name but a few (Organization, 2015). As a result, investments in the maintenance of physical and cognitive capacity are key to ensuring a healthier, longer, and happier old-age that puts less strain on health and long-term care systems. Moreover, since many of these outcomes are highly uncertain, demand for various healthcare and long-term care related insurance products depends on the later-life dynamics of physical and cognitive capacity (Hosseini, Kopecky, and Zhao, 2022). Understanding the later-life dynamics of physical and cognitive capacity is, therefore, a key pre-requisite and input into models aimed at studying the role of later-life human capital on these important later-life outcomes and related investment and insurance decisions.

While physical and cognitive capacity tend to decline during later life (Niccoli and Partridge, 2012). there is considerable heterogeneity in the onset and speed of such aging-related declines across individuals, which is often related to individual differences in exposures and investments (Crimmins, 2020). What is more, several studies have actually shown significant improvements in later life physical and cognitive capacity following targeted investments such as physical exercise programs or cognitive trainings, suggesting that both physical and cognitive function remain malleable even at very high ages (Ball, Berch, Helmers, Jobe, Leveck, et al., 2002; Fiatarone, O'Neill, Ryan, Clements, Solares, et al., 1994). This evidence suggests that aging-related changes in function are not fully pre-determined biologically but can be postponed, slowed down, compensated and in certain instances perhaps even (temporarily) reversed or overcompensated through appropriate later-life investments. These findings highlight the important role of health investments for physical and cognitive capacity throughout the entire life course. This remains true even if early-life health investments into health to build up "reserves" for later life may be more efficient due to a higher degree of malleability early in life, the longer time horizon available to capitalize on early investments, and potentially important complementarities of health investments over time (Cunha, Heckman, and Schennach, 2010).

Besides documenting the continued malleability of physical and cognitive capacity during later life, the more recent literature in gerontological science has also found evidence potentially important cross-effects of physical function on cognitive function and vice versa. These cross-effects may go beyond the responses of physical and cognitive function due to common risk factors such as physical inactivity or diseases affecting both physical and cognitive capacities such as Parkinson's disease and represent more general connections between physical and cognitive capacity (Clouston, Brewster, Kuh, Richards, Cooper, et al., 2013). Evidence for such connections comes from both observational studies and RCTs, often but not always focused on the connection between cognitive and gait (dys-)function (Montero-Odasso, Verghese, Beauchet, and Hausdorff, 2012). In view of these findings, economic models of human capital maintenance and depreciation during later life should thus allow for flexible later-life dynamics of physical and cognitive capacities that can incorporate different forms of investment, and possible cross-effects between physical and cognitive capacities.

Varied existing conceptualizations of physical and cognitive capacity used in the literature and potentially widespread measurement error in physical and cognitive assessments in survey data and self-reported health investments further complicate the already complex task of capturing the joint dynamics of later-life physical and cognitive capacity and related investments (Baker, Stabile, and Deri, 2004; Bound, Brown, and Mathiowetz, 2001; Hosseini, Kopecky, and Zhao, 2022; Kapteyn, Banks, Hamer, Smith, Steptoe, et al., 2018). Physical capacity, for example, is a multifaceted concept that is generally assessed through multiple self-reported and/or performance-based survey items presenting noisy measurements for underlying true physical capacity (Kasper, Chan, and Freedman, 2017). Similarly, cognition comprises a range of different cognitive functions such as such as perception, attention, intelligence, knowledge, memory and working memory, judgement, reasoning, computation, problem solving or comprehension, whose corresponding measurements have signal value for overall cognitive capacity (Salthouse, 2010, 2012). Perhaps more surprisingly, even commonly used survey items for health investments such as self-reported physical activity contain substantial measurement error relative to actual health investments and, therefore, need to be treated with caution (Kapteyn et al., 2018)). Given the large potential for significant measurement error in survey-based assessments of physical and cognitive capacity and corresponding health investments documented in the literature, it seems prudent to employ an analytical framework that can readily accommodate such measurement errors when analyzing the joint dynamics of these outcomes.

The main objective of this paper is to estimate the technology for human capital maintenance and depreciation in later-life focusing on the dynamic interplay between later-life physical and cognitive capacity and corresponding investments among older adults in the US. To this end, we propose the use of a non-linear dynamic latent factor model as first proposed by Cunha, Heckman and Schennach (Cunha, Heckman, and Schennach, 2010) as a framework to model early-life human capital accumulation to study later-life human capital depreciation processes using longitudinal data from the US Health and Retirement Study (HRS). Applying this framework to investigate the joint dynamics of later-life physical and cognitive capacity and related investments is very attractive as such a non-linear dynamic latent factor model can incorporate the main aforementioned stylized facts about human capital depreciation, i.e., (1) allowing for a joint modelling of physical and cognitive capacity and investments that can incorporate potentially important cross-domain effects; (2) integrating the continued malleability of both physical and cognitive capacity into the model to study dynamically optimal investment paths and (3) accounting for error in the measurement of physical and cognitive function and corresponding investments in a context where there are several measurements of each of these domains in many commonly used data set, but each measurement is like to provide only a noisy signal for the underlying construct at hand. In addition to accommodating key stylized facts about human capital maintenance and depreciation into a unified framework, our models also allows us to identify the distribution of latent factors from noisy measurements, simulate the effects of different investment patterns on physical and cognitive capacity, calculate optimal investment patterns, notably the role of investments for human capital maintenance in younger old vs older old individuals, and anchor the results in interpretable metrics such as survival probabilities.

Our paper relates to two strands of research in economics, one methodological one on the use of non-linear dynamic latent factor models for estimating dynamic human capital production, which has-to the best of our knowledge-so far only been applied to the case of human capital accumulation in early life but not to human capital maintenance and depreciation in later life, and one more substantive one on the measurement and modelling of health dynamics during adulthood and later life. From a methodological point of view, our paper transfers widely used methods for the study of early-life human capital accumulation to the study of later-life dynamics of physical and cognitive function and eventual mortality. As a technical contribution, we show how to incorporate mortality into the framework and improve the numerical stability of a well known maximum likelihood estimator. By applying non-linear dynamic latent factor models to questions of aging and later life health dynamics, we show the usefulness of these methods to study human development not just in early life but across the entirely life-course, especially since many of the modelling and measurement issues mentioned above seem common to both ends of the life-course. As a result, we hope that our paper will aspire a larger group of life-course and aging researchers to consider such models in their research both in health economics and related fields.

Substantively, we contribute to the literature on how to measure and model later-life health dynamics in situations were we observe multiple potentially very noisy measurements for fewer latent concepts such as physical and cognitive capacity, which has long challenged empirical analyses in health economics and beyond. More specifically, one important issue in this literature is how to measure health in a comprehensive yet parsimonious way in view of the multifaceted nature of health on the one hand and the common need for dimensionality reduction in econometric models on the other. To address this trade-off, one set of commonly adopted approach to measuring health is to directly use (usually ordered measurements of) self-rated health as summary measure of health as outcome of interest (Contoyannis, Jones, and Rice, 2004; Heiss, 2011; Latham and Peek, 2012). This approach is generally motivated by a high predictive value of self-rated health for mortality (Idler and Benyamini, 1997). Alternatively to directly using self-reports to measure health, a commonly used alternative approach is to "instrument" health via a larger and "more objective" set of individual health measurements,such as information on specific health conditions, functional limitations, performance test results or anthropometric measures. This approach endogenously derives weights for aggregating the more detailed set of individual health measurements into a single health index that can then be used in further analysis (Cutler and Richardson, 1997; Jürges, 2007). Relative to using self-rated health directly as outcome, the approach aims to improve measurement by using "more objective" measures of health to construct an underlying health index, whereby the weights attributed to each detailed and "more objective" health measure in the final health index is determined by the partial association of the respective detailed health measure with self-rated health. While this approach can address some known issues with self-rated health, such as potential age-, sex- or SES-dependent reporting heterogeneity (Dowd and Zajacova, 2007, 2010; Lindeboom and Van Doorslaer, 2004), there is often still considerable measurement error in the "more objective" health measures that cannot be purged using this approach and may require further consideration (Baker, Stabile, and Deri, 2004; Maurer, Klein, and Vella, 2011). A second related approach side-steps the use of self-rated health entirely and instead uses principal component analysis of the more detailed health measurements to derive lower dimensional health indices (Jenicek, Cleroux, and Lamoureux, 1979; Nakazato, Sugiyama, Ohno, Shimoyama, Leung, et al., 2020; Poterba, Venti, and Wise, 2017). A third and increasingly popular approach simplifies the aggregation process for the more detailed health measurements even further by constructing a so-called "frailty index" or "deficit index", which simply consists of the total number of prevalent "health deficits" divided by the total number of potential "health deficits" (Hosseini, Kopecky, and Zhao, 2022; Rockwood and Mitnitski, 2007). A such constructed "frailty index"/"deficit index" is thus bounded to lie between zero and one and represents the percentage of potential "health deficits" already suffered by a given individual. A final set of studies refrains from performing some form of dimensionality reduction and uses the more detailed health measures directly in their analyses, either in isolation or simultaneously. As this is, for example,, the standard approach of disease-based analyses, most published papers on health adopt this latter approach.

While all of the aforementioned approaches have their respective advantages and disadvantages in measuring and modelling health in economic applications and have been employed with some success in the literature, they have mainly been used to describe the dynamic evolution of health during adulthood as inputs for structural models in health economics concerning retirement, housing or insurance decisions rather than studying the production technology of later life health maintenance or depreciation directly. Regarding the latter, the aforementioned approaches have some potential downsides that we aim to address in this paper. First, to the best of our knowledge, our paper is the first to explicitly study the dynamic interplay between physical capacity, cognitive capacity and related investments in the context of a structural non-linear dynamic latent factor model as first proposed by Cunha, Heckman, and Schennach (2010), which can generate new insights on the dynamic relationships between physical and cognitive capacity as well as investment into these important facets of human capital. Second, explicitly distinguishing between physical and cognitive capacity is thereby not only important due to increasing evidence for potentially important cross-effects between the two health domains cited above but also in view of likely differences in the consequences of depleted levels of physical vs cognitive capacity for functioning, participation and other important later life outcomes (Amengual, Bueren, and Crego, 2021; Crimmins, 2020). In the economics literature, there is to date only limited evidence on the potential cross-effects between physical and cognitive capacity maintenance with Schiele and Schmitz (2023) being a notable exception studying the effects of adverse physical health shocks on cognitive capacity in later life using non-structural event study methods. Third, our approach can accommodate a situation where information about a few latent factors needs to be extracted from many measurements of the underlying construct which can potentially suffer from severe measurement error.

Our analysis complements the aforementioned approaches to modelling and analyzing later-life health by delivering new insights on the dynamics of later-life human capital and related investments among older adults in the US. Our approach, thereby, highlights the structural production function of older adults concerning the maintenance and depreciation of physical and cognitive capacity. complementing more descriptive approaches. Our key findings are as follows: 1) There is substantial noise in all observed variables. While most measurements have a high correlation with the latent factor they measure, no single measurement dominates to an extent where it would be justified to just use a single variable and ignore the measurement error in the econometric analysis. 2) Despite a strong decline in means for physical and cognitive capacity, the rank order of these latent factors is remarkably stable. 3) Physical and cognitive capacity can be influenced by investments until very high ages. Cognitive stimulation is a specific investment into cognitive capacity. Physical exercise has a larger effect on physical capacity and a small effect on cognitive capacity.

The remainder of the paper is organized as follows: Section 2 provides information on our main data source and gives detailed description of the factor measurements. Section 3 describes our empirical approach and the challenges associated with it. Section 4 presents and discusses our results, and section 5 concludes.

2 Data and measurements

We base our empirical analysis on the 1992-2016 waves of the Health and Retirement Study (HRS) conducted by The University of Michigan. The HRS offers longitudinal panel data with representative sample of approximately 40,000 individuals living in the U.S. and aged 50 and above. The HRS core questionnaire offers rich set of measures of physical health, mental status, and behaviors. Measures of physical and cognitive capacities include self-reported diagnoses, subjective assessments, and objective biomedical markers. Additional off-wave surveys offer additional measures that are particularly relevant for our analysis. Specifically, we employ the Consumption and Activities Mail Survey (CAMS) (Health and Retirement Study, 2022b) to extract measurements for Exercise and Cognitive Stimulation.

Wherever possible, we include data prepared by the RAND corporation (Health and Retirement Study, 2022c), which provides a harmonized and easy-to-use version of the core HRS data. Out of the many variables we need, several are not included in the RAND HRS data, however, and we recur to the original core files (Health and Retirement Study, 2022b).

We start our analysis at age 68, when most people are retired and we start to see meaningful variation in the measures for physical and cognitive capacity that we have at our disposal. The last age we consider is 93, after which the sample size becomes small. Since the HRS questionnaire is administered biannually, we work with two-year transitions and age groups. For conciseness, we refer to these age groups by the lower bound included – “age 68” thus includes ages 68 and

69, and at the other end of the spectrum "age 92" comprises ages 92 and 93. Because men and women show very different aging patterns, we estimate the model separately for each gender and present all statistics in that way.

We standardize almost all measures to have mean zero and unit variance in the first age group included in our data. Any age trends are thus preserved. For example, until age 90, the mean of (residualized) grip strength declines by around 1.4 original standard deviations. At the same time, the dispersion of grip strength shrinks to around 80% of its original standard deviation. For categorical variables, all of which have numerical values with spacing 1, we add noise using uniform distributions on the unit interval. This preserves the original ordering and improves the numerical stability of the estimator below. Figure 1 shows the age trends in averages of all measurements; Figure A.1 in Appendix A presents the same trends in their standard deviations.

2.1 Physical Capacity

We employ six variables as measurements for physical capacity. Quite naturally, **vital status** is a dummy for being alive, which becomes zero in the first HRS wave after an individual has died. It is set to missing thereafter, so that the average of this variable can be interpreted as the probability of surviving until the next survey wave. The first row of Figure 1 shows the age trends in our measures of physical capacity. Unsurprisingly, survival probabilities decreases in age both for women (Figure 1a) and men (Figure 1b). Note that the level of survival probabilities is depressed because the HRS is very good at tracking respondents' dates of death even when they have not responded to previous waves. In this version of the data preparation, individuals who did not respond to a survey round would not enter the denominator of vital status.

The second measurement shown in Figures 1a and 1b is a version of the **frailty index** used, for example, in Hosseini, Kopecky, and Zhao (2022). The frailty index is the unweighted sum of all recorded medical conditions a doctor has diagnosed in an individual. These conditions comprise high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis. We reverse it so that higher values indicate better health. The reversed frailty index declines by 0.4 (women) and 0.3 (men) original standard deviations until the end of the age range we consider. Note that this trend and all those we will subsequently discuss are conditional on survival. Due to the high predictive power of the frailty index for mortality—as noted by Hosseini, Kopecky, and Zhao (2022) and others—the effect of mortality selection is particularly large here. For individuals still alive at age 80, average frailty at age 68 is 0.39 among women and 0.34 among men. By including vital status among the health measures, our model below will take care of this to some extent, but it is important to keep in mind for the descriptive statistics.

Grip Strength measurements were introduced to the HRS survey in 2006 and consist of in-home physical tests of the hand grip strength, conducted twice for each hand. To obtain our variable of use, we average the four measurements. Our measure of grip strength is then the residual of a regression of average grip strength on individuals' height. We partial height out because of the high correlation between height and grip strength (Steiber, 2016) and we do not expect differences in grip strength associated with differences in height to be indicative of physical capacity. Among all measures pertaining to physical health, grip strength shows the steepest decline.

Mobility summarizes difficulties in performing the various activities of daily living: walking several blocks, walking one block, walking across the room, climbing several flights of stairs, and climbing one flight of stairs. As with the frailty index, we add up indicators for each measurement and reverse the scale so that higher values are associated with greater mobility. Mobility declines strongly in age. At the same time, its standard deviation rises as mobility impairments become more frequent over time.

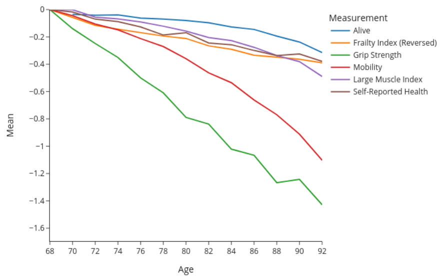
Closely related, the **Large Muscle Index** summarizes difficulties in performing a number of activities associated with large muscles' strength. These activities are sitting for two hours, getting up from a chair, stooping ot kneeling or crouching, and pushing or pulling a large object. Again, we revert the order of the values to have a positive association between the variable and physical capacity.

Finally, **Self-Reported Health** is a measure of health that is based on the respondent's self-assessed rating of their general health status. The values range from 1 (poor) to 5 (excellent). It probably is the most common health measure employed by economists as it provides an individuals' summary of her/his health in a single measure.

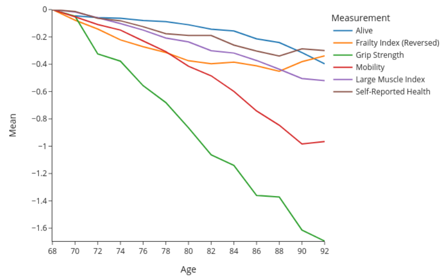
2.2 Cognitive Capacity

HRS interviews include a rich set of tests measuring respondents' cognitive capacity. For respondents that do not answer some of the cognitive test questions, HRS assumes non-random missing values and provides cross-wave imputation data in special data files (Health and Retirement Study, 2022a). Our measures of cognitive capacity are based on these cognitive tests and respondents' subjective ranking of their general memory status. In total, we employ five measures of cognitive capacity.

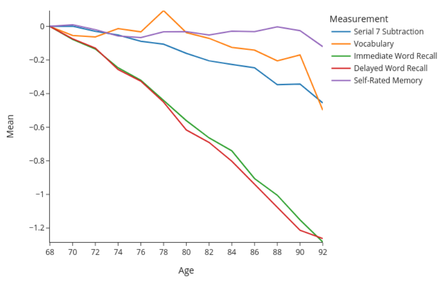
Serial 7 Subtraction is our first measure and is based on the test of serial sevens (SST) during which respondents are asked to subtract 7 from 100 and continue subtracting 7 from each resulting number for a maximum of five times. The respondents are then assigned scores based on the total number of correct answers. In psycho-medical literature SST has widely been used to assess cognitive status of patients with dementia and been generally regarded as a measure of



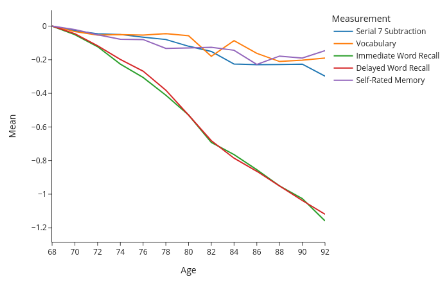
(a) Physical capacity, females



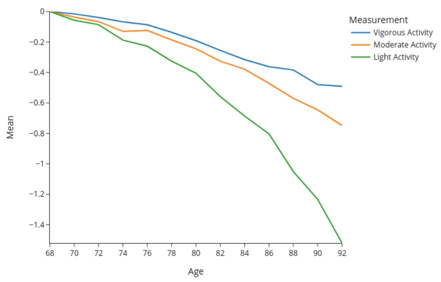
(b) Physical capacity, males



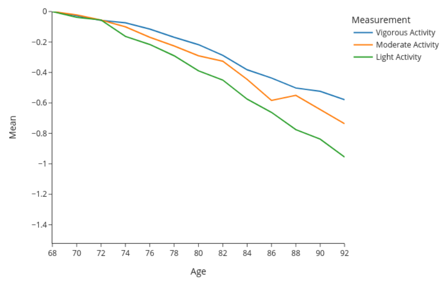
(c) Cognitive capacity, females



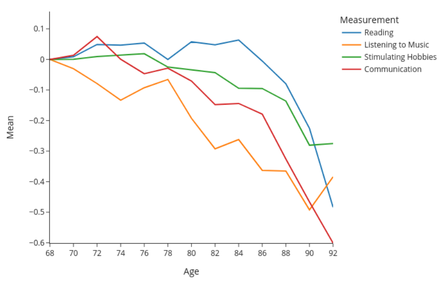
(d) Cognitive capacity, males



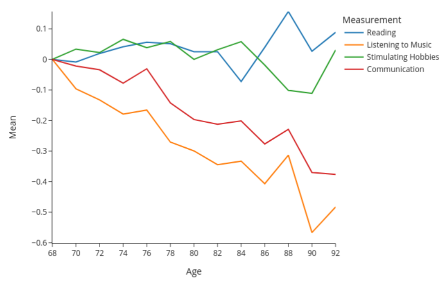
(e) Exercise, females



(f) Exercise, males



(g) Cognitive Stimulation, females



(h) Cognitive Stimulation, males

Figure 1. Average measurements by age

concentration (Karzmark, 2000). Figures 1c and 1d demonstrate a steady decline in concentration, as measured by the serial sevens, for both men and women, from our youngest age group to the oldest one being somewhat larger for women (0.5 units of original standard deviaton) than for men (0.4 units of original standard deviation).

Our second measure of cognitive capability is **Vocabulary** which is a test summarizing respondents' ability to provide correct definitions of words from a list of five words. One of two sets of words is assigned randomly at the first interview, and alternating sets are given during subsequent interviews. The two alternating sets of words are 1) repair, fabric, domestic, remorse, plagiarize; and 2) conceal, enormous, perimeter, compassion, audacious. We can see in Figures 1c and 1d that the vocabulary test has an age trend similar to that of the Serial 7 Subtraction, both in terms of absolute slopes and relative differences between men and women.

Immediate Word Recall is the third variable in Figures 1c and 1d and results from a test that asks the respondents to recall words (in any order) form a list of ten (later waves) or twenty (earlier waves) words, directly after being read the list. Examples of words included in a list are lake, car, army, etc. In in the initial wave, respondents were randomly assigned a list from the set of four lists and during the consequent four waves there were assigned a different list (McCammon, Fisher, Hassan, Faul, Rodgers, et al., 2022). **Delayed Word Recall** has the same structure as immediate word recall. In this task, respondents are asked to recall the same list of words once more, after spending several minutes on answering other survey questions. Word recall tests are widely used as measures of episodic memory frequently administered to patients with alzheimer's disease (see, e.g., Dixon and Frias, 2014; Runge, 2015).

Both of the word recall variables being measures of the same conceptual variable (episodic memory) perhaps explains the similar trends that they display. Of all the measurements of cognitive capacity, word recall variables have the sharpest decline over the age span in our model, and as with other measurements, the decline is larger for women than for men, with the caveat that our data are conditional on survival.

Finally, **Self-Rated Memory**, is our last measure of cognitive capacity and is based on respondents' self-assessed rating of their general memory status. The values range from 1 (poor) to 5 (excellent). Self-Rated Memory displays a moderate decline in both genders, which has a somewhat more pronounced trend among men.

2.3 Exercise and Cognitive Stimulation

We use **Vigorous, Moderate and Light Activities** as measures for investment in physical health. Each of these survey questions asks respondents how often they do vigorous (running, jogging, cycling, etc.), moderate (gardening, cleaning the car, walking at moderate pace, dancing, stretching) and light/mildly energetic (vacuuming, laundry, home repair), respectively.[1] Figures 1e and 1f show that with age people do less of all types of physical activities, with largely similar trends for men and women.

To obtain measures for cognitive stimulation, we utilized the CAMS survey which allowed us to construct measures of time respondents spend on different cognitively stimulating activities. Among these, our first measurement of cognitive stimulation is **Reading** that counts weekly hours spent on reading books, newspapers, or magazines. The association between reading and cognitive decline has been studied in psycho-medical literature, and reading has been found to be positively associated with hampered cognitive decline (Chang, Wu, and Hsiung, 2021). In Figures 1g and 1h we see that Reading has declining trend among women and is rather stable among men.

The second variable in Figures 1g and 1h is **Listening to Music**, and it meausres how many hours weekly respondents listen to music. The effects of music listening on cognitive functioning of at-risk patients have been studied in psycho-medical literature, and listening to music has been found to be beneficial for cognitive functioning (see, e.g., Särkämö and Soto, 2012; Särkämö, Tervaniemi, Laitinen, Forsblom, Soinila, et al., 2008). As with most measurements of cognitive stimulation, we observe a declining age trend for Listening to Music both among men and women.

Our last variables for cognitive stimulation are **Stimulating Hobbies** and **Communication** which summarize how many hours respondents spend weekly on various hobbies that may be expected to stimulate cognition, and the weekly hours spent on interacting with others, respectively. Stimulating Hobbies aggregates the survey variables that ask how many hours respondents spend on: 1) playing cards or solving jigsaw puzzles, 2) singing or playing instruments, 3) doing arts and crafts, and 4) going to movies or lectures. We construct the Communication variable as the sum of hours spent on visiting with others in person and communication via letters/phone/email. Looking at Figures 1g and 1h, Communication has similarly declining tred among men and women, whereas Stimulating Hobbies has a steeper slope for women and than for men.

2.4 Raw correlations in the data

Figures 2 and 3 show correlation matrices for women and men, respectively. Each figure contains two panels. The upper panels show within-period correlations until

1. The continuous availability of these three measures is one of the reasons for not using the first few HRS waves. Up until the sixth wave (year 2002) respondents were only asked if they do vigorous activities at least three times a week. Starting from wave seven, this questionnaire item was replaced by the three activity questions that we use in our study.

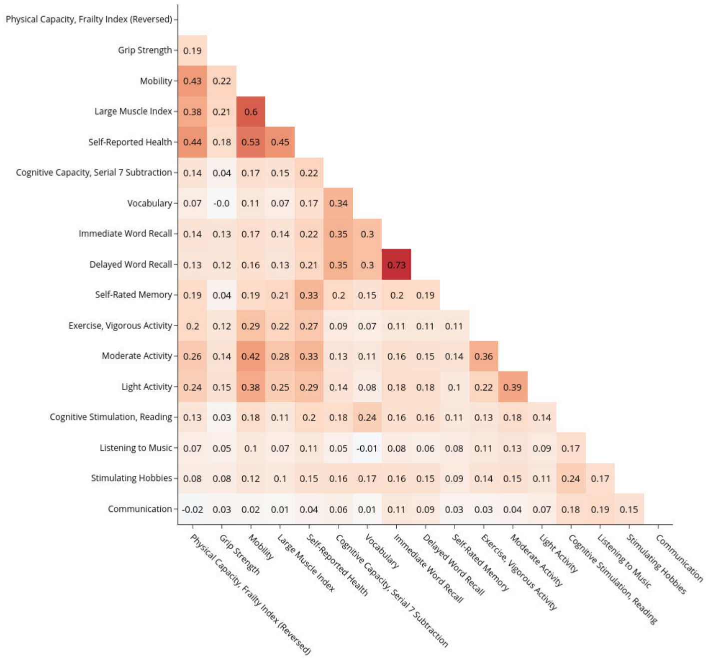
age 79, the lower panels do the same ages 80 and above. We show the lower triangular part of the correlation matrix. We leave out the indicator for being alive because we only measure the other variables whenever it is one. In addition to showing the numbers, we color the matrix' elements such that a correlation of 1 is dark red, 0 is white, and −1 is dark blue. Scaling is linear on both sides of the origin. Variables are ordered by factor, which we include in the label of the first measure pertaining to it. The measures in the first five rows and columns—from the reversed frailty index until self-reported health—load on physical capacity. The subsequent block of five rows and columns load on cognitive capacity. In the lower part of the matrix, exercise and cognitive stimulation load on three and four measures, respectively.

Several patterns are visually apparent in all four correlation matrices. First of all, the blocks of measures pertaining to each factor are clearly visible as having substantial cross-correlation throughout. For example, the first four entries in the first columns are the correlations of the reversed frailty index with the other measures loading on physical capacity. Across all four panels, correlations are at least 0.3 with the exception of the correlation of reversed frailty and grip strength, which is at least 0.1 throughout.

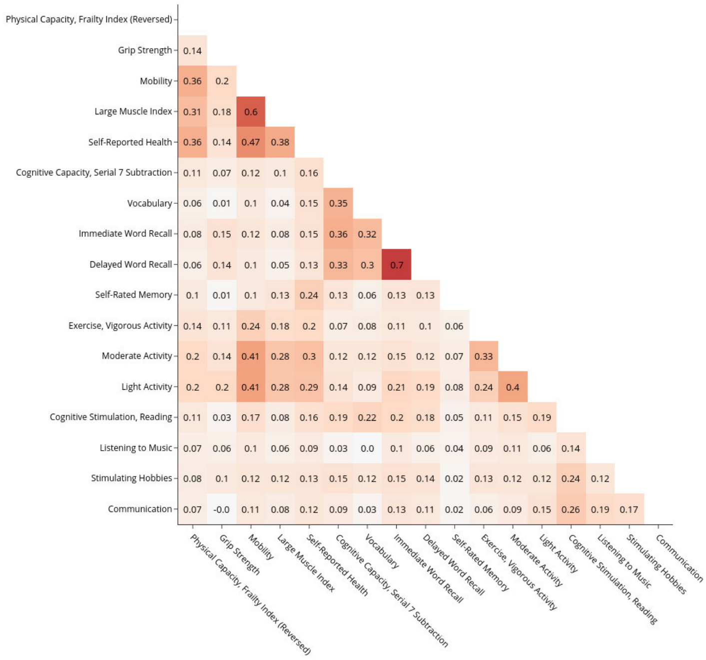
Similarly, the triangle with correlations for measurements pertaining to cognitive capacity—with the three corners (Serial 7 Subtraction, Vocabulary), (Serial 7 Subtraction, Self-Rated Memory), and (Delayed Word Recall, Self-Rated Memory)—has distinctly dark colors throughout. Unsurprisingly, correlations are particularly large between the two word recall tasks. The three correlations between the various types of physical activity are high throughout. The six elements to the bottom right to the matrix contain the correlations among the measures loading on cognitive stimulation. Among all factors, these have the weakest within-factor correlations with values ranging from 0.09 to 0.25. This is not very surprising as the variables do cover a much wider range of activities than, say, the various activity levels that load on exercising.

A second salient feature is that almost all elements are positive. This implies that it is important to model physical and cognitive capacity jointly with each other and with the two types of investments. This being written, there are clear level differences. Maybe unsurprisingly, the largest correlations are between measures of exercise and those of physical capacity. Most measures of cognitive capacity are substantially and positively related to variables measuring physical capacity and exercise, respectively. The correlation patterns are somewhat more mixed when it comes to cognitive stimulation and the other three factors.

This is related to our third broad observation: While the general patterns noted so far hold up across age groups and genders, there are some important differences. For example, the correlations of grip strength with other health measures are higher among women than among men, particularly at higher ages. Correlation patterns of individual measures pertaining to cognitive stimulation and

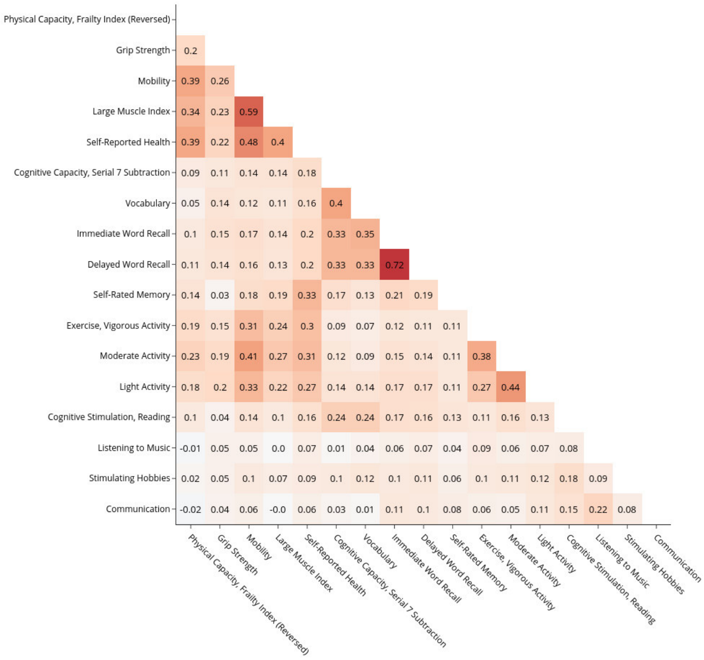


(a) Aged below 80

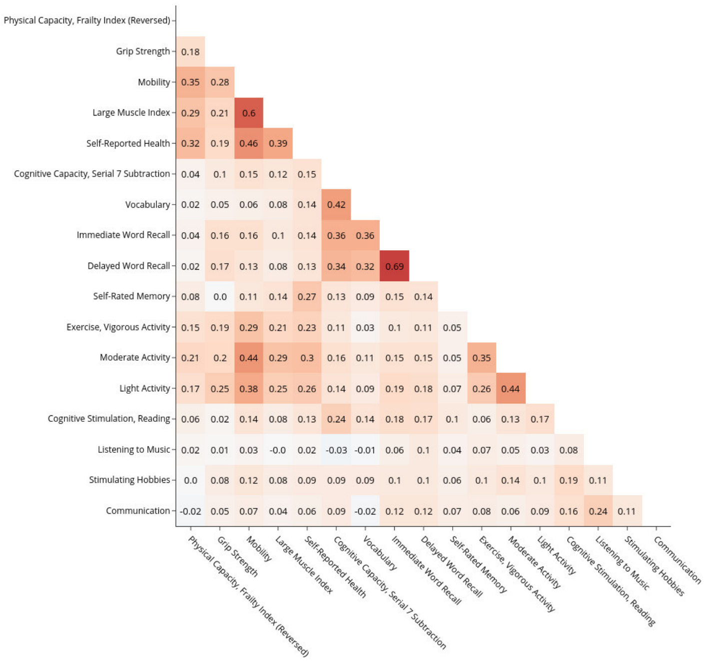


(b) Aged 80 and above

Figure 2. Cross factor measurement correlations (female).



(a) Aged below 80



(b) Aged 80 and above

Figure 3. Cross factor measurement correlations (male).

cognitive capacity are quite distinct among men and women, particularly at older ages. For example, among individuals aged 80 and above, reading and serial 7 subtraction have a correlation of 0.27 among women whereas it is 0.18 among men. Among women in this age group, listening to music is slightly negatively correlated with serial 7 subtraction and vocabulary scores. For men, the same correlations are small and positive.

While these patterns are informative, the $2 \times 2 \times 153$ numbers in Figures 2 and 3 are clearly too many to make sense of directly – and the matrices already reduce the 13 periods we observe in our data to 2. In the next section, we outline a framework that constructs latent variables for our four factors and which allows us to interpret their joint evolution.

3 Model

3.1 The Technology of Aging

Analyzing the joint evolution of physical and cognitive capacity and the effect physical exercise and cognitive stimulation have on both poses many econometric challenges.

(1) As discussed in the previous section, there are many potential observed variables to measure each concept we analyze. In order to make the results interpretable, their dimensionality has to be reduced.

(2) All observed variables are subject to measurement error, which is potentially large in many cases.

(3) Physical and cognitive capacity, exercise, and cognitive stimulation are dynamically intertwined in the sense that each of them has a potential effect on all others. For example, exercise should improve physical capacity. Conversely, it may well be that the cost of exercise might be higher at low levels of physical capacity because physiotherapy is less enjoyable than a walk in nature.

(4) The relationships between variables might change over time.

The Technology of Skill Formation (Cunha and Heckman, 2007; Cunha, Heckman, and Schennach, 2010) is an econometric framework that emerged to deal with very similar challenges in the context of skill formation during childhood. It distinguishes observed variables—for example an IQ test—from latent factors such as cognitive and non-cognitive skills. The technology is the law of motion of latent factors over multiple discrete time periods. Observed variables are stochastic functions of one or more latent factor. In addition to the latent factors of interest, the framework allows for observed or latent investments such as parental investments in skills or schooling.

To account for the multitude of potential effects, each latent factor may depend on lagged values of itself and all other latent factors. The law of motion of the latent factors is usually nonlinear. This is necessary to allow for different productivity of investments at different levels of skills. Moreover, it allows for dynamic complementarity, i.e., the fact that earlier investments may increase the productivity of later investments (Cunha and Heckman, 2007).

The Technology of Skill Formation maps in a straightforward way into our setting. Instead of of cognitive and non-cognitive skills, our Technology of Aging models physical and cognitive capacity. Instead of parental investments, we have exercise and cognitive stimulation.

Transition functions. We assume the following law of motion of our latent factors:

$$\begin{aligned} x_{1,t+1} &= \beta_{1,t} + \sum_{i=1}^{4} \gamma_{1,t,i} x_{t,i} + \sum_{i=1}^{4}\sum_{j=1}^{i} \delta_{1,t,i,j} x_{t,i} x_{j,t} + \eta_{1,t} \\ x_{2,t+1} &= \beta_{2,t} + \sum_{i=1}^{4} \gamma_{2,t,i} x_{t,i} + \sum_{i=1}^{4}\sum_{j=1}^{i} \delta_{2,t,i,j} x_{t,i} x_{j,t} + \eta_{2,t} \\ x_{3,t+1} &= \beta_{3,t} + \sum_{i \in \{1,2,3\}} \gamma_{3,t,i} x_{t,i} + \eta_{3,t} \\ x_{4,t+1} &= \beta_{4,t} + \sum_{i \in \{1,2,4\}} \gamma_{4,t,i} x_{t,i} + \eta_{4,t} \end{aligned} \tag{1}$$

Where x_1, x_2, x_3, and x_4 are physical capacity, cognitive capacity, exercise, and cognitive stimulation, respectively. β, γ and δ denote the technology parameters to be estimated. η denotes a stochastic shock.

The first two equations in (1) mean that physical and cognitive capacity follow a flexible functional form containing all lagged factors, their squares, and their interaction terms. This is known as the translog function in the skill formation literature (because skills are typically assumed to be measured in logs, not levels) and has been used by, for example, Agostinelli and Wiswall (forthcoming). The translog function allows for dynamic complementarity but does not assume it. We view it as a flexible approximation to an arbitrary underlying production function in the spirit of a nonparametric series estimator.

The bottom two equations in (1) relate to exercise and cognitive stimulations, respectively. Both investment factors are assumed to depend on their own lagged values along with the lagged values of physical and cognitive capacity.

Measurement system. We assume the measurement equations to be linear with an additively separable and normally distributed error term. All of them thus have the following form:

$$y_{\ell,t} = \alpha_{\ell,t} + \sum_{i=1}^{4} h_{\ell,t,i} x_{t,i} + \epsilon_{\ell,t} \tag{2}$$

where $y_{\ell,t}$ denotes the ℓ^{th} measurement in period t, α is the intercept of the measurement equation and h are factor loadings. In the empirical application we only have measurements that load on just one factor, so that for all measurements, three out of the potentially four loadings $h_{\ell,t}$ are zero by construction. Subject to identification requirements outlined in Cunha, Heckman, and Schennach (2010), this could easily be relaxed.

In typical applications of the Technology of Skill Formation, the number and type of available measurement variables varies strongly across periods. This is because any test score that is applicable to very young children would not work for older children. In our case, the measurements stay the same across periods and most of them can be assumed to be time-invariant, i.e. to have the same loading, intercept, and standard deviation of measurement error in each period.

3.2 Identification and Interpretation of Parameters

The econometric model implied by the Technology of Skill Formation is a Structural Equation Model or dynamic latent factor model. Linear Structural equation models are widely used since the 1970ies to study relationships between latent and observable variables. However, standard identification results and software for Structural Equation Models are not applicable to our setting because they usually require linearity assumptions or put restrictions on the connectedness of the underlying causal graph, which go beyond those encoded in our system (1).

Cunha, Heckman, and Schennach (2010) provide general nonparametric identification results for nonlinear dynamic latent factor models. The exact conditions for identification depend on the assumptions one is willing to put on the nature of measurement error. Typically, having at least two dedicated continuous measurements for each latent factor in each period is sufficient to identify an arbitrary production function under mild conditions. Doing so requires normalizations of location and scale in each period because latent factors do not have a natural unit of measurement.

A subsequent literature (Agostinelli and Wiswall, 2016; Freyberger, 2024) has shown that much fewer normalizations are required when empirical applications assume the popular constant-elasticity-of-substitution (CES) form, which implies restrictions on the location and scale of its outputs (see Appendix C.1 for details). Our specification of the production function (1) does not impose any such restrictions. However, as discussed previously, we have at least one age invariant measurement for each latent factor. We always use such measurements for normalizations, which pin down the location and scale of each corresponding factor in all periods.

The lack of natural units for the latent factors and the requirement for normalizations also poses challenges for the interpretation of the results. In short: any outcome that depends on transformations of measurements outside of the model, the choice of the measurement being normalized, or the values of the normalized parameters cannot be interpreted without further information. For details and a more formal definition see Freyberger (2024).

In practical applications, different ways of dealing with this have emerged. Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010) propose to anchor the latent factors in terms of observable cardinal variables. For example, they anchor cognitive and non-cognitive skills in terms of years of schooling, wages or the probability to commit a criminal offense. For each anchoring outcome, they re-estimate the model to obtain estimated production function parameters in terms of anchored factors. Attanasio, Meghir, and Nix (2020) do not have access to adult outcomes. Instead they communicate the variables that were normalized and state that results have to be interpreted with respect to the normalizations. Del Bono, Kinsler, and Pavan (2022) propose to simply standardize the variance of the latent factors in logs. This allows for statements such as increasing investment by 1 % increases skills by x %. While this is invariant to any normalization of location and scale in the measurement system, the approach is only valid if one defines that skills are measured in logs not levels. Due to the ordinality of skills, this is a valid but arbitrary definition and thus the approach falls short of its goal to be completely objective. Freyberger (2024) proposes to translate inputs and outputs of the production functions into ranks. This is invariant to any normalization of location and scale, assumptions on whether latent factors are measured in levels or logs and transformations of the measurements outside of the model.

We acknowledge that there is no single natural scale for latent factors and thus see value in all of the above approaches. For example, translating everything to ranks is a natural way of solving a problem that is caused by ordinality. Moreover, it makes the results completely invariant to many decisions made by the econometrician. However, it might not be as interpretable as anchoring approaches. For example, it destroys any time trend that was present in the measurements. To address the shortcomings of any single method, we thus use a combination of all of them.

We standardize age invariant measures with respect to their mean and standard deviation at age 68. We estimate the parameters of the production function, normalizing one age-invariant measure for each factor in period zero. The normalized measures are the reversed Frailty Index, Serial 7 Subtraction, Moderate Activity, and Reading. This preserves the time trend in the measurement variables and means that our estimated parameters and the time trend can roughly be interpreted in terms of standard deviations at age 68. For reference, we also show the marginal distributions of each latent factor and the joint distributions of each factor pair at multiple ages (see D.3).

3.3 Estimation

Multiple estimators for nonlinear dynamic latent factor models are available. Agostinelli and Wiswall (forthcoming) estimate the fist period factor loadings from ratios of covariances between measurements. To estimate production function parameters, they subsequently employ an iterative IV approach. Their method is very tractable; it comes at the cost of statistical efficiency. Our own experiments on simulated data suggest that it works well for models with few periods but becomes imprecise if there are ten or more periods, especially when the correlation between latent factors is high.

Attanasio, Cunha, and Jervis (2019) use linear regression on Bartlett factor scores with a correction approach. This estimator is computationally very attractive. However, it does not deal well with missing observations. Several of our variables are not contained in the core HRS questionnair; they are available for subsets of individuals at different points in time. Because of this, the estimator of Attanasio, Cunha, and Jervis (2019) is unsuitable for our application.

Attanasio, Meghir, and Nix (2020) first estimate the distribution of the latent factors as a mixture of normal distributions and then estimate the parameters of the production functions on a simulated sample from that distribution. This approach is computationally harder than the two previous ones but simpler than the maximum likelihood estimator by Cunha, Heckman, and Schennach (2010). The required assumptions are the same as for the likelihood estimator.

Cunha, Heckman, and Schennach (2010) use a maximum likelihood estimator. For computational tractability, they use nonlinear Kalman Filters to factorize the likelihood function into a product of conditional likelihoods. This estimator is computationally more difficult than the others. In its original formulation, numerical stability is often compromised. However, the estimator is statistically efficient and it can deal well with observations that are missing at random.

We derive a mathematically equivalent but numerically stable version of the likelihood estimator used by Cunha, Heckman, and Schennach (2010). Our versions replaces standard filters by square-root Kalman filters (Prvan and Osborne, 1988; van der Merwe and Wan, 2001), which are numerically more robust. The computational cost is similar to the original approach. The details of the original and the reformulated estimator as well as the exact assumptions required for estimation are described in Appendix B.

To account for mortality, we add a dummy variable for being alive as an additional measurement of physical capacity. This is analogous to a linear probability model of survival. Thus, the estimated health state of survivors is adjusted upwards, while the health state of everyone who has passed away is adjusted downwards compared to a state estimation that ignores mortality.

A flexible implementation of the new estimator can be found in the Python package skillmodels (Gabler, 2025). It uses JAX (Bradbury, Frostig, Hawkins, Johnson, Leary, et al., 2025) for just in time compilation and automatic differentiation. This reduces the computational cost drastically. We use optimagic/estimagic (Gabler, 2024) for numerical optimization and the calculation of standard errors. To generate good start values for the optimization, we first decompose the model into four single factor model with much fewer free parameters. In a second step we estimate a linear model. In the third step we estimate the full nonlinear model. We use pytask (Raabe, 2024) and the Templates for Reproducible Research Projects in Economics (Gaudecker, 2019) to automate our research project and to parallelize many tasks. The full estimation takes approximately four hours on a laptop.

4 Results

We present our results in three stages. First, we describe the measurement system. Next, we describe broad patterns for the transition equations. Finally, we dig deeper into the dynamic effects of changing factors along their distribution.

4.1 Measurement system

Table 1 shows the loadings and standard deviations of measurement errors of the measurement system.[2] The first panel shows the parameters that we constrain to be time-invariant. The three panels below display time-varying parameters of the system at ages 70, 80, and 90. We show loadings and standard deviations for women and men, respectively. Remember from Section 2 that we scale all measures—except for dummy measuring vital status, which retains its natural form—to have mean zero and unit variance in the initial period.

For the measurements loading on physical capacity, we normalize the reversed frailty index to have intercept zero and unit loading. We also restrict the parameters relating to mobility, the large muscle index, and self-reported health to be time-invariant – all of these have fairly similar time trends as seen in Figure 1 (note that mobility has a steeper trend than the others, but making the measurement system time-varying did not change results). All four measurement have similar factor loadings in the 1–1.5 range and the standard deviation in their measurement errors is very similar, too (0.7–0.8). The correlations between these four measurements are high throughout in the 0.6-0.85 range (see the correlation matrices in Section D.2 of the Appendix).

We leave the measurement systems for vital status and grip strength unrestricted across age groups. The standard deviation of measurement error in grip strength decreases over time; the loadings decrease for females and stay roughly

2. Tables D.1–D.8 in Appendix D.1 show the complete set of parameter estimates, including the intercepts.

Table 1. Loadings and Measurement Standard Deviations

			Female		Male	
			Loading	Meas. Std.	Loading	Meas. Std.
Age	Factor	Measurement				
All	Physical Capacity	Frailty Index (Reversed)	1.000	0.745*** (0.002)	1.000	0.804*** (0.002)
		Mobility	1.337*** (0.007)	0.713*** (0.003)	1.500*** (0.009)	0.721*** (0.003)
		Large Muscle Index	1.001*** (0.006)	0.729*** (0.003)	1.134*** (0.009)	0.763*** (0.003)
		Self-Reported Health	1.015*** (0.006)	0.753*** (0.002)	1.040*** (0.008)	0.790*** (0.003)
	Cognitive Capacity	Serial 7 Subtraction	1.000	0.903*** (0.004)	1.000	0.906*** (0.004)
		Vocabulary	0.861*** (0.022)	0.929*** (0.007)	0.971*** (0.026)	0.868*** (0.008)
		Immediate Word Recall	1.816*** (0.018)	0.585*** (0.003)	1.750*** (0.021)	0.595*** (0.004)
		Delayed Word Recall	1.836*** (0.018)	0.579*** (0.003)	1.718*** (0.020)	0.586*** (0.003)
	Exercise	Vigorous Activity	0.695*** (0.010)	0.802*** (0.004)	0.736*** (0.012)	0.813*** (0.005)
		Moderate Activity	1.000	0.796*** (0.004)	1.000	0.811*** (0.004)
		Light Activity	1.068*** (0.012)	0.934*** (0.004)	0.923*** (0.012)	0.860*** (0.004)
	Cognitive Stimulation	Reading	1.000	0.826*** (0.006)	1.000	0.667*** (0.008)
		Listening to Music	0.548*** (0.011)	0.939*** (0.006)	0.208*** (0.011)	1.050*** (0.008)
		Stimulating Hobbies	0.637*** (0.013)	0.895*** (0.005)	0.347*** (0.012)	0.999*** (0.006)
		Communication	0.582*** (0.011)	0.968*** (0.006)	0.306*** (0.011)	0.990*** (0.007)
70	Physical Capacity	Alive	0.068*** (0.006)	0.179*** (0.006)	0.091*** (0.008)	0.202*** (0.007)
		Grip Strength	0.482*** (0.040)	0.886*** (0.015)	0.641*** (0.056)	0.958*** (0.018)
	Cognitive Capacity	Self-Rated Memory	0.633*** (0.038)	0.950*** (0.011)	0.675*** (0.045)	0.943*** (0.013)
80	Physical Capacity	Alive	0.109*** (0.013)	0.262*** (0.013)	0.138*** (0.021)	0.301*** (0.022)
		Grip Strength	0.375*** (0.050)	0.847*** (0.019)	0.650*** (0.065)	0.880*** (0.022)
	Cognitive Capacity	Self-Rated Memory	0.428*** (0.046)	1.006*** (0.015)	0.557*** (0.058)	0.983*** (0.017)
90	Physical Capacity	Alive	0.233*** (0.066)	0.400*** (0.056)	0.271* (0.142)	0.424*** (0.102)
		Grip Strength	0.349*** (0.084)	0.717*** (0.028)	0.397*** (0.113)	0.753*** (0.052)
	Cognitive Capacity	Self-Rated Memory	0.377*** (0.104)	1.054*** (0.029)	0.484*** (0.144)	1.081*** (0.046)

Note: *** $p<0.01$; ** $p<0.05$; * $p<0.1$

constant for males. In sum, this means that the correlation between grip strength and the latent factor representing physical capacity stays roughty constant with age in the 0.3–0.4 range. The loading on vital status increases for both genders. Due to the fact that the dummy for being alive has its natural scale, the coefficient has a meaningful interpretation in terms of survival probabilities. At age 70, the interquartile range of physical capacity is 0.95 for women and 0.78 for men (see Appendix Section D.3). Changing physical capacity from its first to its third quartile thus increases the probability of survival by $0.95 \times 6.8\% = 6.5\%$ for women and $0.78 \times 9.1\% = 7.1\%$ for men. At age 80, the interquartile ranges are just below 1 and the loadings of 0.11 and 0.14, respectively, directly measure changes in survival chances as one moves across the outer quartiles. The same is true at age 90 for men ($\Delta_{\text{survival}} = 0.27$), for women the distribution is less dispersed at that age and an interquartile range of 0.8 implies a increase in survival probabilities of 0.19%. This is in line with the intuition that physical capacity is more predictive of death at older ages, as deterioration of overall health becomes a more important cause of death than fairly sudden shocks such as cancer or heart attacks (Gill, Gahbauer, Han, and Allore, 2010).

For measures pertaining to cognitive capacity, we normalize the results from the serial 7 subtraction task to have intercept zero and unit loading. Each measure is restricted to have the same factor loading and measurement error variance, regardless of age. Serial 7 subtraction and the vocabulary score look very similar in terms of loading and measurement error. For the word recall tasks, loadings are substantially higher and measurement errors are lower than this. Consequently, all correlations between these measures and the cognitive capacity factor are high throughout – around 0.5 for serial 7 subtraction and vocabulary; exceeding 0.8 for the word recall tasks. The measurement system of self-rated memory is allowed to vary with age. For both genders, its loading is estimated to be about 0.6 initially and decreases over time. The standard deviation of measurement error is around unity, with a slightly increasing trend. Consequently, the correlation of self-rated memory with cognitive capacity is declining with age which is consistent with Huang and Maurer (2019).

Given the similarity of our measurements for exercise, it is unsurprising that all three of them load substantially on the underlying factor. Moderate activity—the normalized measurement—has the largest correlation with the exercise factor at all ages. The correlation of vigorous activity and exercise declines over time whereas light activity goes the other direction. Both of these trends are more pronounced among women than among men.

Among the measurements loading on cognitive stimulation, we normalize the parameters on the time spent reading. This is also the dominant one among the four measurements with a standard deviation of its error around 0.8 (women) and 0.67 (men) and correlations with the factor exceeding 0.7 throughout. The errors on the other three measurements are between 0.9 and 1.05; their loadings are estimated to be around 0.6 for women and 0.2 for men. These coefficients translates into correlations with the cognitive stimulation factor of around 0.4 for women and 0.25 for men, which are roughly stable over time.

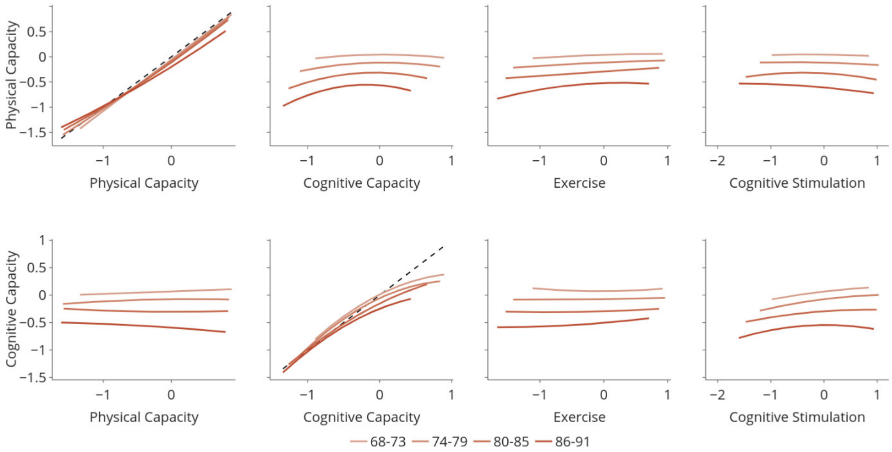
In sum, the measurements show a high correlation with the factors they are supposed to identify. For many measurements, it is sensible to restrict the model parameters to be time invariant and we do so. Measurements that are allowed to be changing with age vary in a way that makes sense in the light of prior literature. Differences between genders are not dramatic, but large enough to treating them separately in the estimation. Having established these direct relations to the data, we now turn to the core contribution of our paper: The joint evolution of physical and cognitive capacity and the impact of exercise and cognitive stimulation.

4.2 Transition equations

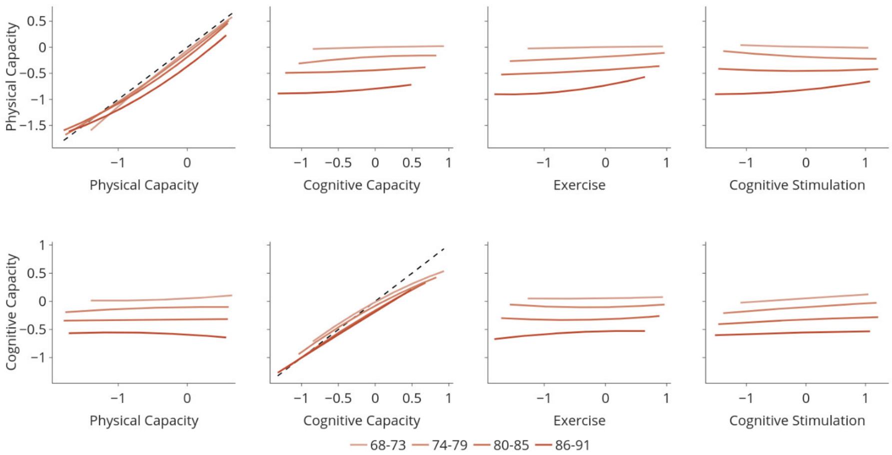
The translog production functions for physical and cognitive capacity have many parameters. In total, we have 15 coefficients per factor, which needs to be multiplied with four age groups (or"stages" in the terminology of Cunha, Heckman, and Schennach, 2010) and two genders. Furthermore, the parameters do not have intuitive interpretations without referring to precise values of the four factors in our model. We thus refrain from listing the parameters in the main text and relegate them to Tables D.9–D.16 in Appendix D.4. We note that the vast majority of parameters is very precisely estimated. The set of model parameters is completed with the initial distribution of states and the standard deviation of period-by-period innovations, which we relegate to Appendix D.5.

As a first pass, Figure 4 shows transition equations for physical capacity (first row of each subfigure referring to women and men, respectively) and cognitive capacity as a function of the input factors. Each of the sixteen panels contains four lines, one for each age group or stage. Input factors are kept at their median except for the one on the x-axis, which is varied from the 1st to the 99th percentile of its distribution in the respective age group.

The top left panel in Figure 4a thus shows the result of the following thought experiment: Conditional on current age, what is a woman's expected value of physical capacity in two years as a function of her current physical capacity while fixing cognitive capacity, exercise, and cognitive stimulation at their median values. The results show that there is a high degree of persistence in all age groups. For the upper part of the distribution of physical capacity, the lines are below the 45°-line (the distributions at ages 70, 80, and 90 are shown in in Appendix D.3, Figures D.7–D.9; as a rough guide to interpret the first panel of Figure 4a, the first quartile at age 90 has a value of −1.2). The transition function is below 45°-line everywhere in the youngest age group, which has the steepest slope throughout. This means that at median levels of cognitive capacity, exercise, and cognitive stimulation, physical capacity will unambiguously decline in expectation regardless of



(a) Transitions, females



(b) Transitions, males

Figure 4. Next period states as a function of current states, other factors evaluated at the median

the initial level. In contrast, for very low values of physical capacity at older ages, there would be some mean reversion – if all other factors were at their median. Of course, there might be substantial costs to reaching median levels of exercise or cognitive stimulation if physical capacity is very low, for example.

Increased cognitive capacity is associated with a slightly more favorable evolution of physical capacity. For example, changing cognitive capacity from its first quartile (−0.63) to its third quartile (−0.17) at age 80 is associated with an increase of age-82 physical capacity of 0.02 units or just under 2 percentiles. The corresponding effects of increased exercise are positive as well and tend to be larger. The same interquartile move for exercise at age 80 (from −0.81 to −0.06) leads to an increase of physical capacity by 0.16 units, which corresponds to almost 5 percentiles. The effects of cognitive stimulation on the dynamics of physical capacity are often slightly negative at median levels of physical capacity, cognitive capacity, and exercise.

The second row of Figure 4a shows the corresponding effects for the evolution of cognitive capacity. We start with the second panel, which contains the own-effects. They are much less persistent than the own-effects for physical capacity as evident by the flatter slopes at all ages. The four lines are also further apart except at the very bottom of the distribution of cognitive capacity. This means that at almost any level of cognitive capacity, the dynamics are worse for higher ages, provided all other factors are at their median.

The first panel in the second row of Figure 4a displays modestly positive effects of physical capacity in the lower age groups; these become zero for higher ages and, in the highest age group, turn out to be negative at very low levels of physical capacity. Exercise has mostly positive effects on the evolution of cognitive capacity at median values of other states with an exception being in the lower half of the exercise distribution during women's upper seventies. Finally, cognitive stimulation has positive effects almost everywhere.

Figure 4b shows the same set of transition functions for men. Again, the broad patterns are fairly similar to women, but there are some important differences. For example, physical capacity is deteriorating more quickly for ages 74 and beyond across the entire distribution of current physical capacity; only at the very bottom of the distribution there is some sign of mean reversion. For the own-effects of physical capacity, there is a similar pattern to what we noted for the own effects of cognitive capacity among women: At almost any level of physical capacity, the dynamics are worse for higher ages, provided all other factors are at their median. In contrast, for cognitive capacity, the same effect is somewhat less pronounced then for women; the lower two age group and the upper two age groups look much more similar to each other there. The signs and magnitudes we noted for the off-diagonal elements generally hold up.

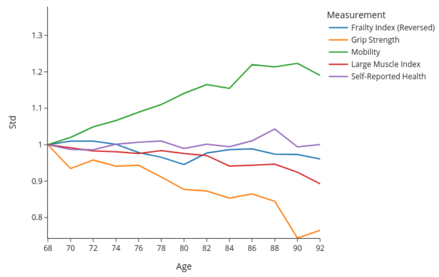
5 Conclusions and Outlook

We adapt a nonlinear dynamic latent factor framework that was developed for skill formation of children to study the physical and cognitive decline between ages 68 and 93. To this end, we incorporate mortality into the model. The model is estimated with a rich set of measures from the Health and Retirement Study.

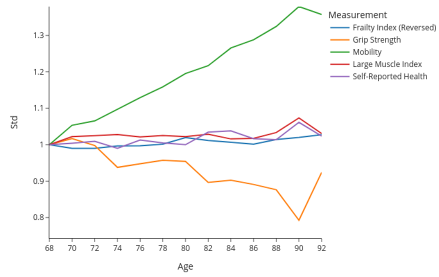
We document a large amount of measurement error in all observed variables. While most measurements have a high correlation with the latent factor they measure, no single measurement is a good enough proxy to use in isolation. A dynamic latent factor model is therefore a good fit for this setting. Having a rich set of time invariant measurements for each latent factor, lets us overcome some of the challenges related to the interpretability of latent factors. To make our results even more interpretable we also present them in terms of population ranks and use survival probabilities to anchor physical capacity.

We find that, despite a strong decline in means for physical and cognitive capacity, the rank order of these latent factors is remarkably stable over periods. Nevertheless, physical and cognitive capacity can be influenced by investments until very high ages. Cognitive stimulation is a specific investment into cognitive capacity. Physical exercise has a larger effect on physical capacity and a small effect on cognitive capacity.

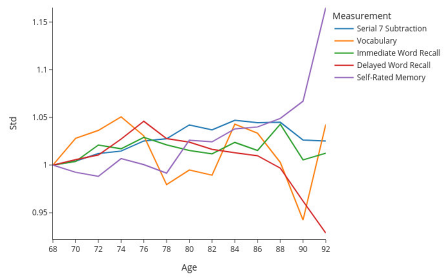
Appendix A Additional Background on the Data and Measurements



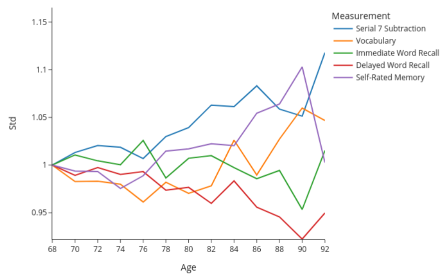
(a) Physical capacity, females



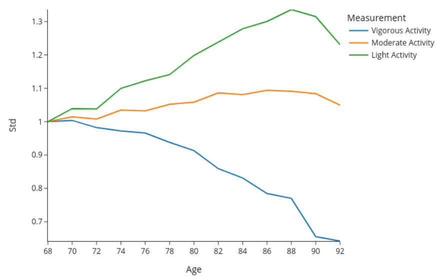
(b) Physical capacity, males



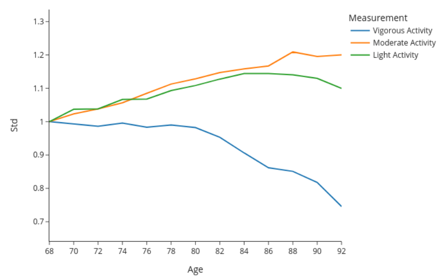
(c) Cognitive capacity, females



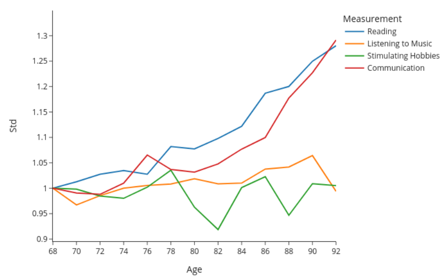
(d) Cognitive capacity, males



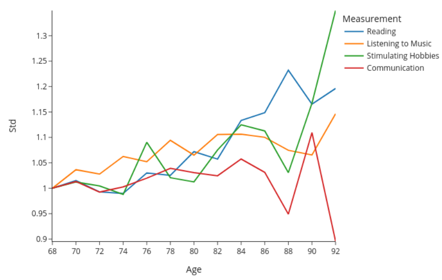
(e) Exercise, females



(f) Exercise, males



(g) Cognitive Stimulation, females



(h) Cognitive Stimulation, males

Figure A.1. Standard deviation of measurements by age

Appendix B The Maximum Likelihood Estimator

B.1 State Estimation

B.1.1 Preliminaries. To discuss the econometric approach used in this paper and potential alternatives it is convenient to express the model in state space notation.

To do so, let $\boldsymbol{x}_t \in \mathscr{R}^N$ denote the vector of latent factors (i.e. physical capacity, cognitive capacity, physical exercise and cognitive stimulation) in period t.

Similarly, let $\boldsymbol{y}_t \in \mathscr{R}^{L_t}$ denote the vector of all observable measurements in period t.

Then the transition function of the latent factors can be written as:

$$\boldsymbol{x}_{t+1} = F_t(\boldsymbol{x}_t) + \eta_t \tag{B.1}$$

where η_t is a vector of error terms with η_t^j on the j^{th} position. Let $\boldsymbol{Q}_t$ denote the covariance matrix of η_t

The linear measurement system can be written as:

$$\boldsymbol{y}_t = \boldsymbol{H}_t\boldsymbol{x}_t + \boldsymbol{\epsilon}_t \tag{B.2}$$

where $\boldsymbol{H_t}$ is a matrix of coefficients known as factor loadings and $\boldsymbol{\epsilon}_t$ is a vector of measurement errors with $\epsilon_{t,l}$ on the l^{th} position. Let $\boldsymbol{R}_t$ denote the covariance matrix of $\boldsymbol{\epsilon}_t$.

Equations B.1 and B.2 define a state space model. Equation B.1 is called transition equation. Equation B.2 is called measurement equation. The vector $\boldsymbol{x}_t$ is called the state of the system. The matrices $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are called process noise and measurement noise, respectively.

To see why it was handy to rewrite the technology of skill formation in state form, assume for a moment that the transition function F_t (including parameters) as well as the matrices $\boldsymbol{H}_t$, $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are known for all $t \in T$ but the state vectors $\boldsymbol{x}_t$ are unknown and have to be estimated from measurements $\boldsymbol{y}_t$. This problem is known as optimal state estimation, which is a well researched topic in physics and engineering.

To efficiently estimate the state vector in period t, an estimator should not only use measurements from this period, but also take the information from all previous measurements into account. For linear systems, Kalman filters are the method of choice for state estimation (Kalman, 1960). For nonlinear systems, several nonlinear variants of the Kalman filter have been developed. Kalman filters treat the state of a system itself as random vector. Therefore, they are sometimes classified as Bayesian filters.

Kalman filters consist of a predict and an update step. They are initialised with an initial estimate for the mean $\bar{\boldsymbol{x}}_0$ and covariance matrix $\boldsymbol{P}_0$ of the distribution of the state vector. Then, in each period, the new measurements are incorporated to update the mean and covariance matrix of the state vector. After that, the transition equation is used to predict the mean and covariance matrix of the state vector in the next period. This predicted state vector can then again be updated with measurements.

For the application of Kalman filters, the following assumptions must hold:

(1) $\eta_t \sim \mathcal{N}(\mathbf{0_N}, \boldsymbol{Q_t})$ where $\mathbf{0}_N$ denotes a vector of zeros of length N, $\boldsymbol{Q_t}$ is a diagonal matrix.

(2) The η_t^j are serially independent over all t.

(3) $\epsilon_t \sim \mathcal{N}(\mathbf{0}_{L_t}, \boldsymbol{R}_t)$ where $\boldsymbol{R}_t$ is a diagonal matrix.

(4) The $\epsilon_{t,l}$ are serially independent over all t.

(5) $\epsilon_{t,l}$ and η_t^j are independent of $\boldsymbol{x}_t$ for all $t = 1, \ldots, T$, $l = 1, \ldots, L$ and each factor j.

(6) The distribution of the state vector $p(\boldsymbol{x}_t)$ can be approximated by a mixture of normal distributions for all $t = 1, \ldots, T$

Due to the assumption of a linear measurement system, the state vector can be estimated by combining the update step of a linear Kalman filter with the predict step of a nonlinear Kalman filter. For computational reasons, it will be convenient not to incorporate all measurements at once but to perform a separate update step for each measurement.

B.1.2 The Update Step of the Kalman Filter. The aim of the Kalman update is to efficiently combine information from measurements in the current period with previous measurements. To do so, the measurement function is used to convert the pre-update state vector into predicted measurements for the current period (equation B.3). The difference between the predicted and actual measurements is called residual (equation B.4). This residual, scaled by the so called Kalman gain, is then added to the pre-update state vector (equation B.8). The Kalman gain is smaller if the variance of the measurement (calculated by equation B.6) is large. This has the intuitive consequence that noisy measurements receive a low weight. The Kalman gain becomes larger if the pre-update covariance matrix has large diagonal entries (equation B.5 and B.7). Thus, measurements receive more weight if the pre-update state is known imprecisely due to bad initial values or a high process noise, for example. After the incorporation of the measurements, the state is always known with the same or more precision than before. This is reflected by subtracting a positive semi-definite matrix from the pre-update covariance matrix (equation B.9).

Let $\bar{\boldsymbol{x}}_{t|y_{t,l}^-}$ denote the mean of the conditional distribution of the state vector given all measurements up to but not including the l^{th} measurement in period t. Let $\boldsymbol{P}_{t|y_{t,l}^-}$ denote the covariance matrix of this distribution. Let $\boldsymbol{h}_{t,l}$ denote the l^{th}

row of $\boldsymbol{H}_t$. Let $r_{t,l,l}$ be the l^{th} diagonal element of $\boldsymbol{R}_t$. The update step that incorporates the l^{th} measurement into the estimate is given by the following equations:

$$\bar{y}_{t,l|y^-_{t,l}} = \boldsymbol{h}_{t,l}\bar{\boldsymbol{x}}_{t|y^-_{t,l}} \qquad \bar{y}_{t,l|y^-_{t,l}} = E(y_{t,l}|y^-_{t,l}) \tag{B.3}$$

$$\delta_{t,l} = y_{t,l} - \bar{y}_{t,l|y^-_{t,l}} \qquad \delta_{t,l} \text{ can be interpreted as residual} \tag{B.4}$$

$$\boldsymbol{f}_{t,l} = \boldsymbol{P}_{t|y^-_{t,l}}\boldsymbol{h}^T_{t,l} \qquad \boldsymbol{f}_{t,l} \text{ is an intermediate result} \tag{B.5}$$

$$\sigma_{t,l} = \boldsymbol{h}_{t,l}\boldsymbol{f}_{t,l} + r_{t,l,l} \qquad \sigma_{t,l} \text{ is the variance of } y_{t,l} \tag{B.6}$$

$$\boldsymbol{k}_{t,l} = \frac{1}{\sigma_{t,l}}\boldsymbol{f}_{t,l} \qquad \boldsymbol{k}_{t,l} \text{ is the (scaled) Kalman gain} \tag{B.7}$$

$$\bar{\boldsymbol{x}}_{t|y_{t,l}} = \bar{\boldsymbol{x}}_{t|y^-_{t,l}} + \boldsymbol{k}_{t,l}\delta_{t,l} \qquad \bar{\boldsymbol{x}}_{t|y_{t,l}} \text{ is the updated mean} \tag{B.8}$$

$$\boldsymbol{P}_{t|y_{t,l}} = \boldsymbol{P}_{t|y^-_{t,l}} - \frac{1}{\sigma_{t,l}}\boldsymbol{f}_{t,l}\boldsymbol{f}^T_{t,l} \qquad P_{t|y_{t,l}} \text{ is the updated covariance matrix} \tag{B.9}$$

B.1.3 The Predict Step of the Kalman Filter. In linear systems, the mean and covariance matrix of the system can be propagated to the next period by simply applying the linear transition equation. With a nonlinear transition function, however, this is not possible, as $E(f(X) \neq f(E(X)$ in general. For the nonlinear predict step, two basic options exist: The *extended Kalman filter* and the *unscented Kalman filter*. Cunha, Heckman and Schennach choose the unscented Kalman filter because it has been shown to be more reliable in a wide range of settings (Van Der Merwe, 2004).

The intuition of the predict step of the unscented Kalman filter is relatively simple: firstly, a deterministic sample of points in the state space, called sigma points (equation B.10), and accompanying weights are chosen (equation B.11). Usually these are $2N+1$ points and weights, where N is the length of the state vector. Secondly, these sigma points are transformed using the true nonlinear transition equation. Thirdly, the weighted sample mean is used as estimate for the next period mean of the state vector (equation B.12). Fourthly, the sum of the covariance matrix of the process noise and the weighted sample covariance of the transformed sigma points is used as estimate of the covariance matrix of the state vector (equation B.13). Intuitively, the addition of the process noise accounts for the fact that the prediction always adds some uncertainty about the state of the system.

For the choice of sigma points and sigma weights, many different algorithms exist. All have in common that some form of matrix square root of the covariance matrix of the state vector is taken. Two definitions of matrix square root exist: 1) $\boldsymbol{A}$ is a matrix square root of $\boldsymbol{P}$ if $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}$. 2) $\boldsymbol{A}$ is a matrix square root of $\boldsymbol{P}$ if $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^T$. The matrix square root is not unique in general and some matrices do not have a square root. However, all symmetric positive semi-definite matrices, i.e. all valid covariance matrices, can be decomposed into $\boldsymbol{P} = \boldsymbol{L}\boldsymbol{L}^T$ where $\boldsymbol{L}$ is lower triangular (Zhang, 1999). For the unscented Kalman filter, both definitions of matrix square root work. Below, the sigma point algorithm proposed by Julier and Uhlmann (1997), is presented without reference to a particular type of matrix square root:

Let $\kappa \in \mathbb{R}$ be a scaling parameter. Usually, κ is set to 2 if the distribution of the state vector is assumed to be normal. Let $\boldsymbol{P}_{t|t}$ denote the covariance matrix of the state vector, conditional on all measurements up to and including period t. Define $\boldsymbol{S}_{t|t} \equiv \sqrt{\boldsymbol{P}_{t|t}}$ as the matrix square root of $\boldsymbol{P}_{t|t}$ and let $\boldsymbol{s}_{t,n}$ denote its n^{th} column.

Sigma points are calculated according to the following equations:

$$\begin{aligned}
\chi_{t,n} &= \bar{\boldsymbol{x}}_{t|t} & \text{for } n &= 0 \\
\chi_{t,n} &= \bar{\boldsymbol{x}}_{t|t} + \sqrt{N+\kappa}\ \boldsymbol{s}_{t,n} & \text{for } n &= 1, \ldots N \\
\chi_{t,n} &= \bar{\boldsymbol{x}}_{t|t} - \sqrt{N+\kappa}\ \boldsymbol{s}_{t,n} & \text{for } n &= N+1, \ldots 2N
\end{aligned} \tag{B.10}$$

where $\chi_{t,n}$ is the n^{th} sigma point at period t that is calculated after incorporating all measurements of that period. The corresponding sigma weights are calculated as follows:

$$w_{t,n} = \frac{\kappa}{N+\kappa} \qquad \text{for } n = 0 \tag{B.11}$$

$$w_{t,n} = \frac{1}{2(N+\kappa)} \qquad \text{for } n = 1, \ldots, 2N$$

where $w_{t,n}$ is the n^{th} sigma weight. Define $\widetilde{\chi}_{t,n} \equiv F_t(\chi_{t,n})$ where $F_t(\cdot)$ is defined as in equation B.1. Then the predict step of the unscented Kalman filter is given by:

$$\bar{\boldsymbol{x}}_{t+1|t} = \sum_{n=0}^{2N} w_{t,n} \widetilde{\chi}_{t,n} \tag{B.12}$$

$$\boldsymbol{P}_{t+1|t} = \left[\sum_{n=0}^{2N} w_{t,n} (\widetilde{\chi}_{t,n} - \bar{\boldsymbol{x}}_{t+1|t})(\widetilde{\chi}_{t,n} - \bar{\boldsymbol{x}}_{t+1|t})^T \right] + \boldsymbol{Q}_t \tag{B.13}$$

B.2 The Likelihood Interpretation of the Kalman Filter

Of course, the parameters of the function F_t and the matrices $\boldsymbol{H}_t$, $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are unknown in reality. However, they can be estimated by maximum likelihood. The direct maximization of the likelihood function would involve the evaluation of high dimensional integrals which is computationally very expensive (Cunha, Heckman, and Schennach, 2010). Instead, Kalman filters can be used to reduce the sumber of computations required for each evaluation of the likelihood function dramatically.

To see how, define $\boldsymbol{\theta}$ as the vector with all estimated parameters of the model. Then, the likelihood contribution of individual i is given by:

$$\mathscr{L}(\boldsymbol{\theta}|\boldsymbol{y}_1,\dots,\boldsymbol{y}_T) \equiv p_\theta(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T) = \prod_{t=1}^{T}\prod_{l=1}^{L_t} p_\theta(y_{t,l}|y_{t,l}^-) \tag{B.14}$$

where $p_\theta(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T)$ denotes the joint density of all measurements for individual i, conditional on the parameter vector θ and $p_\theta(y_{t,l}|y_{t,l}^-)$ is the density of the l^{th} measurement in period t, given all measurements up to but not including this measurement. The subscript i is again omitted for readability.

To see how this relates to the Kalman filter, recall that for each $t = 1,\dots,T$ and each $l = 1,\dots,L_t$, equation B.3 calculates $\bar{y}_{t,l|y_{t,l}^-}$, i.e the expected value of the l^{th} measurement in period t, conditional on all previous measurements. In addition, due to the normality and independence assumptions on the error terms and the factor distribution, $y_{t,l}$ is normally distributed around $\bar{y}_{t,l|y_{t,l}^-}$. Equation B.6 can be used to calculate the variance $\sigma_{t,l}$ of this distribution. Thus, $p_\theta(y_{t,l}|y_{t,l}^-) = \phi_{\bar{y}_{t,l|y_{t,l}^-},\sigma_{t,l}}(y_{t,l})$ where $\phi_{\mu,\sigma}(\cdot)$ is the density of a normal random variable with mean μ and variance σ.

A nice feature of the estimator based on this factorization of the likelihood function is that it can deal very well with missing observations. If measurement $y_{t,l}$ is missing for inidividual i, the corresponding update of the state vector is just skipped. More formally, this means that the missing measurement is integrated out from the likelihood function.

B.3 Numerical stability

B.3.1 Numerical challenges. While the Kalman filter based maximum likelihood estimator is statistically and computationally efficient, it is numerically unstable. The numerical instability caused by floating point imprecision is inherent to Kalman filters and has been discovered soon after Kalman published his original article. Since then, the precision of computers has increased enormously such that nowadays numerical problems are not a big issue for well specified Kalman filters. However, during the maximization of the likelihood function the optimizer might pick parameter combinations that are far from leading to a well specified filter.

The numerical problems manifest themselves in two places:

(1) In the update step, the subtraction in equation B.9 can lead to negative diagonal elements in the updated covariance matrix of the state vector. While this is mathematically impossible in a well specified Kalman filter, numerical imprecisions and badly specified Kalman filters during the maximization process make it possible.

(2) Even if the covariance matrix of the state vector has nonnegative diagonal entries, numerical imprecisions might render it not positive semi-definite. With this the existence of a matrix square root is not guaranteed, which can make the calculation of sigma points impossible.

Cunha, Heckman and Schennach mention the numerical problems in their supplementary material. To solve the first problem, they recommend to find good initial values for the maximization by first constraining some parameters and letting the code find good initial values for the others. For the second problem, they propose to set all off-diagonal elements of $\boldsymbol{P}$ to zero before taking the square root, which then corresponds to taking the element wise square root of the diagonal elements. While this prevents the estimator from crashing, it is not standard practice in Kalman filtering and it is not guaranteed that an estimator based on this type of matrix square root produces reliable results.

B.3.2 Outline of the Solution. A better approach is to use a square root implementation of the Kalman filter. Many different square root Kalman filters exist. They are mathematically equivalent to normal Kalman filters but numerically more stable.

Instead of propagating the full covariance matrix of the state vector, square root Kalman filters propagate the square root of this matrix. This has three advantages:

(1) It avoids overflow errors due to numbers with very small or large absolute values, as taking the square root makes large numbers smaller and small numbers larger.
(2) By using a matrix square root $\boldsymbol{A}$ of the type $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^T$, the problematic covariance matrix is guaranteed to be positive semi-definite (Zhang, 1999), i.e. a valid covariance matrix. In particular, its diagonal entries are sums of squared terms and, consequently, guaranteed to be nonnegative. This solves the first problem.
(3) By choosing an appropriate pair of square root update and predict algorithms, taking matrix square roots can be completely avoided. This eliminates the second problem.

The computational requirements of square root filters are comparable to those of normal Kalman filters. In the nonlinear case, they are even lower. For a maximally robust estimator, we use a pair of square root update and predict algorithms that completely avoid taking matrix square roots. The algorithm for the update was developed by Prvan and Osborne (Prvan and Osborne, 1988). The unscented square root predict step was proposed by Van Der Merwe and Wan (van der Merwe and Wan, 2001). Both propagate the transpose of a lower triangular matrix square root of the state covariance matrix.

B.3.3 The QR Decomposition of a Matrix. Both square root algorithms rely on a matrix factorization called QR decomposition. Note that in this subsection, $\boldsymbol{Q}$ and $\boldsymbol{R}$ do not denote the covariance matrices of the process and measurement noise but factors into which a matrix is decomposed.

$\boldsymbol{QR}$ is called QR decomposition of an $m \times n$ matrix $\boldsymbol{A}$ with $m \geq n$ if:

(1) $\boldsymbol{A} = \boldsymbol{QR}$

(2) $\boldsymbol{Q}$ is an orthogonal $m \times m$ matrix

(3) $\boldsymbol{R}$ is an $m \times n$ matrix and the first n rows of $\boldsymbol{R}$ form a upper triangular matrix and its remaining rows only contain zeros

The QR decomposition of a matrix always exists but is not unique. A useful property of the QR decomposition is that:

$$\boldsymbol{A}^T\boldsymbol{A} = (\boldsymbol{QR})^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{R} \tag{B.15}$$

where the last equality comes from the defining property of orthogonal matrices that $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{QQ}^T = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix. Thus, the upper triangular part of $\boldsymbol{R}$ is the transpose of a lower triangular matrix square root of $\boldsymbol{A}^T\boldsymbol{A}$. For convenience, let $qr(\boldsymbol{A})$ denote the QR decomposition of $\boldsymbol{A}$ that only returns the upper triangular part of the matrix $\boldsymbol{R}$.

B.3.4 The Update Step of the Square-Root Kalman Filter. Let $\boldsymbol{S}_{t|y^-_{l,t}}$ be a lower triangular matrix square root of $\boldsymbol{P}_{t|y^-_{l,t}}$ and keep the rest of the notation as in section B.1. Then, the square root update that incorporates the l^{th} measurement in period t is given by the following equations:

$\bar{y}_{t,l|y^-_{t,l}}$ and $\delta_{t,l}$ are calculated as in equation B.3 and B.4 respectively. Then the following intermediate results are calculated.

$$\boldsymbol{f}^*_{t,l} = \boldsymbol{S}^T_{t|y^-_{t,l}}\boldsymbol{h}^T_{t,l} \tag{B.16}$$

$$\boldsymbol{M}_{t,l} = \begin{bmatrix} \sqrt{r_{t,l,l}} & \boldsymbol{0}^T_N \\ \boldsymbol{f}^*_{t,l} & \boldsymbol{S}^T_{t|y^-_{t,l}} \end{bmatrix} \tag{B.17}$$

It can be shown that:

$$qr(\boldsymbol{M}_{t,l}) = \begin{bmatrix} \sqrt{\sigma_{t,l}} & \frac{1}{\sqrt{\sigma_{t,l}}}\boldsymbol{f}^T_{t,l} \\ \boldsymbol{0}_N & \boldsymbol{S}^T_{t|y_{t,l}} \end{bmatrix} \tag{B.18}$$

where $\boldsymbol{S}^T_{t|y_{t,l}}$ is the transpose of a lower triangular square root of the updated covariance matrix and $\boldsymbol{0}_N$ denotes a column vector of length N that is filled with zeros.

The matrix in equation B.18 also contains $\boldsymbol{f}_{t,l}$ and $\sigma_{t,l}$ such that the Kalman gain can be calculated as in equation B.7 and the mean of the state vector can be updated as in equation B.8.

To see why equation B.18 holds, define $\boldsymbol{U}_{t,l} \equiv qr(\boldsymbol{M}_{t,l})$ and partition it as follows:

$$\boldsymbol{U}_{t,l} = \begin{bmatrix} \boldsymbol{U}_{1,1} & \boldsymbol{U}_{1,2} \\ \boldsymbol{0} & \boldsymbol{U}_{2,2} \end{bmatrix} \tag{B.19}$$

where $\boldsymbol{U}_{1,1}$ is a scalar, $\boldsymbol{U}_{1,2}$ a row vector of length N, $\boldsymbol{0}$ a column vector of length N filled with zeros and $\boldsymbol{U}_{2,2}$ an upper triangular $N \times N$ matrix. Recall from the definition of $\boldsymbol{U}_{t,l}$ and equation B.15 that $\boldsymbol{U}^T_{t,l}\boldsymbol{U}_{t,l} = \boldsymbol{M}^T_{t,l}\boldsymbol{M}_{t,l}$. Multiplying out both sides of this equality yields:

$$\begin{bmatrix} r_{t,l,l} + \boldsymbol{f}^{*T}_{t,l}\boldsymbol{f}^{*}_{t,l} & \boldsymbol{f}^{*T}_{t,l}\boldsymbol{S}^T_{t|y^-_{t,l}} \\ \boldsymbol{S}_{t|y^-_{t,l}}\boldsymbol{f}^{*}_{t,l} & \boldsymbol{S}_{t|y^-_{t,l}}\boldsymbol{S}^T_{t|y^-_{t,l}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}^2_{1,1} & \boldsymbol{U}_{1,1}\boldsymbol{U}_{1,2} \\ \boldsymbol{U}^T_{1,2}\boldsymbol{U}_{1,1} & \boldsymbol{U}^T_{1,2}\boldsymbol{U}_{1,2} + \boldsymbol{U}^T_{2,2}\boldsymbol{U}_{2,2} \end{bmatrix} \tag{B.20}$$

It is obvious from equation B.6 and B.16 that $\boldsymbol{U}_{1,1} = \sqrt{\sigma_{t,l}}$. Using this and noting that $\boldsymbol{f}^{*T}_{t,l}\boldsymbol{S}^T_{t|y^-_{t,l}} = \boldsymbol{f}^T_{t,l}$, where $\boldsymbol{f}_{t,l}$ is defined as in equation B.5, one obtains that:

$$\boldsymbol{U}_{1,2} = \frac{\boldsymbol{f}^T_{t,l}}{\sqrt{\sigma_{t,l}}} \tag{B.21}$$

It remains to show that $\boldsymbol{U}_{2,2} = \boldsymbol{S}^T_{t|y_{t,l}}$. By noting that the the bottom right element of the left hand side of equation B.20 is, by definition, equal to the pre-update covariance matrix $\boldsymbol{P}_{t|y^-_{t,l}}$ and plugging in the value for $\boldsymbol{U}_{1,2}$, one obtains that:

$$\boldsymbol{U}^T_{2,2}\boldsymbol{U}_{2,2} = \boldsymbol{P}_{t|y^-_{t,l}} - \frac{1}{\sigma_{t,l}}\boldsymbol{f}_{t,l}\boldsymbol{f}^T_{t,l} = \boldsymbol{P}_{t|y_{t,l}} \tag{B.22}$$

where the last equality comes comes from equation B.9. Thus $\boldsymbol{U}^T_{2,2}$ is a matrix square root of $\boldsymbol{P}_{t|y_{t,l}}$ and by the definition of the QR decomposition it is lower triangular, which completes the proof. Importantly, no part of the proof requires the lower triangular square roots of $\boldsymbol{P}_{t|y^-_{t,l}}$ or $\boldsymbol{P}_{t|y_{t,l}}$ to be unique or makes reference to a specific type of matrix square root.

B.3.5 The Predict Step of the Square-Root Kalman Filter. For the square root implementation of the unscented predict step in period t, firstly the sigma points are calculated as in equation B.10, where this time $\boldsymbol{S}_{t|t}$ is required to be a lower triangular matrix square root of $\boldsymbol{P}_{t|t}$. Again, $\tilde{\mathcal{X}}_t$ denotes the $(2N+1) \times N$ matrix of the transformed sigma points. The calculation of the predicted mean of the state vector remains the same as before (equation B.12).

Define $\boldsymbol{A}_t$ as stacked matrix of of weighted deviations of the sigma points from the predicted mean and the covariance matrix of the transition shocks:

$$A_t \equiv \begin{bmatrix} \sqrt{w_{t,0}}(\tilde{\mathscr{X}}_{t,0} - \bar{x}_{t+1|t})^T \\ \cdots \\ \sqrt{w_{t,2n}}(\tilde{\mathscr{X}}_{t,2n} - \bar{x}_{t+1|t})^T \\ \sqrt{Q_t} \end{bmatrix} \tag{B.23}$$

Then equation B.13 can be rewritten as:

$$P_{t+1|t} = A_t^T A_t \tag{B.24}$$

and by the relation of the QR decomposition and and the lower triangular matrix square root (equation B.15) a lower triangular matrix square root of $P_{t+1|t}$ is given by $qr(A_t)^T$.

Appendix C Detailed Model Setup

C.1 Background on Identification

Cunha, Heckman, and Schennach (2010) provide very general nonparametric Identification result for nonlinear dynamic latent factor models. The exact conditions for identification depend on the assumptions one is willing to put on the measurement error. However, having at least two dedicated measurements for each latent factor in each period is sufficient to identify an arbitrary production function under mild conditions. Since latent factors do not have a natural unit of measurement, the identification requires normalizations of location and scale. Thus, Cunha, Heckman, and Schennach (2010) normalize one loading of each factor in each period to 1 and one intercept of each factor in each period to 0. While the identification result works for arbitrary production functions, they use a parametric CES function in their empirical application.

Agostinelli and Wiswall (2016) criticize the identification result by Cunha, Heckman, and Schennach (2010) to be flawed. They point out that the CES production function already puts a restriction on the scale and location of its output. Thus, normalization of scale and location are only required in the first period and re-normalizations in each period are actually not normalizations but testable assumptions. Moreover, they show that under the implicit restrictions imposed by the CES production function, identification under a linear measurement system can be achieved with as little as one measurement per latent factor and period as long as there are at least two measurements in the first period.

Freyberger (2024) shows that the CES production function also imposes implicit restrictions on the relative scale of the latent factors and thus identification can be achieved if only the location and scale of a single factor are normalized in the first period.

While the critique by Agostinelli and Wiswall (2016) that over-normalizations are detrimental is correct, it mostly applies to the empirical application and not the general identification result in Cunha, Heckman, and Schennach (2010) nor the maximum likelihood estimator used in the paper. The identification result states that latent factors have no natural scale and location that could be be identified from data and thus their location and scale has to be fixed by restrictions imposed by the econometrician. Cunha, Heckman, and Schennach (2010) restrict factor loadings and intercepts but mention, that instead of factor loadings, the variances of measurement errors could be restricted. Of course, these restrictions are mutually exclusive and it would not be valid to restrict factor loadings and variances of measurement error at the same time. The main contribution of Agostinelli and Wiswall (2016) is to point out that using restrictive functional forms for the production function is yet another way of fixing the location and scale of the latent factors.

Appendix D Additional Tables and Figures for the Main Specification

D.1 Complete Set of Parameters of the Measurement System

Table D.1. Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Females

Age	Measurement	Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.745***
				(0.002)
	Mobility	−0.055***	1.337***	0.713***
		(0.004)	(0.007)	(0.003)
	Large Muscle Index	0.050***	1.001***	0.729***
		(0.004)	(0.006)	(0.003)
	Self-Reported Health	0.034***	1.015***	0.753***
		(0.004)	(0.006)	(0.002)
70	Alive	0.969***	0.068***	0.179***
		(0.029)	(0.006)	(0.006)
	Grip Strength	−0.147***	0.482***	0.886***
		(0.024)	(0.040)	(0.015)
72	Alive	0.966***	0.067***	0.194***
		(0.037)	(0.006)	(0.008)
	Grip Strength	−0.251***	0.413***	0.925***
		(0.028)	(0.042)	(0.015)
74	Alive	0.969***	0.060***	0.190***
		(0.041)	(0.006)	(0.008)
	Grip Strength	−0.337***	0.533***	0.884***
		(0.027)	(0.044)	(0.015)
76	Alive	0.955***	0.093***	0.233***
		(0.040)	(0.009)	(0.010)
	Grip Strength	−0.478***	0.411***	0.911***
		(0.029)	(0.047)	(0.012)
78	Alive	0.951***	0.088***	0.246***
		(0.048)	(0.010)	(0.013)
	Grip Strength	−0.574***	0.420***	0.875***
		(0.030)	(0.048)	(0.018)
80	Alive	0.949***	0.109***	0.262***
		(0.046)	(0.013)	(0.013)
	Grip Strength	−0.747***	0.375***	0.847***
		(0.032)	(0.050)	(0.019)
82	Alive	0.940***	0.109***	0.286***
		(0.057)	(0.016)	(0.019)
	Grip Strength	−0.790***	0.363***	0.846***
		(0.035)	(0.057)	(0.021)
84	Alive	0.933***	0.153***	0.317***
		(0.058)	(0.022)	(0.021)
	Grip Strength	−0.960***	0.353***	0.826***
		(0.037)	(0.059)	(0.023)
86	Alive	0.925***	0.154***	0.337***
		(0.074)	(0.029)	(0.029)
	Grip Strength	−0.994***	0.368***	0.836***
		(0.042)	(0.072)	(0.026)
88	Alive	0.905***	0.187***	0.377***
		(0.091)	(0.044)	(0.042)
	Grip Strength	−1.140***	0.511***	0.787***
		(0.054)	(0.083)	(0.032)
90	Alive	0.906***	0.233***	0.400***
		(0.116)	(0.066)	(0.056)
	Grip Strength	−1.147***	0.349***	0.717***
		(0.050)	(0.084)	(0.028)
92	Alive	0.855***	0.243**	0.440***
		(0.167)	(0.118)	(0.101)
	Grip Strength	−1.289***	0.393***	0.732***
		(0.077)	(0.128)	(0.047)
Note:			***p<0.01;**p<0.05;*p<0.1	

Table D.2. Intercepts, Loadings, and Measurement Standard Deviations for Physical Capacity, Males

Age	Measurement	Intercept	Loading	Meas. Std.
All	Frailty Index (Reversed)	0.000	1.000	0.804***
				(0.002)
	Mobility	0.074***	1.500***	0.721***
		(0.006)	(0.009)	(0.003)
	Large Muscle Index	0.107***	1.134***	0.763***
		(0.005)	(0.009)	(0.003)
	Self-Reported Health	0.121***	1.040***	0.790***
		(0.005)	(0.008)	(0.003)
70	Alive	0.966***	0.091***	0.202***
		(0.033)	(0.008)	(0.007)
	Grip Strength	−0.026	0.641***	0.958***
		(0.032)	(0.056)	(0.018)
72	Alive	0.959***	0.097***	0.227***
		(0.043)	(0.010)	(0.011)
	Grip Strength	−0.264***	0.645***	0.933***
		(0.033)	(0.055)	(0.020)
74	Alive	0.956***	0.091***	0.238***
		(0.054)	(0.011)	(0.014)
	Grip Strength	−0.312***	0.537***	0.888***
		(0.033)	(0.057)	(0.019)
76	Alive	0.955***	0.128***	0.259***
		(0.047)	(0.014)	(0.013)
	Grip Strength	−0.458***	0.668***	0.877***
		(0.034)	(0.058)	(0.019)
78	Alive	0.953***	0.125***	0.272***
		(0.053)	(0.015)	(0.016)
	Grip Strength	−0.567***	0.565***	0.902***
		(0.039)	(0.062)	(0.022)
80	Alive	0.942***	0.138***	0.301***
		(0.066)	(0.021)	(0.022)
	Grip Strength	−0.715***	0.650***	0.880***
		(0.041)	(0.065)	(0.022)
82	Alive	0.934***	0.174***	0.332***
		(0.068)	(0.029)	(0.026)
	Grip Strength	−0.910***	0.538***	0.838***
		(0.043)	(0.065)	(0.025)
84	Alive	0.922***	0.160***	0.349***
		(0.092)	(0.036)	(0.037)
	Grip Strength	−0.964***	0.573***	0.841***
		(0.051)	(0.072)	(0.025)
86	Alive	0.890***	0.186***	0.392***
		(0.128)	(0.060)	(0.064)
	Grip Strength	−1.193***	0.518***	0.839***
		(0.061)	(0.090)	(0.034)
88	Alive	0.924***	0.244***	0.394***
		(0.128)	(0.074)	(0.058)
	Grip Strength	−1.160***	0.622***	0.795***
		(0.068)	(0.099)	(0.040)
90	Alive	0.891***	0.271*	0.424***
		(0.203)	(0.142)	(0.102)
	Grip Strength	−1.434***	0.397***	0.753***
		(0.085)	(0.113)	(0.052)
92	Alive	0.828***	0.287	0.445***
		(0.231)	(0.204)	(0.150)
	Grip Strength	−1.441***	0.740***	0.796***
		(0.123)	(0.186)	(0.072)

Note: ***p<0.01; **p<0.05; *p<0.1

Table D.3. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Females

Age	Measurement	Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.903***
				(0.004)
	Vocabulary	0.048***	0.861***	0.929***
		(0.010)	(0.022)	(0.007)
	Immediate Word Recall	−0.188***	1.816***	0.585***
		(0.008)	(0.018)	(0.003)
	Delayed Word Recall	−0.202***	1.836***	0.579***
		(0.008)	(0.018)	(0.003)
70	Self-Rated Memory	−0.028	0.633***	0.950***
		(0.018)	(0.038)	(0.011)
72	Self-Rated Memory	−0.034*	0.602***	0.949***
		(0.018)	(0.038)	(0.011)
74	Self-Rated Memory	−0.044**	0.626***	0.964***
		(0.019)	(0.038)	(0.012)
76	Self-Rated Memory	−0.023	0.573***	0.964***
		(0.020)	(0.040)	(0.012)
78	Self-Rated Memory	0.035	0.499***	0.964***
		(0.023)	(0.043)	(0.013)
80	Self-Rated Memory	0.053**	0.428***	1.006***
		(0.026)	(0.046)	(0.015)
82	Self-Rated Memory	0.082***	0.508***	0.998***
		(0.031)	(0.053)	(0.016)
84	Self-Rated Memory	0.110***	0.443***	1.018***
		(0.037)	(0.061)	(0.018)
86	Self-Rated Memory	0.099**	0.347***	1.028***
		(0.044)	(0.066)	(0.021)
88	Self-Rated Memory	0.201***	0.447***	1.029***
		(0.057)	(0.080)	(0.024)
90	Self-Rated Memory	0.173**	0.377***	1.054***
		(0.073)	(0.104)	(0.029)
92	Self-Rated Memory	0.107	0.392***	1.151***
		(0.112)	(0.149)	(0.042)
Note:			***p<0.01;**p<0.05;*p<0.1	

Table D.4. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Capacity, Males

Age	Measurement	Intercept	Loading	Meas. Std.
All	Serial 7 Subtraction	0.000	1.000	0.906***
				(0.004)
	Vocabulary	0.024**	0.971***	0.868***
		(0.012)	(0.026)	(0.008)
	Immediate Word Recall	−0.194***	1.750***	0.595***
		(0.010)	(0.021)	(0.004)
	Delayed Word Recall	−0.185***	1.718***	0.586***
		(0.010)	(0.020)	(0.003)
70	Self-Rated Memory	−0.067***	0.675***	0.943***
		(0.021)	(0.045)	(0.013)
72	Self-Rated Memory	−0.071***	0.633***	0.948***
		(0.022)	(0.044)	(0.014)
74	Self-Rated Memory	−0.075***	0.590***	0.935***
		(0.022)	(0.045)	(0.014)
76	Self-Rated Memory	−0.047**	0.542***	0.956***
		(0.024)	(0.049)	(0.015)
78	Self-Rated Memory	−0.061**	0.611***	0.975***
		(0.027)	(0.054)	(0.016)
80	Self-Rated Memory	−0.030	0.557***	0.983***
		(0.030)	(0.058)	(0.017)
82	Self-Rated Memory	−0.020	0.395***	1.006***
		(0.037)	(0.064)	(0.020)
84	Self-Rated Memory	−0.011	0.406***	1.003***
		(0.044)	(0.073)	(0.023)
86	Self-Rated Memory	−0.032	0.524***	1.027***
		(0.056)	(0.090)	(0.028)
88	Self-Rated Memory	0.037	0.509***	1.038***
		(0.074)	(0.113)	(0.032)
90	Self-Rated Memory	0.038	0.484***	1.081***
		(0.096)	(0.144)	(0.046)
92	Self-Rated Memory	0.073	0.425**	0.983***
		(0.132)	(0.196)	(0.055)
Note:			***p<0.01;**p<0.05;*p<0.1	

Table D.5. Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Females

Age	Measurement	Intercept	Loading	Meas. Std.
All	Vigorous Activity	−0.005	0.695^{***}	0.802^{***}
		(0.005)	(0.010)	(0.004)
	Moderate Activity	0.000	1.000	0.796^{***}
				(0.004)
	Light Activity	-0.137^{***}	1.068^{***}	0.934^{***}
		(0.006)	(0.012)	(0.004)
Note:			$^{***}p<0.01;^{**}p<0.05;^{*}p<0.1$	

Table D.6. Intercepts, Loadings, and Measurement Standard Deviations for Exercise, Males

Age	Measurement	Intercept	Loading	Meas. Std.
All	Vigorous Activity	-0.014^{**}	0.736^{***}	0.813^{***}
		(0.006)	(0.012)	(0.005)
	Moderate Activity	0.000	1.000	0.811^{***}
				(0.004)
	Light Activity	-0.076^{***}	0.923^{***}	0.860^{***}
		(0.007)	(0.012)	(0.004)
Note:			$^{***}p<0.01;^{**}p<0.05;^{*}p<0.1$	

Table D.7. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Females

Age	Measurement	Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.826***
				(0.006)
	Listening to Music	−0.142***	0.548***	0.939***
		(0.006)	(0.011)	(0.006)
	Stimulating Hobbies	−0.036***	0.637***	0.895***
		(0.008)	(0.013)	(0.005)
	Communication	−0.064***	0.582***	0.968***
		(0.007)	(0.011)	(0.006)
Note:			***p<0.01;**p<0.05;*p<0.1	

Table D.8. Intercepts, Loadings, and Measurement Standard Deviations for Cognitive Stimulation, Males

Age	Measurement	Intercept	Loading	Meas. Std.
All	Reading	0.000	1.000	0.667***
				(0.008)
	Listening to Music	−0.200***	0.208***	1.050***
		(0.008)	(0.011)	(0.008)
	Stimulating Hobbies	0.018*	0.347***	0.999***
		(0.010)	(0.012)	(0.006)
	Communication	−0.105***	0.306***	0.990***
		(0.007)	(0.011)	(0.007)
Note:			***p<0.01;**p<0.05;*p<0.1	

D.2 Correlations between measurements and factors

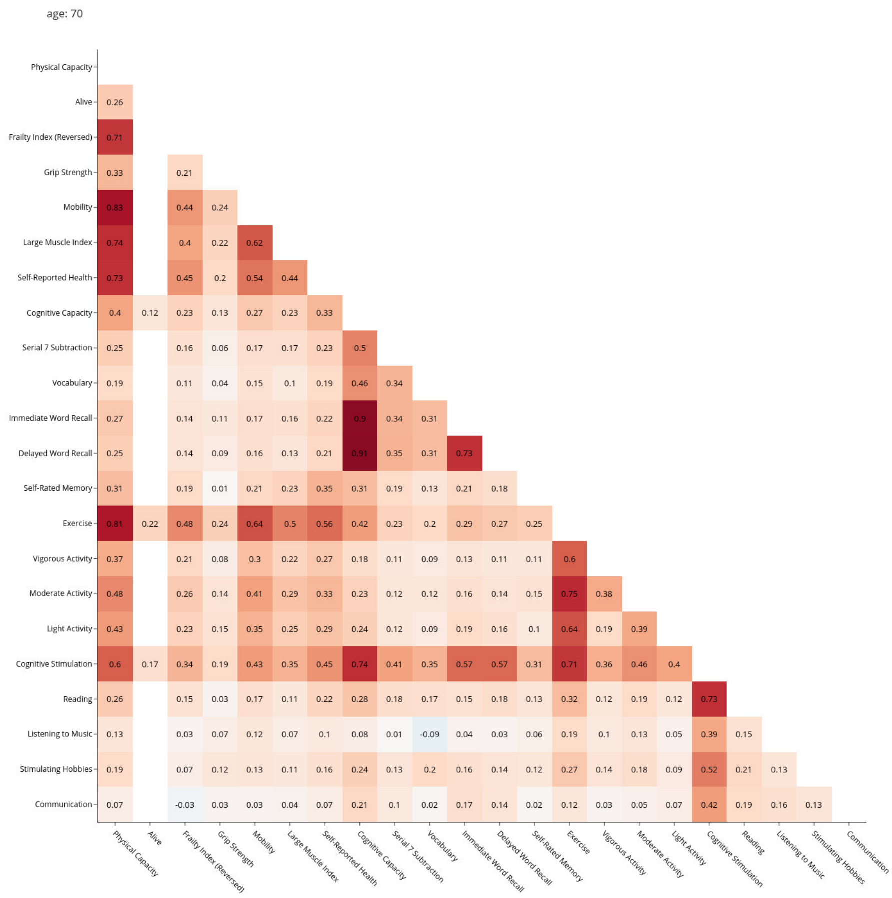


Figure D.1. Correlations across implied factors and measurement correlations – females aged 70

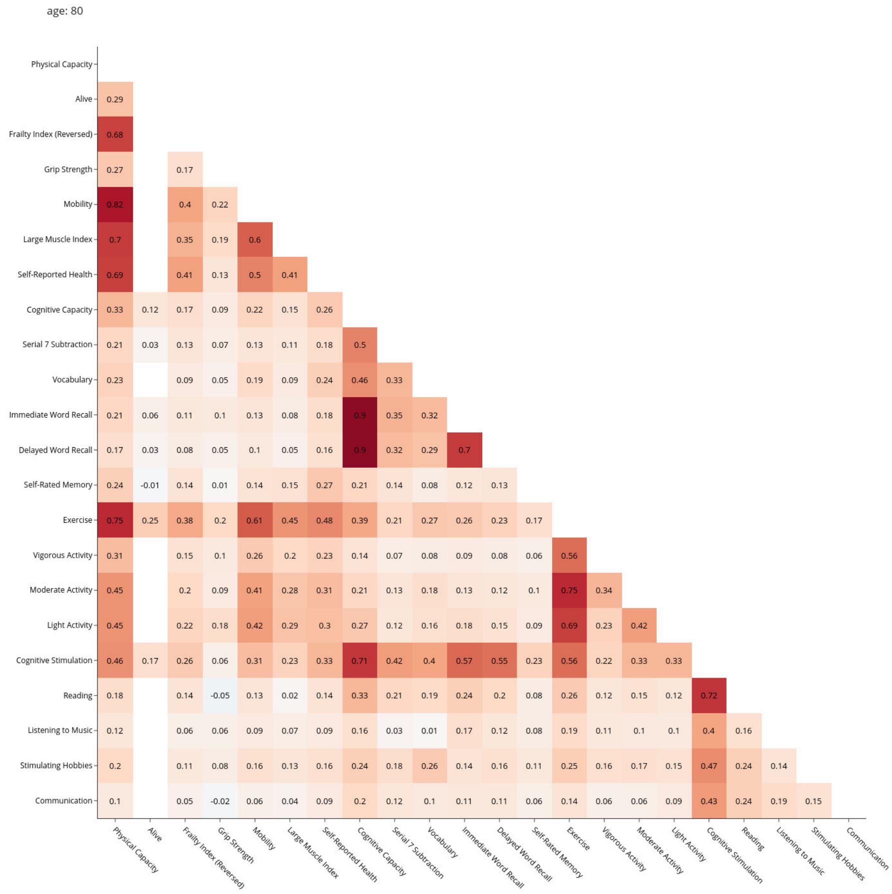


Figure D.2. Correlations across implied factors and measurement correlations – females aged 80

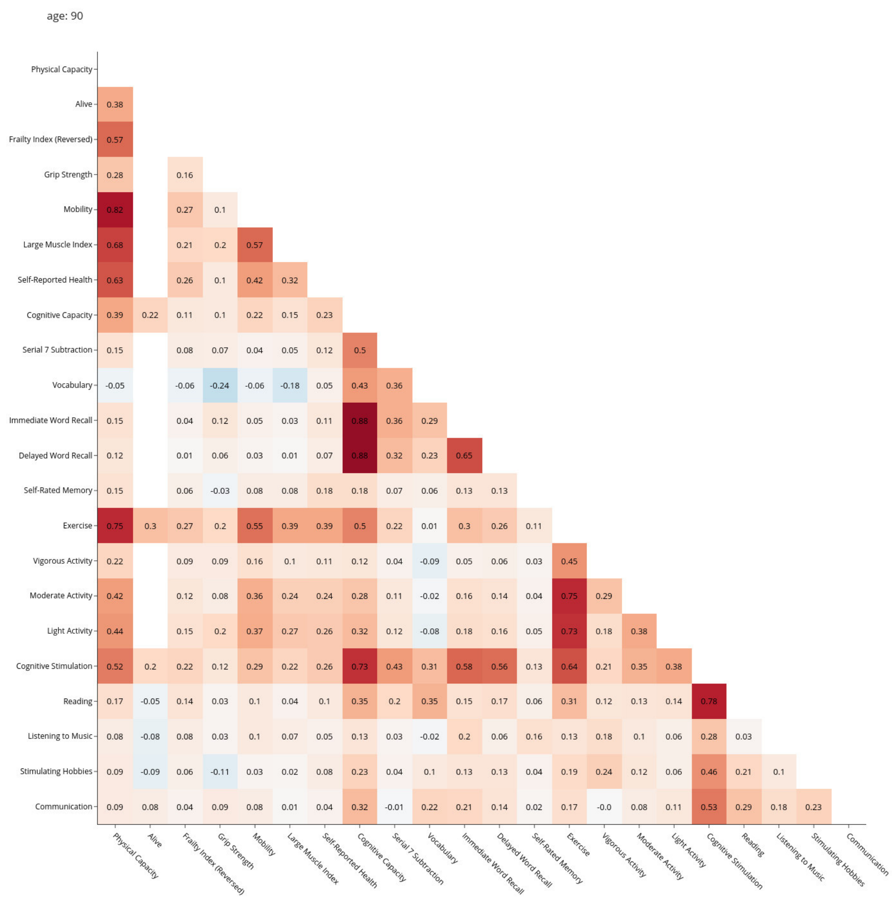


Figure D.3. Correlations across implied factors and measurement correlations – females aged 90

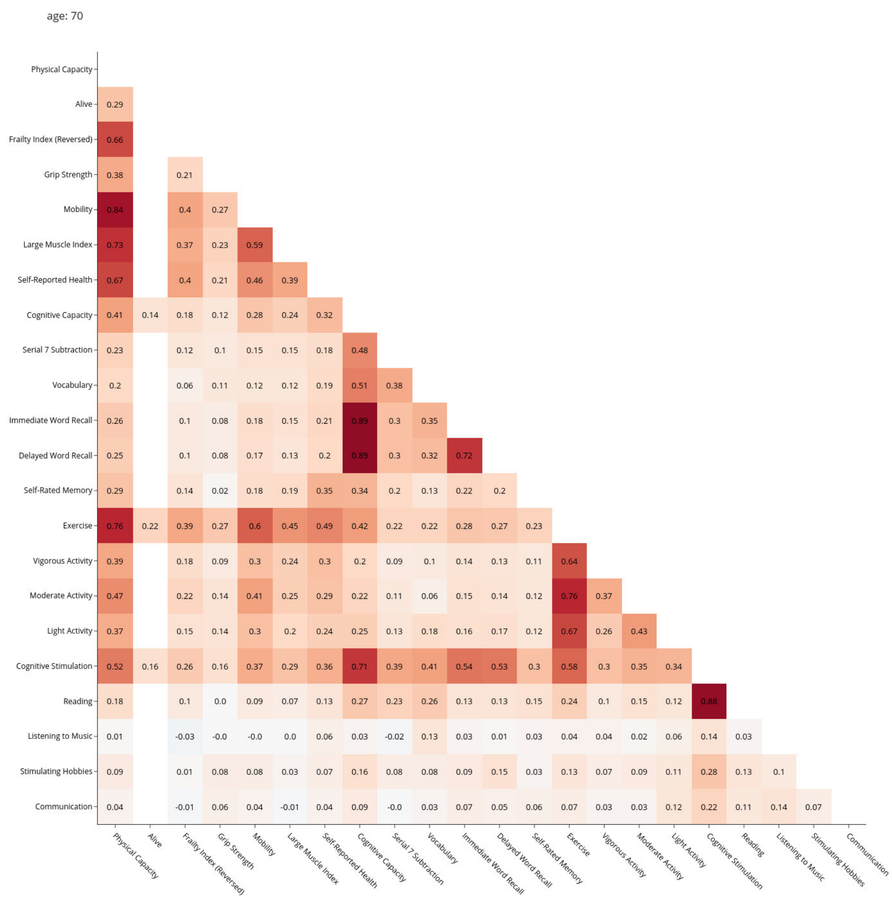


Figure D.4. Correlations across implied factors and measurement correlations – males aged 70

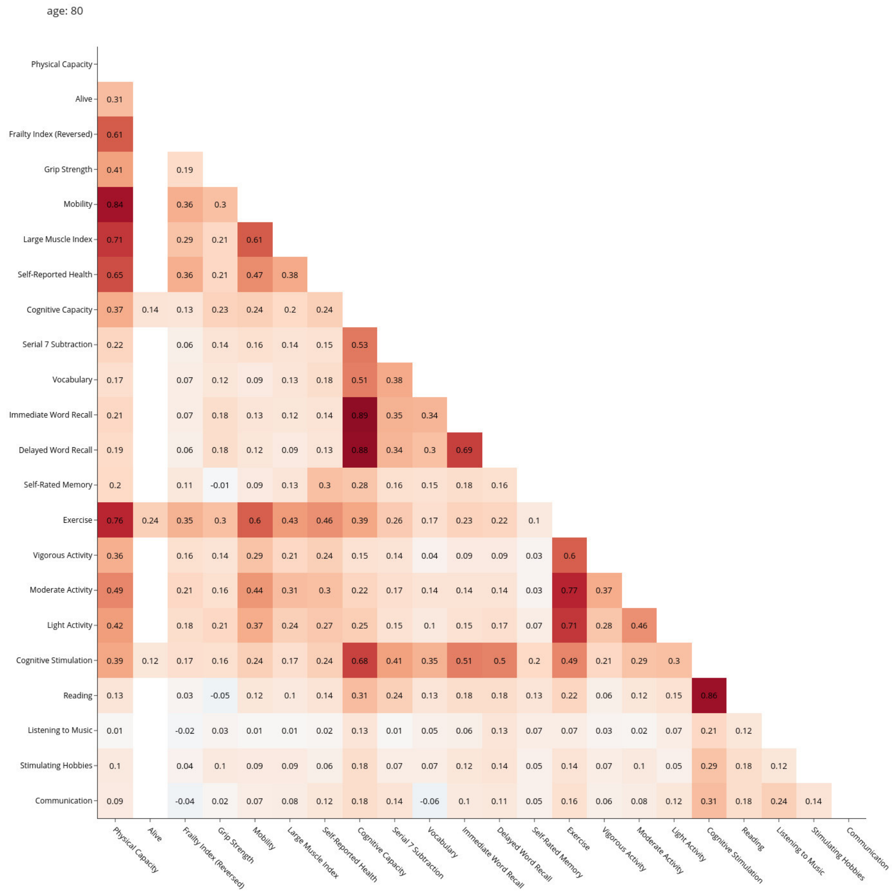


Figure D.5. Correlations across implied factors and measurement correlations – males aged 80

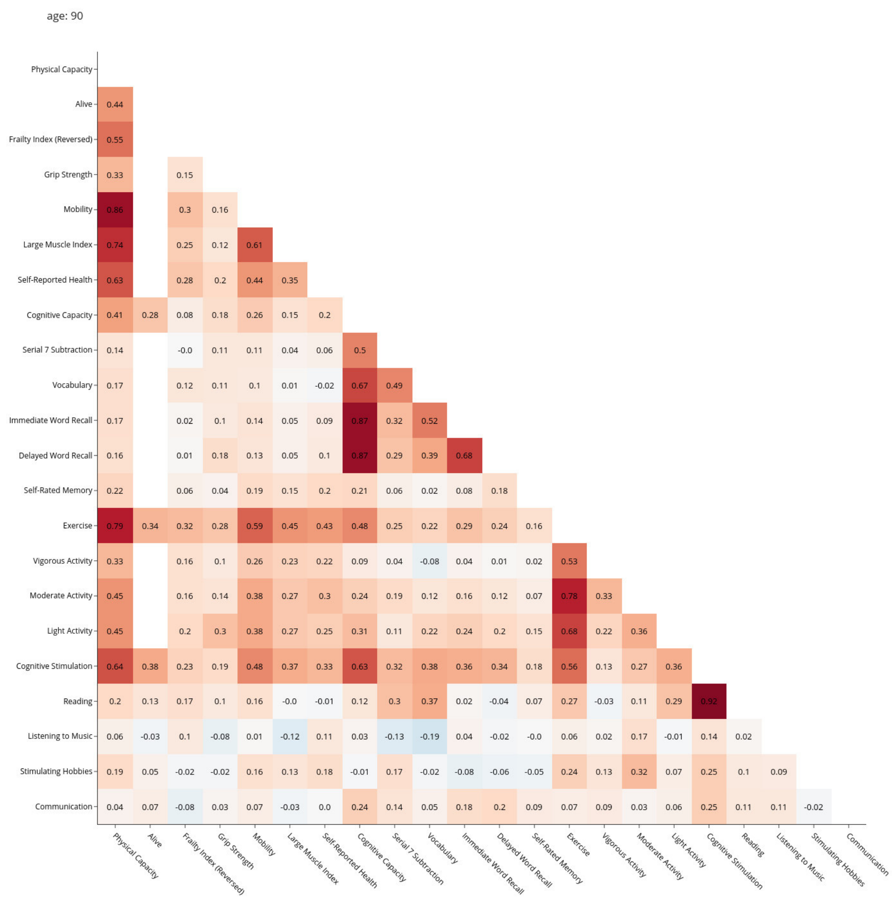


Figure D.6. Correlations across implied factors and measurement correlations – males aged 90

D.3 Factor distributions

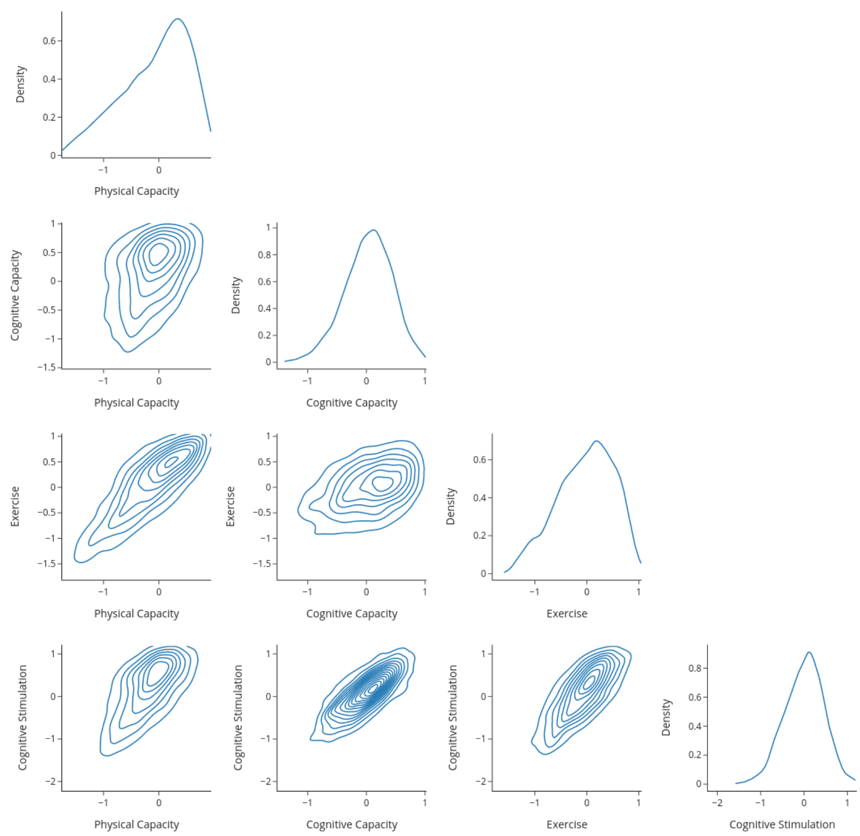


Figure D.7. Factor distributions – females aged 70

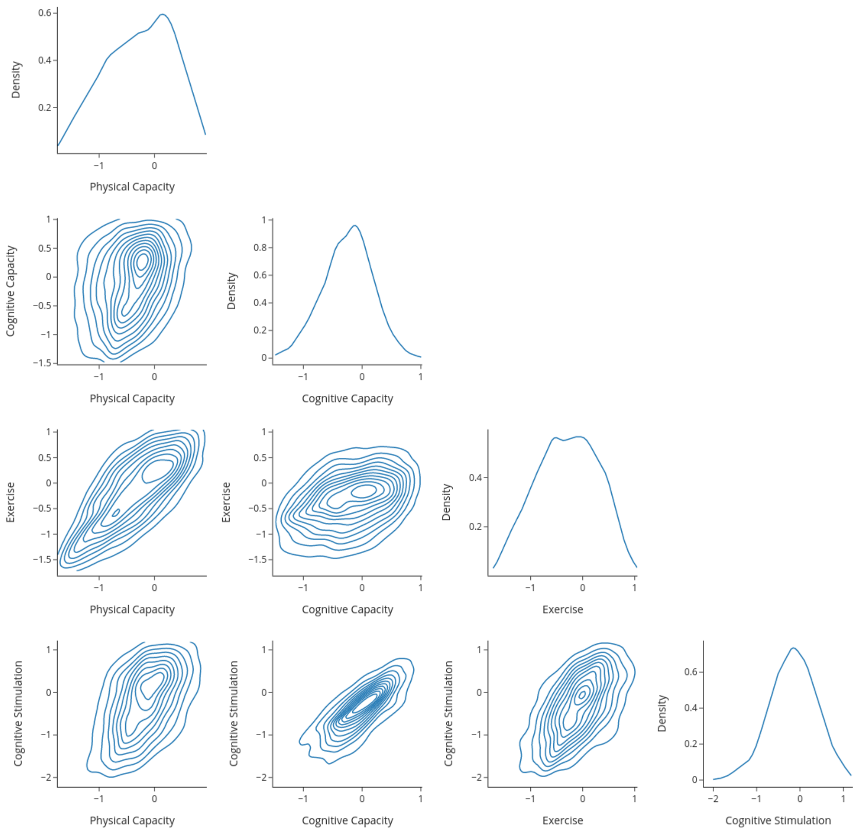


Figure D.8. Factor distributions – females aged 80

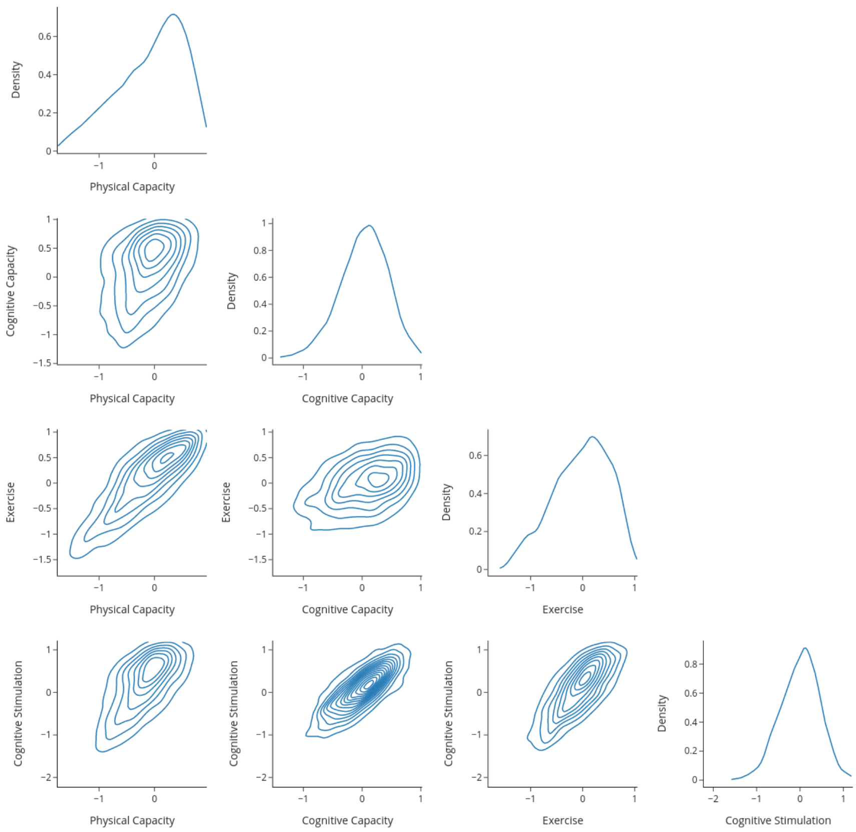


Figure D.9. Factor distributions – females aged 90

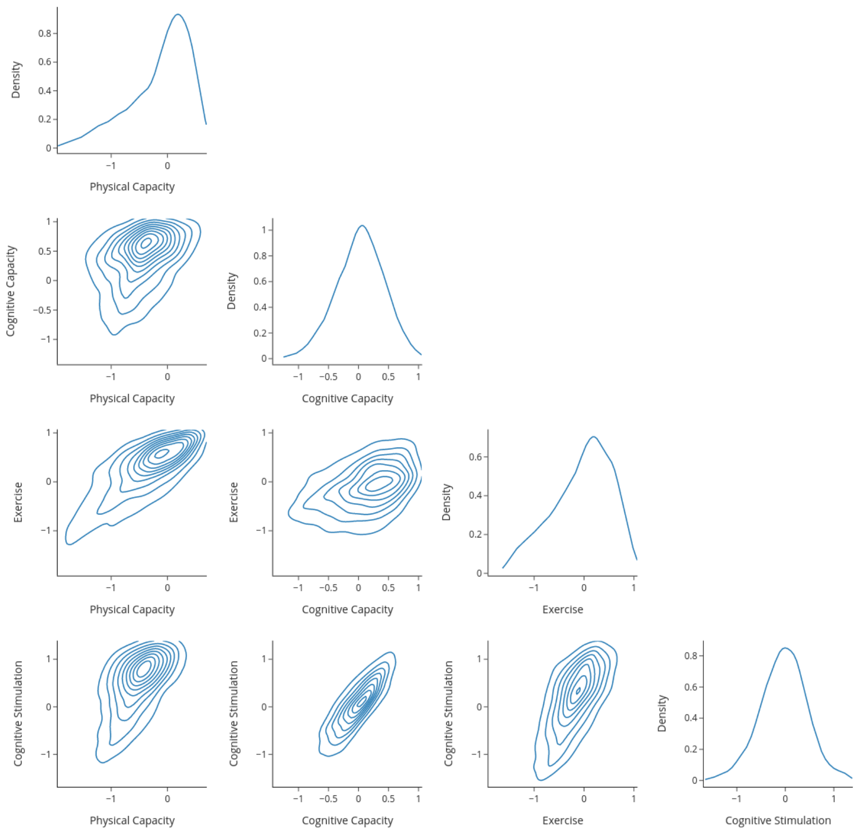


Figure D.10. Factor distributions – males aged 70

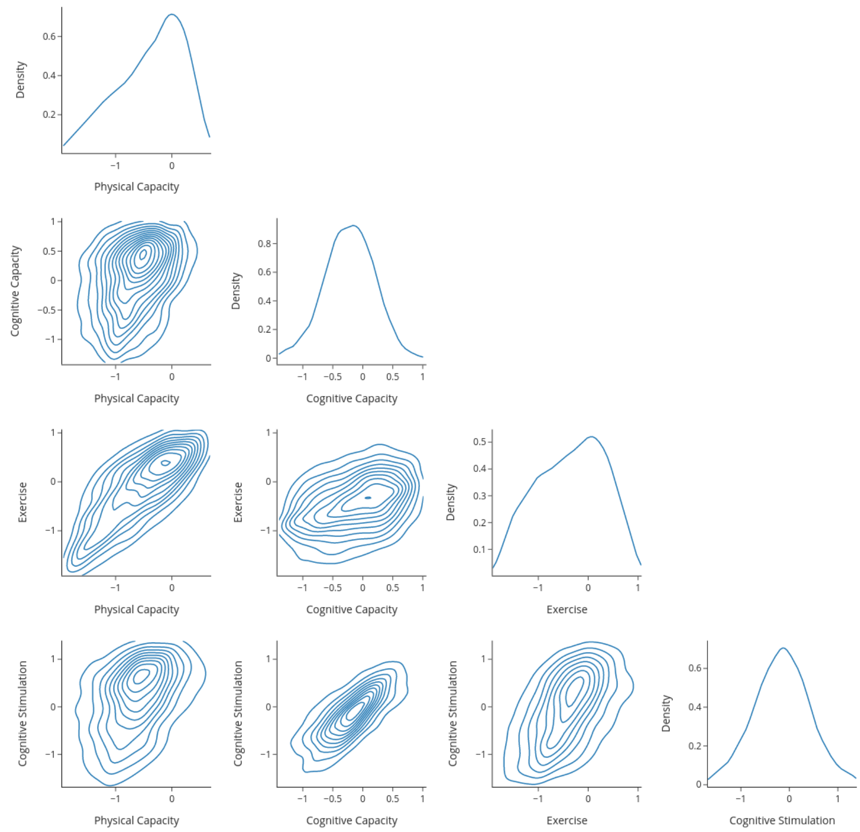


Figure D.11. Factor distributions – males aged 80

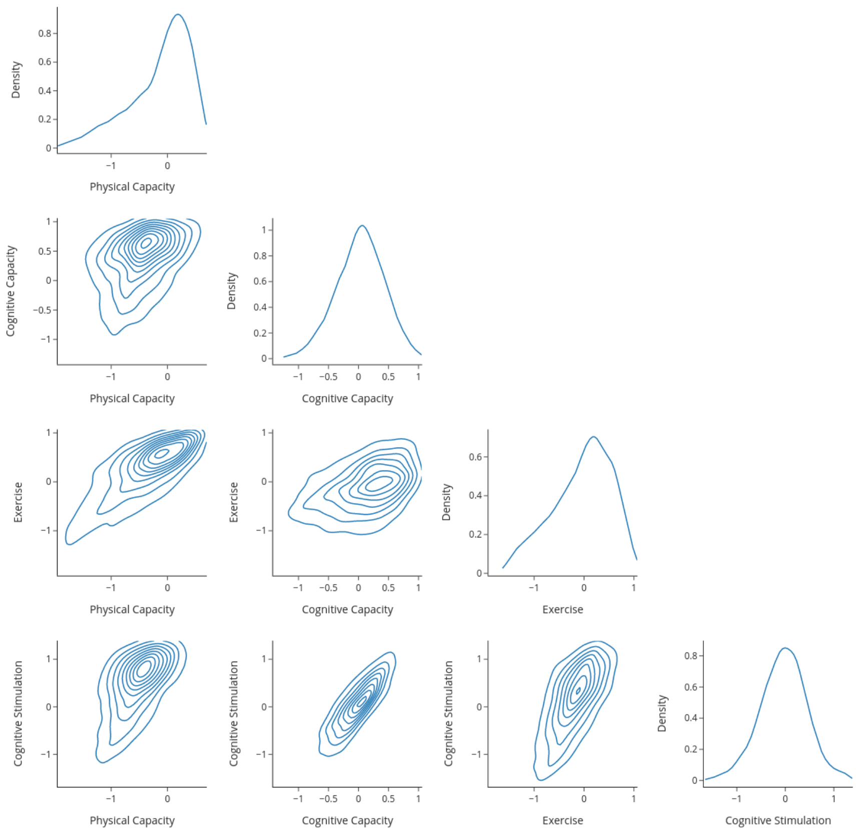


Figure D.12. Factor distributions – males aged 90

D.4 Transition Equations

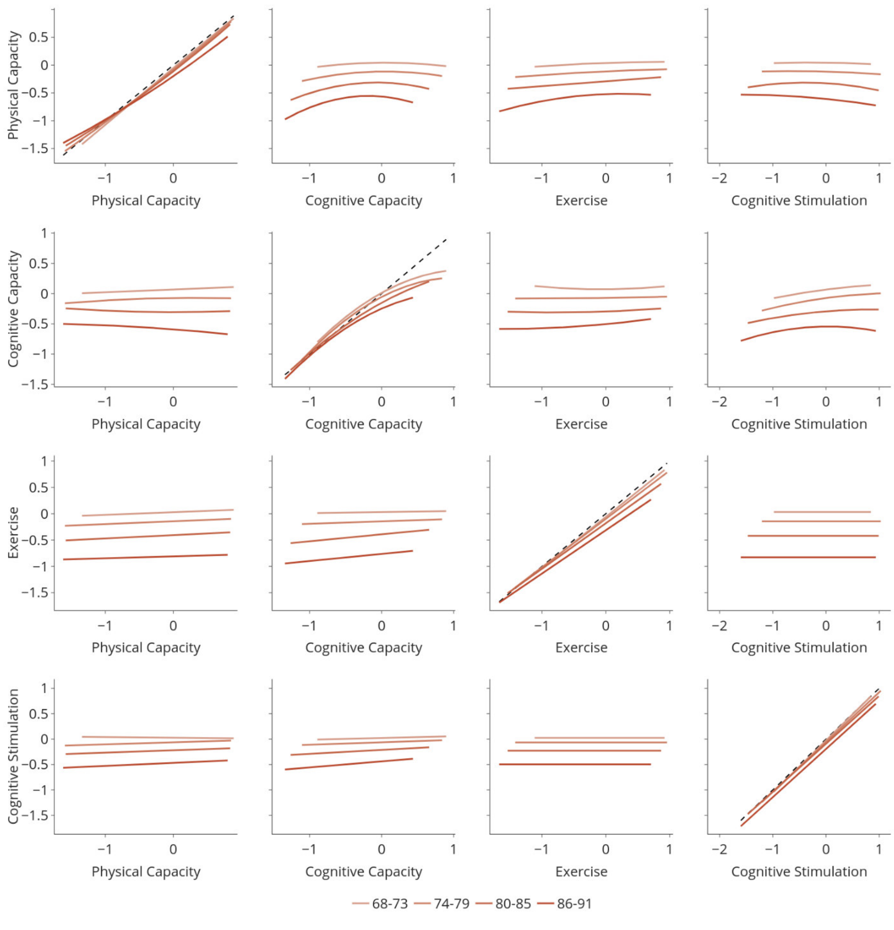


Figure D.13. Transition equations for all factors (other factors evaluated at the median), females

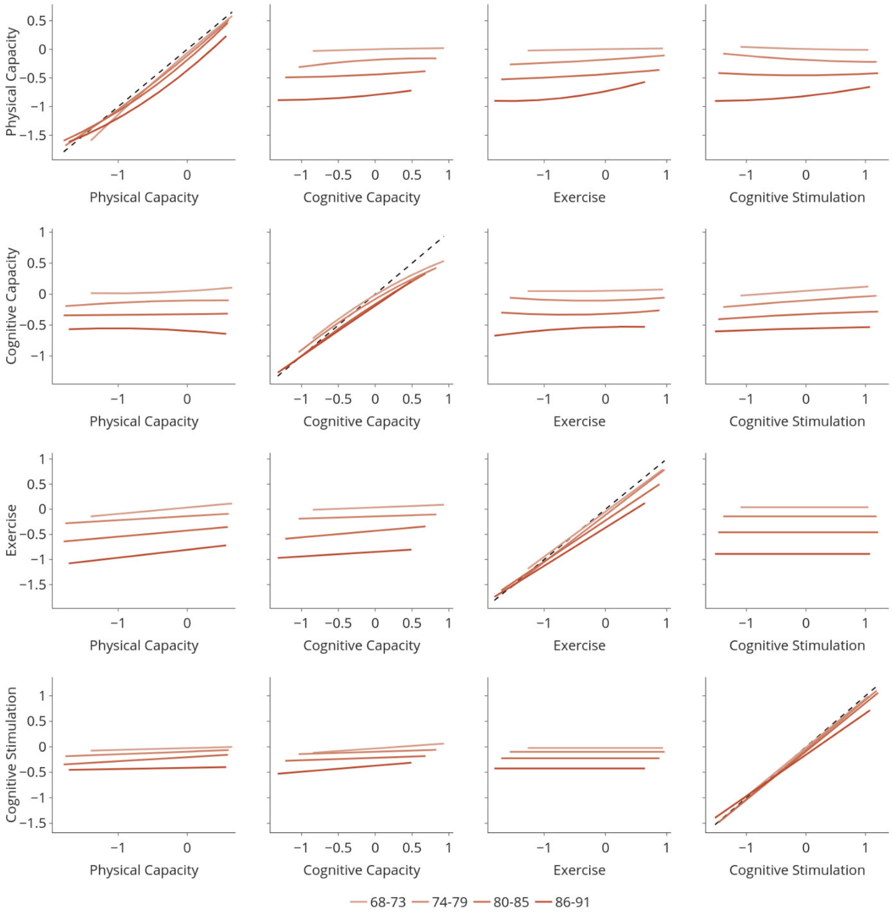


Figure D.14. Transition equations for all factors (other factors evaluated at the median), males

Table D.9. Transition Parameters for Physical Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	1.010***	0.989***	0.960***	0.973***
	(0.008)	(0.010)	(0.015)	(0.038)
Cognitive Capacity	0.003	0.008	−0.018	−0.110*
	(0.009)	(0.012)	(0.022)	(0.063)
Exercise	0.043***	0.050***	0.063***	0.002
	(0.007)	(0.009)	(0.015)	(0.035)
Cognitive Stimulation	−0.024***	−0.021**	0.016	0.033
	(0.008)	(0.009)	(0.016)	(0.035)
Physical Capacity Squared	−0.006	0.044***	0.068***	0.063*
	(0.012)	(0.014)	(0.019)	(0.034)
Cognitive Capacity Squared	−0.085***	−0.129***	−0.220***	−0.315***
	(0.017)	(0.020)	(0.031)	(0.067)
Exercise Squared	−0.019	−0.013	0.001	−0.086**
	(0.014)	(0.015)	(0.018)	(0.037)
Cognitive Stimulation Squared	−0.022	−0.021	−0.073***	−0.030
	(0.014)	(0.014)	(0.019)	(0.029)
Physical Capacity × Cognitive Capacity	0.062***	0.058**	0.083**	0.212***
	(0.021)	(0.023)	(0.034)	(0.069)
Physical Capacity × Exercise	0.004	0.000	−0.018	0.103*
	(0.020)	(0.022)	(0.027)	(0.052)
Physical Capacity × Cognitive Stimulation	−0.051***	−0.021	−0.042*	−0.109**
	(0.018)	(0.019)	(0.025)	(0.048)
Cognitive Capacity × Exercise	−0.078***	−0.134***	−0.161***	−0.266***
	(0.023)	(0.025)	(0.035)	(0.066)
Cognitive Capacity × Cognitive Stimulation	0.124***	0.148***	0.249***	0.224***
	(0.023)	(0.027)	(0.037)	(0.069)
Exercise × Cognitive Stimulation	0.047**	0.051**	0.099***	0.118**
	(0.023)	(0.021)	(0.028)	(0.049)
Constant	−0.055***	−0.083***	−0.087***	−0.112***
	(0.006)	(0.008)	(0.011)	(0.024)
Note:				***p<0.01;**p<0.05;*p<0.1

Table D.10. Transition Parameters for Physical Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	1.020***	1.000***	1.020***	0.996***
	(0.011)	(0.014)	(0.026)	(0.065)
Cognitive Capacity	0.035***	0.070***	0.059*	0.084
	(0.011)	(0.014)	(0.031)	(0.087)
Exercise	0.019**	0.062***	0.054***	0.124**
	(0.007)	(0.009)	(0.018)	(0.054)
Cognitive Stimulation	−0.028***	−0.048***	−0.005	−0.028
	(0.008)	(0.010)	(0.020)	(0.049)
Physical Capacity Squared	−0.054***	0.064***	0.135***	0.147**
	(0.015)	(0.018)	(0.030)	(0.060)
Cognitive Capacity Squared	−0.012	−0.053**	0.024	0.049
	(0.019)	(0.023)	(0.038)	(0.092)
Exercise Squared	−0.005	0.009	0.012	0.066
	(0.011)	(0.013)	(0.021)	(0.048)
Cognitive Stimulation Squared	0.008	0.020*	0.021	0.036
	(0.010)	(0.011)	(0.017)	(0.034)
Physical Capacity × Cognitive Capacity	−0.026	−0.008	0.030	0.102
	(0.025)	(0.029)	(0.050)	(0.108)
Physical Capacity × Exercise	0.006	−0.068***	−0.058	−0.086
	(0.019)	(0.022)	(0.040)	(0.077)
Physical Capacity × Cognitive Stimulation	0.015	0.056***	−0.010	0.084
	(0.018)	(0.020)	(0.035)	(0.065)
Cognitive Capacity × Exercise	−0.040*	−0.037	−0.031	−0.026
	(0.021)	(0.024)	(0.043)	(0.093)
Cognitive Capacity × Cognitive Stimulation	0.002	0.019	−0.061	−0.258***
	(0.022)	(0.025)	(0.039)	(0.089)
Exercise × Cognitive Stimulation	0.025	0.002	0.022	−0.093
	(0.016)	(0.016)	(0.031)	(0.062)
Constant	−0.066***	−0.112***	−0.160***	−0.231***
	(0.007)	(0.009)	(0.015)	(0.038)

Note: ***p<0.01;**p<0.05;*p<0.1

Table D.11. Transition Parameters for Cognitive Capacity, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.040***	0.021**	0.003	−0.013
	(0.010)	(0.011)	(0.015)	(0.027)
Cognitive Capacity	0.646***	0.614***	0.696***	0.599***
	(0.012)	(0.013)	(0.020)	(0.053)
Exercise	0.018	0.013	0.029*	0.056*
	(0.011)	(0.011)	(0.015)	(0.029)
Cognitive Stimulation	0.097***	0.127***	0.094***	0.166***
	(0.012)	(0.011)	(0.015)	(0.031)
Physical Capacity Squared	0.000	−0.025	0.024	−0.018
	(0.018)	(0.018)	(0.021)	(0.029)
Cognitive Capacity Squared	−0.283***	−0.294***	−0.182***	−0.253***
	(0.021)	(0.025)	(0.030)	(0.067)
Exercise Squared	0.048**	0.007	0.022	0.030
	(0.024)	(0.021)	(0.022)	(0.032)
Cognitive Stimulation Squared	−0.033	−0.044**	−0.041**	−0.088***
	(0.023)	(0.018)	(0.020)	(0.028)
Physical Capacity × Cognitive Capacity	0.077***	0.194***	0.074*	0.117*
	(0.027)	(0.028)	(0.038)	(0.062)
Physical Capacity × Exercise	−0.015	0.022	−0.032	0.024
	(0.032)	(0.028)	(0.034)	(0.042)
Physical Capacity × Cognitive Stimulation	−0.018	−0.068***	−0.024	0.006
	(0.029)	(0.025)	(0.028)	(0.043)
Cognitive Capacity × Exercise	−0.071**	−0.151***	0.015	−0.128*
	(0.032)	(0.031)	(0.039)	(0.073)
Cognitive Capacity × Cognitive Stimulation	0.245***	0.240***	0.164***	0.305***
	(0.031)	(0.032)	(0.039)	(0.064)
Exercise × Cognitive Stimulation	−0.065*	0.037	−0.026	0.021
	(0.036)	(0.029)	(0.034)	(0.050)
Constant	0.002	−0.053***	−0.127***	−0.177***
	(0.010)	(0.010)	(0.012)	(0.021)
Note:			***p<0.01;**p<0.05;*p<0.1	

Table D.12. Transition Parameters for Cognitive Capacity, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.060^{***}	0.018	0.040^{*}	−0.009
	(0.014)	(0.017)	(0.021)	(0.045)
Cognitive Capacity	0.717^{***}	0.709^{***}	0.781^{***}	0.806^{***}
	(0.014)	(0.015)	(0.025)	(0.050)
Exercise	0.021^{*}	0.016	0.014	0.045
	(0.011)	(0.011)	(0.016)	(0.043)
Cognitive Stimulation	0.064^{***}	0.071^{***}	0.066^{***}	0.017
	(0.011)	(0.011)	(0.016)	(0.027)
Physical Capacity Squared	0.030	−0.019	0.002	−0.034
	(0.024)	(0.025)	(0.029)	(0.051)
Cognitive Capacity Squared	-0.136^{***}	-0.116^{***}	-0.068^{*}	−0.009
	(0.023)	(0.030)	(0.038)	(0.062)
Exercise Squared	0.010	0.030	0.031	−0.029
	(0.020)	(0.019)	(0.021)	(0.047)
Cognitive Stimulation Squared	−0.003	−0.006	−0.009	−0.005
	(0.015)	(0.015)	(0.018)	(0.023)
Physical Capacity × Cognitive Capacity	0.046	0.063	0.126^{***}	0.056
	(0.035)	(0.039)	(0.045)	(0.088)
Physical Capacity × Exercise	0.012	−0.007	−0.022	0.057
	(0.034)	(0.034)	(0.037)	(0.078)
Physical Capacity × Cognitive Stimulation	−0.026	−0.018	0.034	−0.005
	(0.029)	(0.030)	(0.033)	(0.055)
Cognitive Capacity × Exercise	-0.059^{*}	-0.079^{**}	-0.093^{**}	−0.029
	(0.031)	(0.033)	(0.040)	(0.077)
Cognitive Capacity × Cognitive Stimulation	0.066^{**}	0.058^{*}	0.054	−0.026
	(0.030)	(0.033)	(0.040)	(0.061)
Exercise × Cognitive Stimulation	−0.002	0.030	0.006	0.011
	(0.028)	(0.025)	(0.028)	(0.053)
Constant	−0.018	-0.075^{***}	-0.138^{***}	-0.152^{***}
	(0.011)	(0.012)	(0.015)	(0.029)

Note: $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$

Table D.13. Transition Parameters for Exercise, Females

	68-73	74-79	80-85	86-91
Physical Capacity	0.051***	0.053***	0.063***	0.037*
	(0.011)	(0.012)	(0.015)	(0.022)
Cognitive Capacity	0.020*	0.046***	0.133***	0.134***
	(0.011)	(0.012)	(0.016)	(0.026)
Exercise	0.969***	0.917***	0.864***	0.822***
	(0.014)	(0.013)	(0.017)	(0.025)
Constant	−0.067***	−0.096***	−0.139***	−0.233***
	(0.004)	(0.005)	(0.008)	(0.015)
Note:				***p<0.01;**p<0.05;*p<0.1

Table D.14. Transition Parameters for Exercise, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.125***	0.080***	0.121***	0.158***
	(0.014)	(0.015)	(0.022)	(0.038)
Cognitive Capacity	0.057***	0.045***	0.129***	0.090**
	(0.014)	(0.015)	(0.022)	(0.041)
Exercise	0.896***	0.934***	0.818***	0.759***
	(0.015)	(0.015)	(0.021)	(0.041)
Constant	−0.062***	−0.107***	−0.162***	−0.241***
	(0.005)	(0.006)	(0.011)	(0.022)
Note:				***p<0.01;**p<0.05;*p<0.1

Table D.15. Transition Parameters for Cognitive Stimulation, Females

	68-73	74-79	80-85	86-91
Physical Capacity	−0.013	0.041***	0.047**	0.060
	(0.015)	(0.015)	(0.022)	(0.048)
Cognitive Capacity	0.034	0.049**	0.077**	0.119*
	(0.025)	(0.023)	(0.037)	(0.063)
Cognitive Stimulation	1.030***	0.981***	0.943***	0.949***
	(0.021)	(0.018)	(0.027)	(0.042)
Constant	−0.028***	−0.055***	−0.062***	−0.109***
	(0.008)	(0.009)	(0.015)	(0.033)
Note:				***p<0.01;**p<0.05;*p<0.1

Table D.16. Transition Parameters for Cognitive Stimulation, Males

	68-73	74-79	80-85	86-91
Physical Capacity	0.036*	0.051**	0.079**	0.023
	(0.021)	(0.022)	(0.039)	(0.096)
Cognitive Capacity	0.100***	0.045	0.049	0.119
	(0.030)	(0.035)	(0.049)	(0.134)
Cognitive Stimulation	0.958***	0.975***	0.949***	0.811***
	(0.019)	(0.020)	(0.033)	(0.070)
Constant	−0.023**	−0.049***	−0.057**	−0.088
	(0.010)	(0.012)	(0.024)	(0.072)
Note:				***p<0.01;**p<0.05;*p<0.1

D.5 Distributions of initial factors and of shocks to factors

Table D.17. Distribution of the initial states, females

Factor	Mean	Standard Deviation	Correlation with			
			Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Physical Capacity	0.10	0.62	1.00	0.34	0.66	0.40
Cognitive Capacity	0.12	0.45	0.34	1.00	0.30	0.52
Exercis	0.11	0.60	0.66	0.30	1.00	0.50
Cognitive Stimulation	0.07	0.64	0.40	0.52	0.50	1.00

Table D.18. Distribution of the initial states, males

Factor	Mean	Standard Deviation	Correlation with			
			Physical Capacity	Cognitive Capacity	Exercise	Cognitive Stimulation
Physical Capacity	0.03	0.55	1.00	0.32	0.60	0.31
Cognitive Capacity	0.11	0.47	0.32	1.00	0.27	0.43
Exercise	0.10	0.65	0.60	0.27	1.00	0.33
Cognitive Stimulation	0.03	0.78	0.31	0.43	0.33	1.00

Table D.19. Standard deviations of shocks

Age	Factor	Female	Male
68-73	Physical Capacity	0.061***	0.091***
		(0.010)	(0.008)
	Cognitive Capacity	0.308***	0.294***
		(0.004)	(0.005)
	Exercise	0.167***	0.251***
		(0.010)	(0.010)
	Cognitive Stimulation	0.001	0.162***
		(2.783)	(0.029)
74-79	Physical Capacity	0.155***	0.146***
		(0.005)	(0.007)
	Cognitive Capacity	0.299***	0.286***
		(0.005)	(0.006)
	Exercise	0.252***	0.267***
		(0.009)	(0.011)
	Cognitive Stimulation	0.139***	0.212***
		(0.021)	(0.024)
80-85	Physical Capacity	0.184***	0.218***
		(0.008)	(0.008)
	Cognitive Capacity	0.269***	0.242***
		(0.006)	(0.007)
	Exercise	0.275***	0.330***
		(0.012)	(0.015)
	Cognitive Stimulation	0.222***	0.352***
		(0.030)	(0.034)
86-91	Physical Capacity	0.234***	0.257***
		(0.012)	(0.019)
	Cognitive Capacity	0.240***	0.249***
		(0.011)	(0.011)
	Exercise	0.303***	0.362***
		(0.017)	(0.025)
	Cognitive Stimulation	0.271***	0.546***
		(0.046)	(0.052)
Note:		***p<0.01;**p<0.05;*p<0.1	

References

Agostinelli, Francesco, and Matthew Wiswall. 2016. "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development." Working Paper 22441. National Bureau of Economic Research. [18, 38, 39]

Agostinelli, Francesco, and Matthew Wiswall. Forthcoming. "Estimating the Technology of Children's Skill Formation." *Journal of Political Economy*, [17, 20]

Amengual, Dante, Jesús Bueren, and Julio A Crego. 2021. "Endogenous health groups and heterogeneous dynamics of the elderly." *Journal of Applied Econometrics* 36 (7): 878–97. [6]

Attanasio, Orazio, Flávio Cunha, and Pamela Jervis. 2019. "Subjective Parental Beliefs. Their Measurement and Role." NBER Working Papers 26516. National Bureau of Economic Research. [20]

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2020. "Human Capital Development and Parental Investment in India." *The Review of Economic Studies* 87 (6): 2511–41. [19, 20]

Baker, Michael, Mark Stabile, and Catherine Deri. 2004. "What do self-reported, objective, measures of health measure?" *Journal of Human Resources* 39 (4): 1067–93. [3, 5]

Ball, Karlene, Daniel Berch, Karin Helmers, Jared Jobe, Mary Leveck, Michael Marsiske, John Morris, George Rebok, David Smith, Sharon Tennstedt, Frederick Unverzagt, and Sherry Willis. 2002. "Effects of Cognitive Training Interventions With Older Adults: A Randomized Controlled Trial." *JAMA: The Journal of the American Medical Association* 288: 2271–81. [2]

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of Econometrics*. Vol. 5, Elsevier, 3705–843. [3]

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. 2025. *JAX: Transformable numerical computing at scale*. Version 0.5.0. URL: http://github.com/google/jax. [20]

Chang, Yu-Hung, I-Chien Wu, and Chao A. Hsiung. 2021. "Reading activity prevents long-term decline in cognitive function in older people: evidence from a 14-year longitudinal study." *International Psychogeriatrics* 33 (1): 63–74. [12]

Clouston, Sean A. P., Paul Brewster, Diana Kuh, Marcus Richards, Rachel Cooper, Rebecca Hardy, Marcie S. Rubin, and Scott M. Hofer. 2013. "The Dynamic Relationship Between Physical Function and Cognition in Longitudinal Aging Cohorts." *Epidemiologic Reviews* 35 (1): 33–50. [3]

Contoyannis, Paul, Andrew M. Jones, and Nigel Rice. 2004. "The dynamics of health in the British Household Panel Survey." *Journal of Applied Econometrics* 19 (4): 473–503. [5]

Crimmins, Eileen M. 2020. "Social hallmarks of aging: Suggestions for geroscience research." *Ageing Research Reviews* 63: 101136. [2, 6]

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. [16, 17]

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources* 43 (4): 738–82. [19]

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931. [1, 3, 4, 6, 16, 18–20, 24, 33, 38, 39]

Cutler, David M, and Elizabeth Richardson. 1997. "Measuring the health of the U.S. population." *Brookings Papers on Economic Activity. Microeconomics* 1997: 217–82. [5]

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan. 2022. "Identification of dynamic latent factor models of skill formation with translog production." *Journal of Applied Econometrics* 37: (6), 1256–65. [19]

Dixon, Roger A., and Cindy M. de Frias. 2014. "Cognitively elite, cognitively normal, and cognitively impaired aging: Neurocognitive status and stability moderate memory performance." *Journal of Clinical and Experimental Neuropsychology* 36 (4): 418–30. [11]

Dowd, Jennifer Beam, and Anna Zajacova. 2007. "Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US?" *International Journal of Epidemiology* 36 (6): 1214–21. [5]

Dowd, Jennifer Beam, and Anna Zajacova. 2010. "Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data." *Annals of Epidemiology* 20 (10): 743–49. [5]

Fiatarone, Maria A., Evelyn F. O'Neill, Nancy Doyle Ryan, Karen M. Clements, Guido R. Solares, Miriam E. Nelson, Susan B. Roberts, Joseph J. Kehayias, Lewis A. Lipsitz, and William J. Evans. 1994. "Exercise Training and Nutritional Supplementation for Physical Frailty in Very Elderly People." *New England Journal of Medicine* 330 (25): 1769–75. [2]

Freyberger, Joachim. 2024. "Normalizations and misspecification in skill formation models." Working Paper. [18, 19, 38]

Gabler, Janos. 2024. *A Python Tool for the Estimation of large scale scientific models.* Version 0.5.1. URL: https://github.com/optimagic-dev/optimagic. [21]

Gabler, Janos. 2025. *skillmodels: A Python Library to Estimate Nonlinear Dynamic Latent Factor Models.* URL: https://github.com/OpenSourceEconomics/skillmodels. [20]

Gaudecker, Hans-Martin von. 2019. "Templates for Reproducible Research Projects in Economics." URL: https://doi.org/10.5281/zenodo.2533241. [21]

Gill, Thomas M., Evelyne A. Gahbauer, Ling Han, and Heather G. Allore. 2010. "Trajectories of Disability in the Last Year of Life." *New England Journal of Medicine* 362 (13): 1173–80. [23]

Health and Retirement Study. 2022a. "Cross-Wave Imputation of Cognitive Functioning Measures 1992-2018." URL: https://hrsdata.isr.umich.edu/data-products/cross-wave-imputation-cognitive-functioning-measures-1992-2018 (visited on 07/05/2022). [9]

Health and Retirement Study. 2022b. "HRS Public Survey Data." URL: https://hrsdata.isr.umich.edu/data-products/public-survey-data?_ga=2.253716109.582073791.1631194254-1744477196.1631018779 (visited on 07/05/2022). [7]

Health and Retirement Study. 2022c. "RAND HRS Products." URL: https://hrsdata.isr.umich.edu/data-products/rand?_ga=2.178170657.582073791.1631194254-1744477196.1631018779 (visited on 07/05/2022). [7]

Heckman, James J, and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Reviews in Economics* 6 (1): 689–733. [2]

Heiss, Florian. 2011. "Dynamics of self-rated health and selective mortality." *Empirical Economics* 40 (1): 119–40. [5]

Hosseini, Roozbeh, Karen A. Kopecky, and Kai Zhao. 2022. "The evolution of health over the life cycle." *Review of Economic Dynamics* 45: 237–63. [2, 3, 6, 8]

Huang, Zhiyong, and Jürgen Maurer. 2019. "Validity of Self-Rated Memory Among Middle-Aged and Older Chinese Adults: Results From the China Health and Retirement Longitudinal Study (CHARLS)." *Assessment* 26 (8): 1582–93. [23]

Idler, Ellen L., and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38 (1): 21–37. [5]

Jenicek, Milos, Robert Cleroux, and Michel Lamoureux. 1979. "Principal component analysis of four health indicators and construction of a global health index in the aged." *American Journal of Epidemiology* 110 (3): 343–49. [5]

Julier, Simon J., and Jeffrey K. Uhlmann, editors. 1997. *New extension of the Kalman filter to nonlinear systems*. International Society for Optics and Photonics. [33]

Jürges, Hendrik. 2007. "True health vs. response styles: exploring cross-country differences in self-reported health." *Health Economics* 16 (2): 163–78. [5]

Kalman, Rudolph Emil. 1960. "A new approach to linear filtering and prediction problems." *0021-9223*, [30]

Kapteyn, Arie, James Banks, Mark Hamer, James P Smith, Andrew Steptoe, Arthur Van Soest, Annemarie Koster, and Saw Htay Wah. 2018. "What they say and what they do: comparing physical activity across the USA, England and the Netherlands." *Journal of Epidemiology and Community Health* 72 (6): 471–76. [3, 4]

Karzmark, Peter. 2000. "8." *International Journal of Geriatric Psychiatry* Validity of the serial seven procedure (15): [11]

Kasper, Judith D, Kitty S Chan, and Vicki A Freedman. 2017. "Measuring physical capacity: an assessment of a composite measure using self-report and performance-based items." *Journal of Aging and Health* 29 (2): 289–309. [3]

Latham, Kenzie, and Chuck W. Peek. 2012. "Self-Rated Health and Morbidity Onset Among Late Midlife U.S. Adults." *The Journals of Gerontology: Series B* 68 (1): 107–16. [5]

Lindeboom, Maarten, and Eddy Van Doorslaer. 2004. "Cut-point shift and index shift in self-reported health." *Journal of Health Economics* 23 (6): 1083–99. [5]

Maurer, Jürgen, Roger Klein, and Francis Vella. 2011. "Subjective health assessments and active labor market participation of older men: evidence from a semiparametric binary choice model with nonadditive correlated individual-specific effects." *Review of Economics and Statistics* 93 (3): 764–74. [5]

McCammon, Ryan J., Gwenith G. Fisher, Halimah Hassan, Jessica D. Faul, Willard L. Rodgers, and David R. Weir. 2022. "Health and Retirement Study – Imputation of Cognitive Functioning Measures: 1992-2018 Data Description." Techreport. Version 7.0. Survey Research Center University of Michigan. [11]

McFadden, Daniel. 2008. "Human capital accumulation and depreciation." *Applied Economic Perspectives and Policy* 30 (3): 379–85. [2]

Montero-Odasso, Manuel, Joe Verghese, Olivier Beauchet, and Jeffrey M Hausdorff. 2012. "Gait and cognition: a complementary approach to understanding brain function and the risk of falling." *Journal of the American Geriatrics Society* 60 (11): 2127–36. [3]

Nakazato, Yuichi, Tomoko Sugiyama, Rena Ohno, Hirofumi Shimoyama, Diana L Leung, Alan A Cohen, Riichi Kurane, Satoru Hirose, Akihisa Watanabe, and Hiromi Shimoyama. 2020. "Estimation of homeostatic dysregulation and frailty using biomarker variability: a principal component analysis of hemodialysis patients." *Scientific Reports* 10 (1): 1–12. [5]

Niccoli, Teresa, and Linda Partridge. 2012. "Ageing as a Risk Factor for Disease." *Current Biology* 22 (17): R741–R752. [2]

Organization, World Health. 2015. *World report on ageing and health*. World Health Organization, 246 p. [2]

Poterba, James M, Steven F Venti, and David A Wise. 2017. "The asset cost of poor health." *The Journal of the Economics of Ageing* 9: 172–84. [5]

Prvan, Tania, and M. R. Osborne. 1988. "A Square-Root Fixed-Interval Discrete-Time Smoother." *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 30 (1): 57–68. [20, 35]

Raabe, Tobias. 2024. *A Python tool for managing scientific workflows.* Version 0.5.2. URL: https://github.com/pytask-dev/pytask. [21]

Rockwood, Kenneth, and Arnold Mitnitski. 2007. "Frailty in relation to the accumulation of deficits." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62 (7): 722–27. [6]

Runge, Shannon K. 2015. "Word Recall: Cognitive Performance Within Internet Surveys." *JMIR Mental Health* 2 (2): [11]

Salthouse, Timothy. 2010. "Selective review of aging." *Journal of the International Neuropsychological Society* 16: 754–60. [3]

Salthouse, Timothy. 2012. "Consequences of age-related cognitive declines." *Annual Review of Psychology* 63: 201. [3]

Schiele, Valentin, and Hendrik Schmitz. 2023. "Understanding cognitive decline in older ages: The role of health shocks." *European Economic Review* 151 (104320): [6]

Steiber, Nadia. 2016. "Strong or Weak Handgrip? Normative Reference Values for the German Population across the Life Course Stratified by Sex, Age, and Body Height." *PLOS One* 10 (11): [9]

Särkämö, Teppo, and David Soto. 2012. "Music listening after stroke: beneficial effects and potential neural mechanisms." *Annals of the New York Academy of Sciences* 1252 (1): 266–81. [12]

Särkämö, Teppo, Mari Tervaniemi, Sari Laitinen, Anita Forsblom, Seppo Soinila, Mikko Mikkonen, Taina Autti, Heli M. Silvennoinen, Jaakko Erkkilä, Matti Laine, Isabelle Peretz, and Marja Hietanen. 2008. "Music listening enhances cognitive recovery and mood after middle cerebral artery stroke." *Brain* 131 (3): 866–76. [12]

Van Der Merwe, Rudolph. 2004. "Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models." Phdthesis. OGI School of Science & Engineering at Oregon Health & Science University. URL: http://www.cslu.ogi.edu/publications/ps/merwe04.pdf. [32]

van der Merwe, Rudolph, and Eric A. Wan, editors. 2001. *The square-root unscented Kalman filter for state and parameter-estimation.* IEEE. [20, 35]

World Bank. 2018. *World development report 2019: The changing nature of work.* The World Bank. [2]

Zhang, Fuzhen. 1999. *Matrix Theory: Basic Results and Techniques.* Universitext (Berlin. Print). Springer. [33, 35]