Retirement and cognitive fonctioning: A longitidinal analysis

Auteur : Habets, Martin
Promoteur(s) : Perelman, Sergio
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en sciences économiques, orientation générale, à finalité spécialisée en Economics and Society
Année académique : 2017-2018
URI/URL : http://hdl.handle.net/2268.2/4789

LIÈGE université
HEC LIÈGE École de gestion

Retirement and cognitive functioning: A longitudinal analysis

Martin Habets

A thesis submitted for the degree of
Master in Economics

Supervised by

Sergio Perelman

Readers:

Joseph Tharakan
Jérôme Schoenmaeckers

Academic year 2017–18

Abstract — Retirement induces major changes in the lifestyle of elderly people and understanding its influences is crucial to promote successful cognitive ageing. In this analysis, I estimate the effect of retiring on cognitive functioning using an instrumental variable strategy, instrumenting for retirement with legal eligibility ages. I exploit the panel dimension of the Survey on Health, Ageing and Retirement in Europe (SHARE) to control for individual heterogeneity and find a positive effect of retirement on memory. I also show suggestive evidence that the beneficial effect of retirement on cognition is larger for women and high-educated indivuals, altough heterogeneous across regions of Europe.

Keywords — Cognition, Ageing, Retirement, SHARE, Panel estimation, IV fixed-effects

This thesis has greatly benefitted from the time and support of many people.

First of all, I gratefully acknowledge the help of my supervisor, Sergio Perelman. Throughout this research, he has been extremely approachable and constantly available, offering valuable guidance and thoughtful advice. For the many stimulating discussions we had and his enthusiasm, I thank Jérôme Schoenmaeckers. Further, I thank Joseph Tharakan for being a reader of this thesis. I am also grateful to Julien Jacqmin for introducing me to economic researching.

Additionally, I would like to use this opportunity to express my sincere appreciation to ECLAC researchers and to my peers of the summer school, who have taught me to always question my own thoughts and to be my own best sceptic.

Finally, I am deeply indebted to my family. I thank my brother, whose unfaltering commitment to his own undertakings has proven to be contagious, and my parents, whose encouragement and support were pivotal during this journey.

Contents

List of Figures

List of Tables

Introduction and review of empirical literature

The "use it or lose it" hypothesis refers to the assumption that an individual can impede cognitive ageing – the deterioration of mental performance with age – by engaging in cognitively demanding activities. Conversely, the hypothesis holds that an undemanding environment will exert opposite effects.

While this psychological hypothesis is not unequivocally proven, Hertzog, Kramer, Wilson, and Lindenberger (2008) infer from the available evidence that maintaining an intellectually engaged and physically active lifestyle promotes successful cognitive ageing. Based on this assumption, the elderly are urged to maintain an engaged lifestyle, as cognitive decline is not exogenous and can be affected by their behaviour.

The aim of this study is to address the impact of retirement on cognitive functioning, as retirement is a major life event that induces changes in individuals' lifestyles. Yet, although the changes in lifestyles are expected to affect cognitive functioning, the direction of the effect is ambiguous. On the one hand, retirement may result in harmful effects, such as a decrease in stimulating activities, a reduction of social interactions, or a loss of market incentive to invest in cognitive repair activities, which, in turn, would lead to a less mentally challenging lifestyle. On the other hand, retirement may have a positive impact on cognitive functioning, in that, for example, an increase in free time may lead individuals to raise their investment in cognitive stimulating leisure activities, or that the removal of work-related constraints may bring about positive spillovers on retirees' mental health.

This study addresses the question of which of those two conflicting effects prevails by estimating the causal impact of retirement on cognitive functioning, measured by word recall, using a survey of older persons in 13 European countries.

Review of empirical literature

Particular attention has been devoted to the impact of retirement on cognition in the economic literature since the seminal paper by Adam, Bonsang, Germain, and Perelman (2007). Still, in a systematic literature review, Meng, Nexø, and Borg (2017) reveal the existence of a major knowledge gap in regards to the impact of retirement on cognitive decline. Indeed, there is no clear consensus, as the evidence is conflicting. Some studies find that retirement leads to a decline in cognition (*e.g.* Rohwedder & Willis, 2010; Bonsang, Adam, & Perelman, 2012; Mazzonna & Peracchi, 2012) while other find mixed or positive effects (*e.g.* Coe, von Gaudecker, Lindeboom, & Maurer, 2012; Bianchini & Borella, 2016).

Although a majority of papers finds that retirement has a detrimental effect on cognition, the size and effects vary widely depending on the specification used. Indeed, the type of specification seems to lead to large changes in the magnitude and significance of the estimated effects. Fonseca, Kapteyn, and Zamarro (2016) replicate several studies and find the results to be very sensitive to differences in econometric specifications.

Most of the studies rely on a cross-sectional approach. Rohwedder and Willis (2010) find negative effects when using data from various surveys, such as SHARE, ELSA and HRS. Mazzonna and Peracchi (2012) also find a negative effect, while Coe and Zamarro (2011), using the first wave of SHARE, do not find a significant effect of retirement on cognition. These studies suggest that there is a negative association between retirement and cognitive functioning. However, their cross-sectional nature does not allow for an evidence of a causal relationship. Cross-sectional designs only inform about differences in cognitive functioning, not changes in cognitive functioning over time, because they only measure cognitive functioning at one point in time.

To help overcome the limitations and shortcomings encountered in these studies, Bonsang et al. (2012) exploit the longitudinal information provided by the American HRS dataset to estimate a fixed-effects specification. The authors find a significant negative effect of retirement on cognition. Following this innovation, three papers exploited the European longitudinal information of SHARE in a similar fashion.

First, Mazzonna and Peracchi (2014) find substantial heterogeneity in the effect of retirement across occupational groups. Using a principal-component analysis to express cognitive functioning, the effect is found to be negative for most workers, but positive for those employed in highly physically demanding jobs. Second, Bianchini and Borella (2016) support that time spent in retirement has a positive effect on cognition by using an instrumental variables fixed-effects (IV-FE) estimation. Third, Celidoni, Dal Bianco, and Weber (2017) construct a measure of cognition based on word recall and find that retirement has a long-term detrimental effect on cognition for individuals that retire at the statutory eligibility age, but plays a protective role for those who retire as soon as possible.

Overall, I consider these four papers to be the closest to this study. However, only Bonsang et al. (2012) and Bianchini and Borella (2016) use a similar empirical approach to the one developed here, namely, the IV-FE estimation.

I contribute to the literature in several dimensions. First, I use different data that all of the above-mentioned studies. Although different papers rely on SHARE, all exploit a maximum of the first four waves and not the information available in the most recent ones. As a result, the studies restrict attention to mostly Western European countries, and none include respondents from Estonia, Czech Republic or Slovenia, for example. Second, most importantly, I investigate different sources of heterogeneity in the effect of retirement on cognition.

Main findings

My findings suggest that retirement has an overall positive effect on cognitive functioning. Specifically, the estimates show that retiring implies a 4% increase in memory score, accounting for a delayed effect. It appears that retirement does not affect everyone the same. When exploring different sources of heterogeneity, I find a clear education gradient. The effect of retirement is positive for low and medium-educated individuals, but twice larger for high-educated respondents. Also, retirement displays a higher and more significant effect on women. Finally, retirement is found to affect cognitive functioning differently in the various regions and countries of Europe.

Organisation of the thesis

This thesis is organised as follows. Chapter 1 presents the data from the SHARE project. The first section defines data selection and following sections discuss the dependent variables, endogenous regressors and control variables. Chapter 2 delves into regression analyses. It starts by exposing the empirical approach in the first section and then details the results of the empirical analyses. I develop the analysis by accounting for a delayed effect, and introduce population heterogeneity. Finally, I discuss the findings and their implications on public policy and then conclude.

CHAPTER 1

SHARE project data

In this chapter, I introduce the data from the SHARE project and its use in this study. Section 1.1 explains the sample selection process and introduces its particularities. Further, I go through the different measures of cognitive abilities that are provided by four brief tests in section 1.2. Then, section 1.3 defines the retirement variable and details the construction of the instruments. I also present some statistics that suggest a wide heterogeneity across retirement schemes in Europe. After that, the control variables are specified in section 1.4. Finally, table 1.4 reports the descriptive statistics of the variables described throughout this chapter and used in the regression analysis of chapter 2.

The data used are from the waves 4, 5 and 6 of the Survey of Health, Ageing and Retirement in Europe (SHARE)[1], a multidisciplinary and cross-national population representative longitudinal survey which collects information on health, socioeconomic status and social and family networks. The target population is individuals aged 50 or over who speak the official language(s) of their country. Their partners, regardless of age, are also included.

The collection of the data is made through personal interviews that were conducted for waves 4, 5 and 6, respectively, in 2011, 2013 and 2015. Respondents were interviewed in multiple waves, but the sample was also refreshed to keep it representative of the ageing population at each wave.

[1]This paper uses data from SHARE Waves 4, 5 and 6 (DOIs: 10.6103/ SHARE. w4.600, 10.6103/ SHARE. w5.600, 10.6103/ SHARE. w6.600), see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6 - CT - 2001 - 00360), FP6 (SHARE - I3: RII- CT - 2006 - 062193, COMPARE: CIT5 - CT - 2005 - 028857, SHARELIFE: CIT4 - CT - 2006 - 028812) and FP7 (SHARE - PREP: N.211909, SHARE - LEAP: N.227822, SHARE M4: N.261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Ageing (U01 AG 09740 - 13S2, P01 AG 005842, P01 AG 08291, P30 AG 12815, R21 AG 025169, Y1 - AG - 4553-01, IAG BSR06-11, OGHA 04-064, HHSN 271201 300071 C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

The interview mode is Computer Assisted Personal Interview (CAPI). The questionnaire consists of 20 modules covering several aspects of life circumstances. This work uses data from the modules on demographics, physical and mental health, employment and pensions which complements the information supplied by the easySHARE data set[2]. The data processing and analysis was made using the statistical software package Stata. The code executed is shown in appendix C.

1.1 Sample selection

I restrict attention to individuals interviewed in the countries that contributed in every of the waves 4, 5 and 6. The selected countries represent different regions of continental Europe. The final sample includes data from 12 countries. Drawing on previous research, cultural roots, attitudes and welfare regimes (*e.g.* Tosi & Grundy, 2018), in some analyses I divided countries into four groups: Western European countries (Austria, Belgium, France, Germany and Switzerland), Nordic countries (Denmark and Sweden) a Southern Europe group (Italy and Spain) and a cluster of former Eastern bloc countries (Estonia, Czech Republic and Slovenia). Although representative of the population, the sample size is not proportional to the number of people living in each country.

In the study, the sample is restricted to individuals that participated in at least two of the three waves. Among these, I keep individuals that participated to wave 6. This leaves me with respondents that first appear either on wave 4 or 5, and provides a common measurement on wave 6. I further keep only individuals that were aged more than 50 when first interviewed or less than 75 when last interviewed. At this point, 34,597 individuals are included the sample.

Retired and employed people represent 85% of the sample. Since I am interested in studying the transition between work and retirement, I exclude individuals that declared belonging to any other categories, namely, individuals who reported themselves permanently sick or disabled, unemployed and homemakers or who did not provide information about their employment status. For these 7,963 people, the instruments would not be sufficiently relevant and the effect of retirement on cognitive abilities may be peculiar (Bianchini & Borella, 2016). Additionally, I exclude from the analysis 361 individuals who reported returning to work during the sample period. As explained and assessed in Bonsang et al. (2012), including those individuals would require the assumption that leaving the workforce and going back exert symmetrical effects on cognitive functioning.

Besides, I exclude 1,009 individuals for which word recall test results are missing. This is primarily a result of proxy interviews, during which cognitive tests are not performed. I then drop 2 respondents for whom the retirement year is missing and 231 individuals that retired before the age of 45.

[2]Generated easySHARE data set (DOI: 10.6103/ SHARE. easy. 600), see Gruber, Hunkler, and Stuck (2014) for methodological details

Table 1.1
Gender distribution of selected individuals by country

	Men		Women		Total
	Obs.	%	Obs.	%	Obs.
Austria	810	45.1	988	54.9	1798
Belgium	1273	53.2	1122	46.8	2395
Czech Republic	1154	39.1	1797	60.9	2951
Denmark	1066	48.2	1147	51.8	2213
Estonia	796	35.3	1458	64.7	2254
France	920	47.8	1006	52.2	1926
Germany	1307	50.5	1282	49.5	2589
Italy	1000	57.1	750	42.9	1750
Slovenia	642	42.0	888	58.0	1530
Spain	1080	61.4	679	38.6	1759
Sweden	1071	45.5	1283	54.5	2354
Switzerland	785	51.9	727	48.1	1512
All countries	11904		13127		25031

The final sample corresponds to an unbalanced panel including 65,120 observations for 25,031 individuals. Among those, 60% participated to the three waves, 35% to the waves 5 and 6, while the remaining 5% participated only to waves 4 and 6. Wave 6 is thus the wave common to every respondent.

Table 1.1 shows the distribution of the final sample by country and gender. We notice that women are undersampled in some countries, such as Spain (38.6%) and Italy (42.9%), whereas the opposite is true in others. Eastern European countries make a good example. Women represent 64.7%, 60.9% and 58% of the final sample in Estonia, Czech Republic and Slovenia, respectively. However, the original SHARE study consistently sampled more women than men, as shows table A.1 (page 39).

We can thus conclude that the selection is the main cause of the differences in gender representation in the final sample. In order to understand the role of the selection process in this sampling heterogeneity, table 1.2 shows the proportion of selected men and women from the original sample, by country. For example, in Spain, 41% of men were selected, whereas only 21% of women of the initial sample were selected in the studied sample. The selection process excluded (included) 20% more (less) women than men in Spain. This difference in selection is shown in the last column of table 1.2. The numbers are particularly high in Spain and Italy. In those two countries, Mazzonna and Peracchi (2012) explain that about half of women never worked. This lower female attachment to labor could explain this selection result, as only current and past workers are included in the sample. The gender difference is also high in Belgium and Switzerland, although, to a lesser extent. This under representation of women in the final sample will exert considerable influence on the regression estimates, as we will see in section 2.6.

Table 1.2
Selection process: Percentages of selected men and women from the original sample, by country

	Men		Women		Gender diff.
	Not selected	Selected	Not selected	Selected	
Austria	57	43	62	38	5
Belgium	47	53	63	37	16
Czech Republic	49	51	45	55	-4
Denmark	36	64	43	57	7
Estonia	67	33	60	40	-7
France	53	47	62	38	9
Germany	38	62	46	54	8
Italy	48	52	68	32	20
Slovenia	49	51	49	51	0
Spain	59	41	79	21	20
Sweden	40	60	39	61	-1
Switzerland	45	55	58	42	13
All countries	49	51	56	44	7

1.2 Cognitive ability measures

I want the dependent variable to be a measure of cognitive abilities. However, cognitive decline is a multidimensional phenomenon and multiple aspects of the respondent's cognitive functioning are assessed in SHARE. This section describes each test and exposes the reasons that lead me to focus on the measure of memory.

The cognitive function module contains subjective and objective measures that assess four aspects of the respondent's cognitive functioning: numeracy, orientation in time, verbal fluency and memory. These four measures are the outcome of brief tests, included in the CAPI questionnaire, that follow a protocol aimed at minimising the potential influence of the interviewer and the interview process[3].

First, the numeracy test involves a simple arithmetical calculation based on percentages and gives information on the mathematical performance of the respondents. If the respondents had already participated in one of the panel waves, they would be asked a similar test, based on subtractions. The final scores have a narrow range, from 1 (bad) to 5 (good). In the sample under study, variability is very low. 14% of the observations are a score of 4 and 71% are a score of 5, the maximum score.

Second, analogously, the test of orientation in time displays very little variability. 91% of the respondents answered correctly to the four questions about the interview date (day, month and year) and day of the week, and thus scored the maximum. For this reason, I did

[3] See, for example, Mazzonna and Peracchi (2012) for more details

not use the results of the orientation and numeracy tests[4]. Indeed, I doubt they would yield to a realistic representation of the population's cognitive abilities distribution.

Third, verbal fluency is assessed through a simple test. The respondent is asked to name as many different animals as she or he can think of in one minute. The number of distinct animals enumerated among selected respondents – which is the score of the test – has a maximum of 100, while the mean is 23. Half of the results lie between 18 and 27.

Finally, memory is challenged with a ten-word-list learning test that consists of verbal registration and recall of the list[5]. The speed at which the words are read out is controlled by the CAPI. The respondent hears the list only once but is asked to recall as many words as possible on two occasions; first, right after the enumeration of the list (immediate recall) and, then, after an interference period (delayed recall), about 5 minutes later, at the end of the cognitive functioning module. The maximum score for each recall is 10. In waves 1 and 2, the test suffered from a drawback. Indeed, the tests were administered to respondents of the same household using the same list of words. While the CAPI clearly asks no third person to be present during the module, it has been reported that, for example, individuals were present during their partner tests. Repeated exposure can only be thought of as skewing the results. Incidentally, the same list was used with the same individual over time. Thus, in further waves (with the exception of wave 3, which, due to its peculiar nature did not include the cognitive functioning module), this issue was solved by administering a different list of words[6], to inhibit learning effects that would be likely to improve the cognitive scores of some respondents (see Malter and Börsch-Supan (2013) for methodology details). This drawback motivates my choice to work with waves 4, 5 and 6.

Figures 1.1 to 1.3 display cross-sectional age profiles of the cognitive tests that were not rejected due to low variability. They represent the age distribution of the averages of the fluency and the two recall tests. I differentiate the sample according to three criteria, namely, gender, education attainment, and region. Indeed, Schmitz-Scherzer and Thomae (1983) have found, for example, that individuals of lower socioeconomic status experience greater age-related decreases in performance on cognitive tests, so that socioeconomic status was shown to be positively related to better performance by older adults.

First, figure 1.1 displays the average test scores differentiating for gender. A substantial difference appears. Women correspond to better scores than men at all ages in the recall tests. In the fluency test, the confidence bands of the average lines overlap, but, again,

[4]Note that to be able to better discriminate respondents, Mazzonna and Peracchi (2012) used the time spent to answer to those questions to construct an adjusted test score. This information is made available by the advanced system used to conduct the CAPI interview but is not public, and could therefore not be used.

[5]Specifically, the question asked is : «Now, I am going to read a list of words from my computer screen. We have purposely made the list long so it will be difficult for anyone to recall all the words. Most people recall just a few. Please listen carefully, as the set of words cannot be repeated. When I have finished, I will ask you to recall aloud as many of the words as you can, in any order ».

[6]The CAPI randomly assigns one of the following ten-word lists : hotel, river, tree, skin, gold, market, paper, child, king, book — sky, ocean, flag, dollar, wife, machine ,home, earth, college, butter — woman, rock, blood, corner, shoes, letter, girl, house, valley, engine — water, church, doctor, palace, fire, garden, sea, village, baby, table

Figure 1.1

Age profiles of average test scores by gender

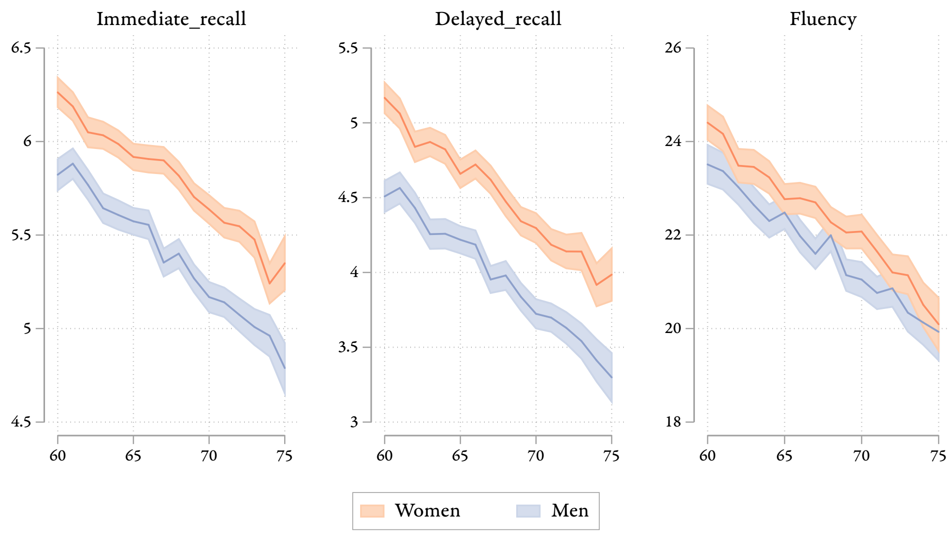


Figure 1.2

Age profiles of average test scores by macro region

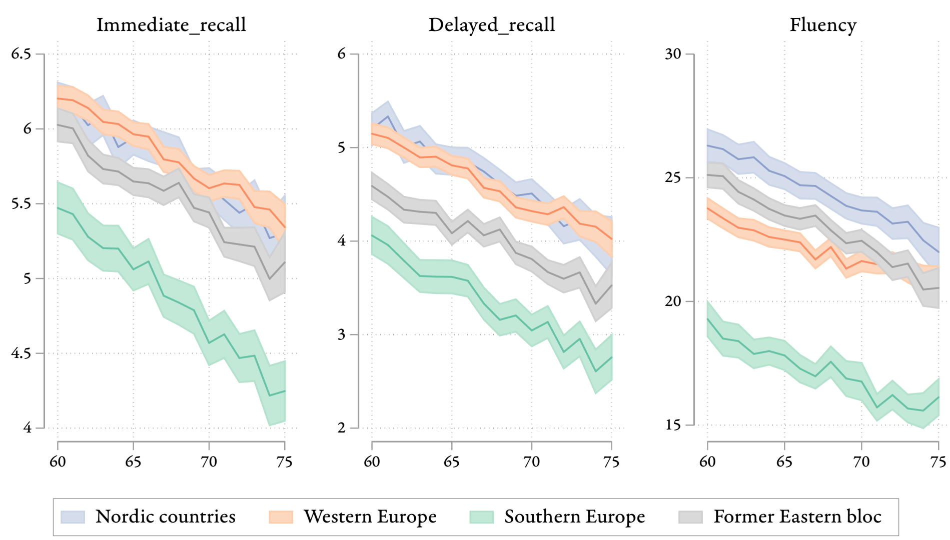
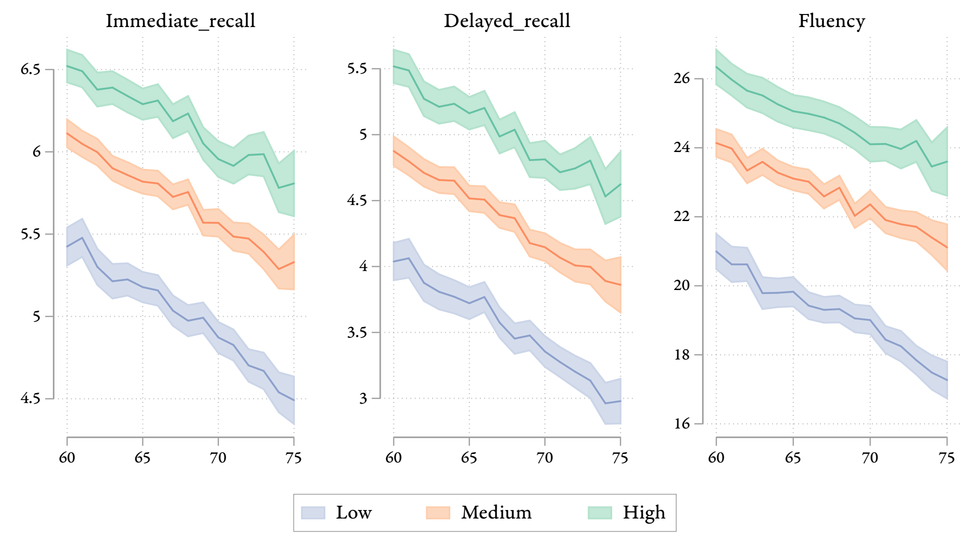


Figure 1.3
Age profiles of average test scores by education attainment



women tend to score better. Also, figure A.1 (p. 43) shows the average results of the recall tests by country and gender. The better results of women hold for every country sampled.

Furthermore, the figure (A.1) also indicates differences across countries in the level of the scores. To investigate this heterogeneity, figure 1.2 displays the average test score for macro-regions of continental Europe (see page 6 for the classification of countries). Western, Northern and Eastern Europe countries show little divergence in tests, although the group of Eastern European countries seems to lag behind in the recall tests. However, the most noticeable feature of these graphs is the results of Southern Europe that contrast with the rest. They consistently lie below those of the other European regions.

Finally, figure 1.3 displays the average test scores differentiating for levels of education. International Standard Classification of Education coding is used to account for country specificities in the educational system and is generated by the SHARE team. I separate people according to their educational attainment level. Following the methodology of Eurostat (2016), I create three aggregates: low, medium and high levels of education. The figure displays a clear education gradient in all tests. This is largely consistent with the results of Banks and Mazzonna (2012), who, using cognitive tests from the English Longitudinal Survey on Ageing (an English survey broadly similar to SHARE) provide evidence of a large positive and significant causal effect of a year of education on cognitive abilities at older ages. Also, Dal Bianco, Garrouste, and Paccagnella (2013) showed that cognitive decline and education are strongly – negatively – related [7]. Still, in figures 1.1 to 1.3, the main difference across groups lies in the level of the test and not in the rate of decline.

With these three measures of cognitive abilities, I ought to construct a single variable. Morris, Evans, Hebert, and Bienias (1999) highlighted some general constraints of great importance in measuring cognitive abilities, that need to be taken into account. SHARE is a multidisciplinary survey and thus suffers from three specific constraints: (i) the survey must cover a wide range of cognition levels, from excellent function to severe disease, (ii) it must provide a uniform measurement of cognition across respondents, (iii) the cognitive functioning module being only a part of the survey, it must be restricted to brief tests, to ease the respondent burden. These constraints lead to two potential issues. First, they cause the tests to summarize multiple facets of cognition. Second, they increase the likelihood of floor and ceiling effects.

As in most papers on cognitive decline[8], I will up to this point focus exclusively on the measure of memory score, as it allow to bypass those two issues. Indeed, first, as pointed out by Bingöl, Crespo, and Mira (2016), the memory test is a measure of pure fluid intelligence. This opposes to the verbal fluency test which measures both fluid and crystallised intelligences, that are crucial determinants of the verbal fluency score: crystallised intelligence is mainly responsible for knowing about a large number of distinct elements, while fluid intelligence allows one to remember them rapidly. Second, Bonsang et al. (2012) explain that memory score does not suffer from floor and ceiling effects. This contrasts with the orientation and numeracy tests that, as a result of ceiling effects, displayed low variability. Naturally, since individuals with the highest possible score can only change in one direction, random variation is not evenly distributed around initial scores.

Additionally, episodic memory, measured by memory score, is particularly affected by ageing. Some neuroscience studies argue it to be among the first cognitive function to decline with age (Anderson & Craik, 2000; Prull, Gabrieli, & Bunge, 2000; Souchay, Isingrini, & Espagnet, 2000). Because of these advantages, I consider memory score to allow for a more realistic representation of real innate cognitive ability levels over all other cognitive measures available in SHARE. Concretely, I construct the final indicator of cognitive functioning by summing the immediate and the delayed recall tests scores. The measure thus ranges from 0 to 20.

---

[7] Dal Bianco et al. studied the link between individual schooling and cognitive decline. They explain that Schneeweis, Skirbekk, and Winter-Ebmer (2012), working with SHARE data, assess the causal effect of education on old-age memory and fluency and find a positive impact of schooling on memory – with one year of education increasing the delayed memory score by about 0.3. However, they consider that the findings of Cavapozzi, Garrouste, and Paccagnella (2011) might drive this positive correlation. Indeed, Cavapozzi et al. shows that parental socio-economic background plays a crucial role in determining the average number of years in full-time education of an individual and their later socio-economic condition. For that reason, Dal Bianco et al., controlling for this positive correlation between socio-economic condition of parental households and educational attainment of respondents, confirmed Schneeweis et al.'s results of strong negative relation between cognitive decline and education.

[8] Every paper referred to in the literature review (*e.g.* Bonsang et al., 2012; Mazzonna & Peracchi, 2014; Bianchini & Borella, 2016; Celidoni et al., 2017) uses memory score as the dependent variable, or a measure constructed directly from it.

Figure 1.4
Average memory score by age

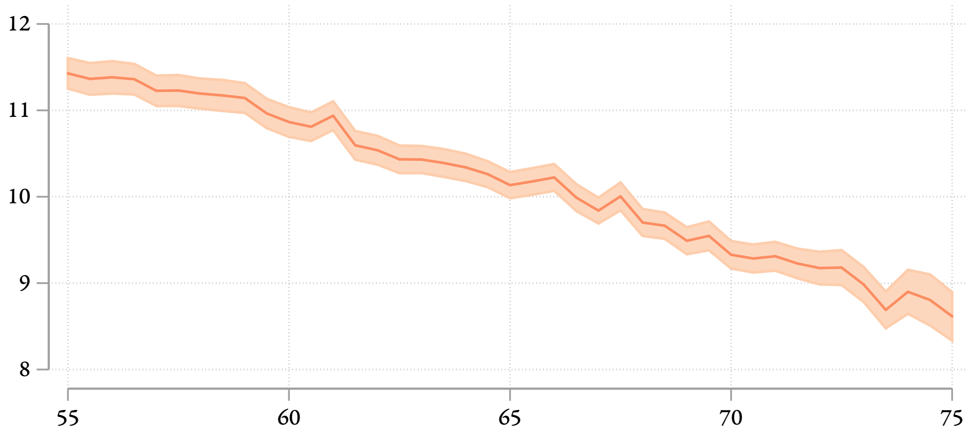


Figure 1.4 provides a first insight into this measure. Similarly to figures 1.1 to 1.3, it displays the variation of the average memory score measure with age. While it does not provide evidence of a systematic difference in the rate of decline of cognitive abilities with age, the figure clearly highlights the negative association between memory and age. Actually, on average, the number of words recalled decreases almost linearly with age. Even though this measure of episodic memory allows for great variability, we perceive a steady but slow decrease in score. Over the span of 20 years, the average score declines from 11.4 at age 55 to 8.6 at age 75.

Nonetheless, the cross-sectional nature of this kind of analysis does not allow us to infer a causal relationship, since we would observe time, age and cohort effects combined. Accordingly, to understand the link between retirement and word recall, I exploit longitudinal information. This allows me to look at the change in performance, a superior outcome measure, for it estimates the decline directly.

1.3 Retirement and instrumental variables

In the study of the effect of retirement on cognitive decline, the definition of retirement is naturally of high relevance. I define an individual as retired if he or she self-reports being retired from work. Thereby, I construct a dummy variable for retirement status using the employment information provided by the variable *ep005* from the module *EP*. This variable classifies the individual in each period, either as retired, or employed/self-employed as I excluded from the sample individuals that reported being sick, disabled or homemaker (see section 1.1).

Over the three waves, 54% of the selected individuals are retired during the entire period of the survey. Apart from those, every individual was working when first interviewed.

Table 1.3
Retirement related summary statistics by country

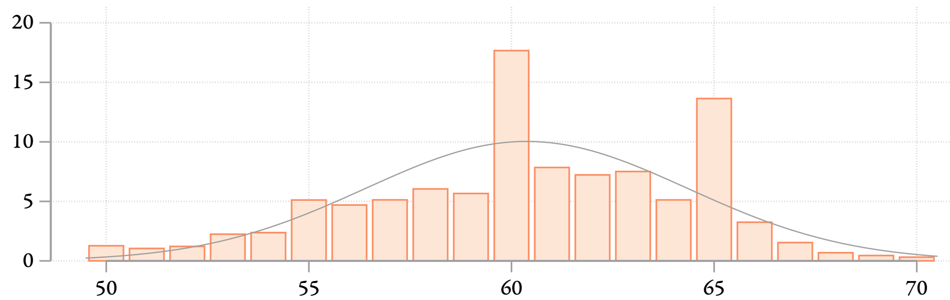
	Age of retirement	Employment status evolution: Working	Transition	Retired
	Mean	%	%	%
Austria	58.5	20	12	68
Belgium	60.3	38	12	50
Czech Republic	58.4	21	9	70
Denmark	62.3	52	9	39
Estonia	61.0	39	16	45
France	59.5	27	12	61
Germany	61.8	46	8	46
Italy	58.2	34	6	60
Slovenia	56.1	20	6	74
Spain	62.9	41	11	47
Sweden	63.7	37	11	52
Switzerland	62.8	49	15	36
All countries	60.5	35	11	54

35% of the respondents remained employed throughout all waves and the remaining 11% of respondents transited from employment to retirement during the period under study. Those 2,676 respondents are the most relevant type of observation to this longitudinal study. Table 1.3 exposes the relative size of each of those three categories in every country, as well as the average age of retirement.

Across retirement ages, there is a large cross-country variability. Average age of retirement goes from 56.1 in Slovenia to 63.7 in Sweden, while over the entire sample, the unweighted average is about 60. Figure 1.5 shows the retirement age distribution by age across the entire sample, while figure 1.6 displays the same distribution for each country separately. Overall, we notice a large prevalence of ages 60 and 65, but examining the country dissimilarities already suggests that broadly different retirement schemes are in application in every country.

Figure 1.6 shows a strong surge of probability for certain years of the retirement age distribution, but displays a large variability across countries. In order to account for the presumed endogeneity of the retirement decision, as in Bonsang et al. (2012) and most of the literature that followed, I use retirement eligibility ages as instruments, for both statutory and early retirement. The instruments must influence the retirement decision but should not be having a direct effect on cognitive functioning. In other words, they must be correlated with cognitive functioning only through the effect of retirement. Here, I assume that most of the cross-country variation is a result of national policies. Since it is highly unlikely that these policies have been set up in response to age-related cognitive performance in the country population, I argue that these policies provide valid instruments.

Figure 1.5
Retirement age distribution (percentage by age)



These key retirement ages are thus used as instruments for the retirement decision. Appendix B (page 45) explains the different retirement schemes of every country of the analysis and details the early and statutory retirement ages. I differentiate for country, gender and cohorts. The instrumental variables are constructed as two dummy variables that take the value zero if the individual's age is less than the legal retirement age, for both early and statutory retirement.

Finally, another retirement related variable of interest is the length of retirement. This variable measures the time elapsed between interview's and retirement's dates. Interview dates are available for all respondents, but retirement date, being based on a self-report, lacks precision. While retirement year[9] was largely provided, retirement month was not disclosed by half of the respondents. In those cases, I assumed the individuals to have retired in June. The final variable, *ret_length*, is computed for every interview date and is set to missing for respondents that remained employed throughout the panel period.

1.4 Control variables

Using the longitudinal dimension of the data to model a fixed effect specification allows to eliminate most of the individual heterogeneity, given that it controls for time-invariant observed and unobserved characteristics. However, Bingley and Martinello (2013) demonstrated that using the right controls is very important when using cross-country variation in retirement eligibility ages as instruments. Failing to do so leads to a bias in the instruments.

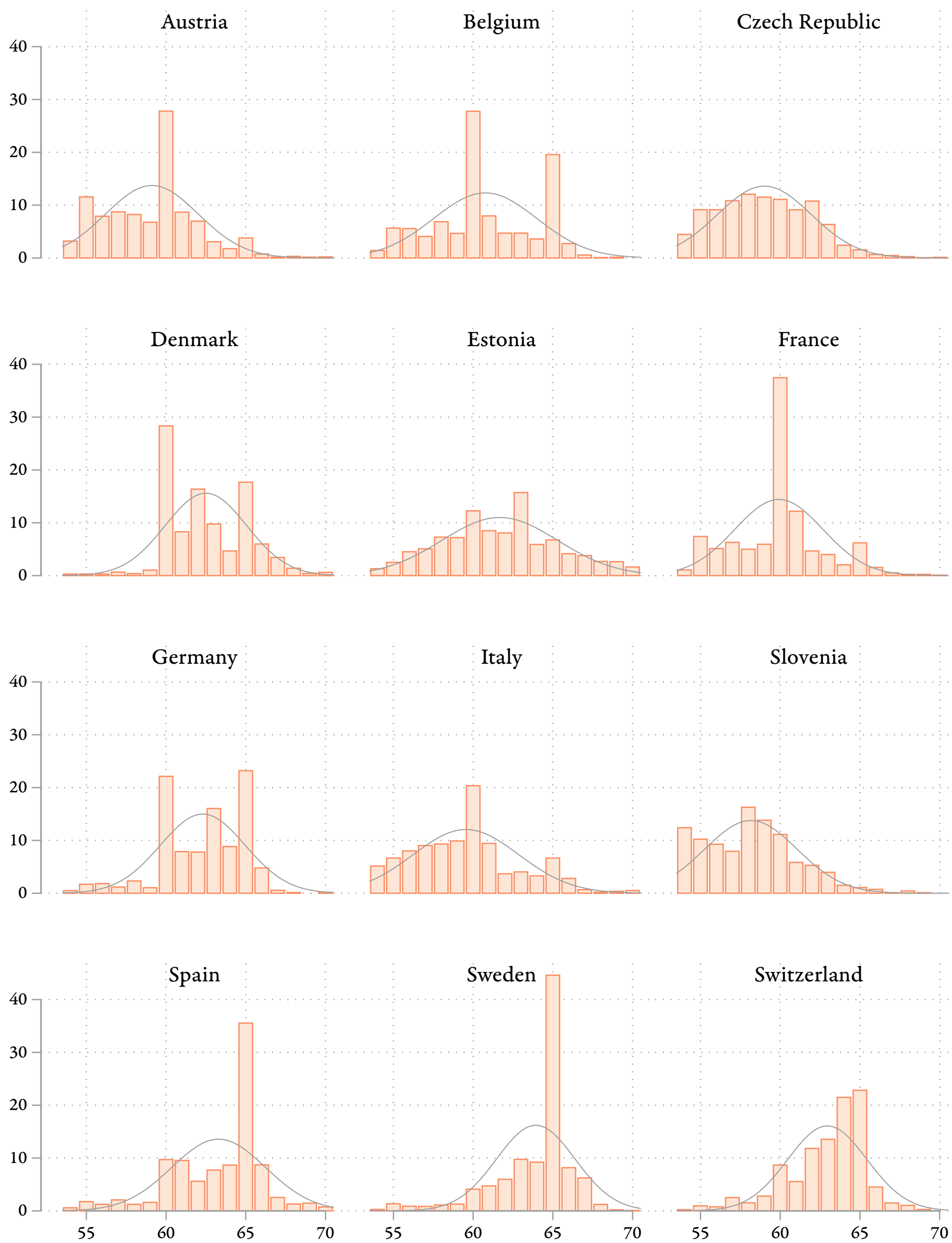
First, I include age controls. Indeed, the ageing process is a prominent cause of cognitive decline. I check for this natural process using a quadratic polynomial of age.

Second, I include a dummy variable that takes the value one when the respondent is living with a spouse or partner in the same household. Bingöl et al. (2016) called this control

[9] Only 2 individuals were dropped due to missing data about retirement year. Still, for about a third of the respondents, retirement year is inconsistent across the three waves. Thus, based on the variable *ep329*, I computed the mode, the value appearing most frequently across waves 1, 2, 4, 5 and 6, to use as the final retirement year. When two years appeared in an equal number of waves, I kept the most recent one.

Figure 1.6

Retirement age distribution by country (percentage by age)



the *hearing spouse* effect. In reality, its total expected effect is ambiguous. Indeed, on the one hand, living alone means having less social interactions on average, which is expected to have a negative impact on cognitive abilities. On the other hand, it might be a cognitive stimulant in that it compels the individual to take care of complex activities related to the maintaining of the household, and this, on its own. Overall, adding this dummy variable allows to control for the important change in lifestyle that living alone implies.

Third, it is important to consider a form of learning effect from retesting. Ferrer, Salthouse, Stewart, and Schwartz (2004) studied various cognitive abilities tests and suggested that both age and retest should be modelled simultaneously when analysing longitudinal data. Indeed, part of the change across measurements may be attributable to practice. Here, the respondents retake the memory test on every wave. Even though the lists of words are different, individuals may learn how to respond to the test and prepare themselves to it. Hence, I include a dummy variable, *retest*, that captures the learning effect, differentiating for respondents that take the test for the first time from others.

Table 1.4
Descriptive statistics

	Variable name	%	Mean	SD
Cognitive abilities:				
Numeracy	*numeracy_2*		4.44	1.08
Orientation in time	*orienti*		3.90	0.35
Verbal fluency	*cf010*		22.88	7.34
Immediate recall	*recall_1*		5.80	1.60
Delayed recall	*recall_2*		4.53	2.05
Memory score	*mem_sc*		10.33	3.35
Retirement variables:				
Retired	*es_retired*	60.2		
Empl. status evolution: working	*es_evol_working*	11.3		
Empl. status evolution: transition	*es_evol_fromWtoR*	34.2		
Empl. status evolution: retired	*es_evol_retired*	54.5		
Age at retirement	*ret_age*		60.17	4.35
Length of retirement	*ret_length*		7.62	5.27
Descriptive and control variables:				
Gender	*female*	52.8		
Age	*age*		63.67	6.41
Living with spouse/partner	*partnerinhh*	76.6		
Already interviewed before	*retest*	72.7		
Education attainment: low	*degree*	28.2		
Education attainment: medium	*degree*	43.4		
Education attainment: high	*degree*	28.4		
Countries and regions:				
Nordic group:	*region*	16.74		
- Denmark	*country*	8.40		
- Sweden	*country*	8.35		
Western Europe:	*region*	41.68		
- Austria	*country*	7.94		
- Belgium	*country*	9.81		
- France	*country*	8.39		
- Germany	*country*	8.77		
- Switzerland	*country*	6.77		
Southern Europe:	*region*	13.29		
- Italy	*country*	6.84		
- Spain	*country*	6.45		
Former Eastern bloc:	*region*	28.28		
- Czech Republic	*country*	12.16		
- Estonia	*country*	10.01		
- Slovenia	*country*	6.11		

CHAPTER 2

Data analysis

This chapter focuses entirely on regression analyses. I use the data described in the previous chapter to estimate the effect of retirement on cognitive functioning. In the first section, I present the empirical approach. In section 2.2, I report the results of the initial model's estimation. The next section assesses the reliability of the instrumental variable approach. Section 2.4 explores wether retirement has an instantaneous effect, or impacts cognitive functioning with a delay. The section progresses trough various steps, and ultimately introduces the final model: first, I introduce an alternative model, then, I justify the length of the delay chosen, and finally I present the final model, that accounts for a delayed effect of retirement on cognition. Then, section 2.5 analyses the sensitivity of the final model to different age functional forms. Finally, section 2.6 evaluates the variability of the results among subgroups of the population. Overall, I find that retirement has a positive effect on cognitive functioning.

2.1 Empirical approach

The empirical analysis aims to test the hypothesis that retirement affects cognitive functioning. I first estimate the following equation:

$$y_{i,t} = \beta(\mathit{Retired}_{i,t}) + f(\mathit{age}_{i,t}) + \gamma X_{i,t} + \mu_i + \varepsilon_{i,t} \tag{2.1}$$

This model assumes that memory score – $y_{i,t}$ – depends on (i) the retirement status – the dummy variable $\mathit{Retired}_{i,t}$ – (ii) a function of age – $f(\mathit{age}_{i,t})$ – and (iii) individual specific characteristics that might affect cognition – $X_{i,t}$. It also depends on an error term that can be decomposed into individual specific unobserved time-invariant heterogeneity – μ_i – and idiosyncratic error – $\varepsilon_{i,t}$. Consequently, the identification of the causal effect of retirement on cognitive functioning relies on the observations of individuals who transit from work to retirement during the sample period. The panel data set includes 2,676 individuals in that situation.

Endogeneity considerations have to be taken into account when considering this model. Indeed, endogeneity can be the source of inconsistent estimation of the parameter β, the causal effect of retirement on cognitive functioning. In such a case, an OLS estimation of equation 2.1 would likely lead to a biased β for two reasons. First, the presence of individual specific unobserved time-invariant heterogeneity would result in a correlation between the retirement decision and the error term, between $\textit{Retired}_{i,t}$ and $\varepsilon_{i,t}$. The second potential source of bias lies in the presence of reverse causality. Indeed, workers that suffer from a decrease in cognitive decline may be selected or select themselves into retirement. This goes against the hypothesis that aims to test if cognitive decline is a result of retirement. In the presence of reverse causality, retirement would be associated with lower cognitive performance, even if it had no effect on cognition.

To circumvent these barriers, I adopt a specific approach. First, I control for individual specific unobserved time-invariant heterogeneity (μ_i) through a fixed effects estimation. Second, I account for the presumed endogeneity of the retirement decision by implementing an instrumental variables strategy. I use the instruments described in section 1.3 – indicators for being above retirement eligibility ages – through the following first-stage equation, where $K_{i,t}$ represents a vector of the instrumental variables:

$$Retired_{i,t} = \delta K_{i,t} + f(age_{i,t}) + \tau X_{i,t} + \rho_i + \epsilon_{i,t} \tag{2.2}$$

In both the first- and second-stage equations, the vector $X_{i,t}$ includes the control variables detailed in section 1.4: *retest*, a dummy variable that captures the learning effect and *partnerinhh*, that captures the effect of living with a partner in the household. Finally, I control for the ageing process by adding function of age that takes the form of a quadratic polynomial of age ($\textit{age} + (\textit{age}/10)^2$).

In conclusion, the empirical strategy consists of a 2SLS specification in fixed effects, with one endogenous variable and two instruments, allowing to control for two sources of endogeneity, unobserved time-invariant heterogeneity and reverse causality.

2.2 Preliminary results

In this section, I estimate the model described above in which the variable memory score – the number of words recalled in the immediate and delayed test – is regressed on the retirement status indicator.

First, the model is simply estimated by fixed-effects (FE). In this case, the average effect of retirement is measured taking account of all the respondents that transited from work to retirement during the sample period – whether they retired smoothly or were incentivised by the retirement schemes and retired more suddenly. Moreover, endogeneity in the retirement decision is not taken into account by the FE estimation, whereas, as explained above, it should raise serious concerns.

Table 2.1
Instrumental variable estimation of the effect of retirement on cognition

	Coefficient	*p-Value*
First stage		
(Dependent variable: retired)		
Above early retirement age	0.142***	0.000
Above statutory retirement age	0.196***	0.000
Second stage		
(Dependent variable: memory score)		
Retired	0.419**	0.028
Statistics		
Under-identification (*p*-Value)	0.245	
Weak identification	103.9	
Over-identification (*p*-Value)	0.495	

Notes: Fixed-effects two-stage least-squares estimates (FE-2SLS). See table A.2 on page 40 for more detailed results and 2-way cluster-robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Thus, the full-fledged model is then estimated by using the two-stage least-squares fixed effects (2SLS-FE) estimation. This estimation is fragmented into two stages. In the first stage (equation 2.2), the endogenous regressor – retirement status – is used as the dependent variable while the instruments are the explanatory variables, along with the age polynomial and the control variables. In the second stage (equation 2.1), memory score is the dependent variable, while the main explanatory variable is the retirement status of each respondent as predicted by the first-stage regression. Consequently, the explanatory variable estimates of this second-stage includes the effect on retirement behaviour induced solely by the national retirement schemes. Besides, this strategy aims to provide an unbiased estimate of the effect of retirement on cognitive decline by accounting for its expected heterogeneous nature across the population. The IV estimates can, in this scenario, be interpreted as the local average treatment effect.

For every analysis, I use cluster-robust standard errors (SE), following suggestions of Wooldridge (2003). SEs and statistics are thus robust to heteroskedasticity and intra-group correlation. Specifically, I use the 2-way cluster-robust SEs proposed by Cameron, Gelbach, and Miller (2011). This provides SEs that are robust to within-panel (within one respondent) autocorrelation and to contemporaneous cross-panel correlation (clustering on wave).

Table 2.1 provides a summary of the full-fledged regression estimation results[1]. According to those, and conditional to the non-linear age specification, the average short-term effect of being retired is positive and statistically significant on cognitive functioning. Notably, the results suggest that retirement increases the memory score by about half a point (the coefficient estimate is 0.419 and the 95% confidence interval ranges from 0.044 to 0.794). This increase of 0.4 words represents an improvement of approximately 4 percent in the average respondent's score.

Finally, as regards the control variables, first, the results show that the chosen age functional form controls for the cognitive ageing process accurately. Further, in section 2.5, I test the robustness of the results by adopting different specifications for age trend and conclude that the quadratic polynomial of age, by capturing the non linearity in age, is adequate to the study. Second, the effect of living with a partner in the household is estimated to be very close to null. Third, the variable that captures the effect of having already been interviewed before, *retest*, proves to display significant effects on memory score. A non-negligible part of the increase in memory score is attributed to practice. The coefficient estimates is significant and its magnitude is more than half of the effect of retirement. Its presence is thus seemingly justified.

2.3 Instruments reliability

The soundness of the instruments rely on two major conditions. First, in order for the instruments to fulfil the relevance condition, they need to have a causal impact on the retirement decision. Here, the set of instruments always pass the F test of excluded instruments in the first stage, individually and jointly. Since I use cluster-robust statistics, I report the LM version of the Kleibergen and Paap (2006) *rk* statistic and the *p*-Value of the under-identification test. The null of under-identification is accordingly rejected. However, the instruments could still be weak. For this reason, I test for weak identification of the model. I report the Wald version of the Kleibergen-Paap rk F statistic. The Stock-Yogo test critical value at 10% is 19.93. I thus reject the null hypothesis of the estimator being weakly identified. In conclusion, the instruments – that indicate whether individuals are eligible for retirement – are correlated with the retirement decision. Those results find evidence to suggest a strong first stage.

Second, the instruments must not be correlated with the error term of equation 2.1. Regarding this validity condition, I analyse the results of the second stage. Since the model has two excluded instruments and one endogenous regressor, I can test for over-identification of the restrictions. Given the assumption of heteroskedasticity and autocorrelation, I report Hansen's J statistic. This specification test leads to not rejecting the null that the instruments are valid and correctly excluded from the estimated equation. Altogether, the instruments meet the diagnostic tests of both relevance and validity.

[1]Table A.2 (page 40) displays the complete results of the two types of estimation. Column 1 shows the simple FE estimation that does not control for endogeneity. Then, columns 2 and 3 show the 2SLS-FE estimation results that account for endogeneity by using the instrumental variable approach.

2.4 Delayed effect?

Previous results suggest that retirement has an instantaneous effect on cognitive functioning. This section explores the prospect of a delayed effect through the adaptation of the initial model.

Atchley (1976) suggested the existence of a pattern of adjustment in retirement across time that reflect the different stages of retirement. Accordingly, the change of lifestyle implied by retirement may not display its effects on cognitive decline immediately. Gall, Evans, and Howard (1997), based in part on Atchley's stage model of retirement, found the impact of retirement on psychological health to be positive in the short term, up to one year post retirement. This confirmed Atchley's suggestion of a honeymoon phase early on in retirement. During the honeymoon stage, due to the release from daily pressures of work, retirees are considered to feel more energetic and satisfied as they pursue desired projects and activities which were earlier put off because of work-related constraints. Bonsang et al. (2012) illustrate this effect and provide support to the hypothesis that the effect of retirement is not instantaneous. They find that the effect of the change in lifestyle on cognition might differ according to the time spent in retirement.

(i) Alternative (lagged) model

Following the strategy of Bonsang et al., and in order to account for this process, I alternatively model memory score as resulting from a function of the lagged environment. Specifically, I define the retirement status indicator as a dummy variable that takes the value one when the respondent has been retired for at least one year and the value zero otherwise. I also compute a new set of instruments, following an analogous procedure of that described above, but increasing the age-threshold by one year, to account for the presumably delayed effect of retirement. This approach allows the model to account for the honeymoon stage of retirement. The estimates are displayed in table 2.2. Under this alternative approach, the effect remains very similar to the one found in section 2.2. Although minimally lower, the impact of retirement on cognitive functioning is positive, inducing an increase of 0.406 points in memory score.

Table 2.2
Instrumental variable estimation of the *delayed* effect of retirement on cognition

	Coefficient	*SE*	*p-Value*
Second stage *(Dependent variable: memory score)*			
Retired for at least one year	0.406**	0.172	0.018

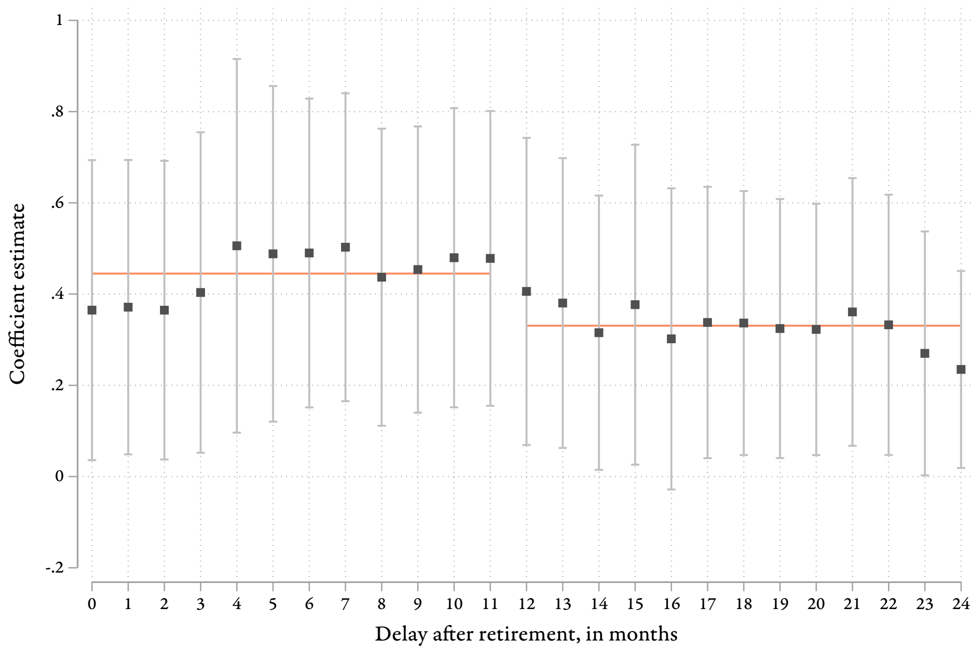
Notes: FE-2SLS estimation. SE : 2-way cluster-robust standard errors.
See table A.3 on page 41 for more detailed results.

(ii) Length of the delay

The introduction of this alternative model raises a new consideration. The model considers a honeymoon effect that lasts for one year. I thus empirically justify the choice of this specific delay between retirement and its effect on memory score. To shed some light on this matter, I estimate the alternative model using different delay periods, ranging from 0 to 24 months. I thus construct a set of new endogenous regressors and instrumental dummy variables for each of the 25 specifications. This allows me to run a sequence of regressions, each accounting for a different delay of retirement on cognitive functioning. I report the estimates of the coefficient of main interest – being retired for at least 0 to 24 months – in figure 2.1. It should be noted that the specification in which the delay is set to 12 months corresponds to the specification estimated in table 2.2.

Figure 2.1

Sensitivity to the length of the delay after retirement



Notes: Markers are the 2SLS-FE coefficient estimates of *being retired for at least 0—24 months*, under the alternative (lagged) model. Vertical bars represent the 95% confidence intervals of the estimates. Horizontal lines show the means, for up-to-one- and two-year delays. Note that the specification in which the delay is set to 12 months is the one estimated in table 2.2.

Figure 2.1 shows that the effect of retirement on cognition is comparatively higher when accounting for delays of up to one year than when the delays are set to longer periods. The horizontal lines represent the 12-month average of the coefficients estimated. When accounting for a delay of up to one year, the effect of retirement on cognitive functioning is comparatively higher than when accounting for longer period delays. These results confirm the gerontological hypothesis of the honeymoon phase, which brings forth a positive boost in cognitive functioning during the first year of retirement that seems to gradually wear off over time. Such a finding substantiates the approach followed and the choice of a 12-month delay. Accordingly, I account for a one year delay in the effect of retirement on cognition in further analyses.

(iii) Second endogenous regressor - final model

Under the alternative (lagged) model, the control group is made of a weighted average of workers and recently retired individuals. In order to isolate the effect of the recently retired individuals from the estimates, I specify a slightly different model that includes an additional dummy variable as an endogenous regressor. *Retired less than 1 year* equals one if the individual is retired for less than one year and zero otherwise. I also construct new instruments to fit into this model. Those are the two dummy variables that indicate if the individual is aged exactly either the early or the statutory retirement age. Under this specification, the control group thus only includes working individuals.

The results of this new estimation are shown in table 2.3. They allow the effects of retirement on cognition to differ between individuals that recently retired and those that retired more than a year ago. The first endogenous regressor – the dummy for being retired for less than one year – captures the effect of retirement that occurs within the year following the retirement decision. This is the immediate effect of retirement. The coefficient estimate is positive but statistically insignificant. Its large standard error does not allow for an accurate interpretation. However, the coefficient estimation of the second endogenous regressor – the dummy for being retired for at least one year – is statistically significant. It captures the average effect of retirement in the longer run, one year post-retirement. Its coefficient estimate is of slightly greater magnitude, but remains very similar to those of anterior specifications. These results suggest a significant positive effect of retirement on cognition. Still, the effect is less likely to be immediate or happen in the very short run.

Table 2.3
Instrumental variable estimation of the *immediate* and *delayed* effects of retirement on cognition

	Coefficient	*SE*	*p-Value*
Second stage			
(Dependent variable: memory score)			
Retired for less than one year	0.220	0.226	0.329
Retired for at least one year	0.450**	0.204	0.027

Notes: FE-2SLS estimation. SE : 2-way cluster-robust standard errors.
See table A.4 on page 42 for more detailed results.

2.5 Sensitivity to age functional forms

In all of the preceding analyses, I have used a quadratic trend to control for the cognitive ageing process. In order to check for the robustness of the results to different age trends, I estimate the model with different functional forms of age as control variables. The results of those various estimations are reported in table 2.4.

Table 2.4
Robustness to age trends

		Age polynomial:			
	Log	Degree 1	Degree 2	Degree 3	Degree 4
Endogenous regressors:					
Retired less than 1y	0.558**	0.597**	0.220	0.156	-0.053
	[0.250]	[0.247]	[0.226]	[0.249]	[0.259]
Retired at least 1y	0.526**	0.621***	0.450**	0.306	0.222
	[0.206]	[0.203]	[0.204]	[0.310]	[0.307]
Age specifications:					
Log of age	-0.487	–	–	–	–
	[0.586]				
Age	–	-0.020**	0.513***	-0.463	-21.274***
		[0.009]	[0.050]	[0.876]	[5.585]
$Age^2/10$	–	–	-0.407***	1.162	51.024***
			[0.036]	[1.425]	[13.161]
$Age^3/100$	–	–	–	-0.083	-5.357***
				[0.076]	[1.370]
$Age^4/1000$	–	–	–	–	0.208***
					[0.053]

Robust standard errors are reported between brackets.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Overall, the coefficient of main interest's estimation is quite insensitive to the variations of age trends. The coefficient estimates of the dummy *retired at least one year* is positive for every different age trend. Still, its magnitude decreases as the order of the polynomial trend increases. With the polynomials of orders 3 and 4, the standard errors become considerably larger and the estimation is not statistically significant. However, it is significant for the other specifications, including the log of age.

When excluding any non-linear term in age (using only the first order polynomial), the coefficients of main interest tend to be overestimated. As noted by Bianchini and Borella (2016), failing to recognise the non-linearity of the average age trend induces a bias in the estimate. Overall, the age trends that best suit the model are the logarithmic and the quadratic ones. The wide use of the latter in the literature motivated my choice to use it as the control for the cognitive ageing process.

2.6 Heterogeneity across subsamples

The empirical results of the previous sections lead me to adopt the approach consisting in a model that includes a quadratic polynomial of age and two endogenous regressors, which allow to control for the delayed effect of retirement on cognition. The estimation results support the hypothesis that retirement, conditional on the non-linear age profile, has a significant positive causal effect on cognitive functioning. Nonetheless, this effect might be heterogenous across group of individuals. This section explores this question.

Figures 1.1, 1.2 and 1.3 (pp. 10-11) show large differences in memory score across groups. However, these differences lie in the level of the score. To investigate the potential heterogeneity of the retirement effect across individuals, I fit the model described above separately with different subsamples of the population. I use the sources of heterogeneity described in section 1.2, namely, gender, education and region. The estimates of the coefficient of main interest, *retired at least one year*, when using only a part of the population to run the regression, are reported in table 2.5.

Table 2.5
Heterogeneity across subsamples of the population

	Coefficient	*SE*	*p-Value*
Level of education			
Low and medium	0.381**	0.192	0.047
High	0.737**	0.337	0.029
Gender			
Men	0.063	0.163	0.699
Women	0.806**	0.353	0.022
Macro region			
Southern Europe	0.724	0.708	0.307
Former Eastern bloc	0.345	0.351	0.326
Nordic countries	0.971**	0.498	0.049
Western Europe †	0.448***	0.171	0.009
- Belgium	*-0.630***	*0.308*	*0.041*

Notes: FE-2SLS estimation of the model with two regressors, ran separately on subsamples of the population. Coefficient of *retired for at least one year*. SE : 2-way cluster-robust standard errors.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$
† Western Europe does not include Belgium, which is estimated separately, due to its peculiar results.

Education

First, when splitting up the sample between lower and higher educated individuals, I find a clear education gradient. Retirement has a significant positive effect on both groups but the magnitude for individuals that completed an education programme of tertiary education (*i.e.* postsecondary education) is double the one of other respondents, with low or medium level of education attainment. This difference could be the result of reasonable yet simplistic hypotheses. For instance, Bingöl et al. (2016) note that higher educated individuals might have a stronger preference for more cognitive stimulating activities. The change of lifestyle entailed by retirement – for example, more free time – would then lead these individuals to engage in more cognitive repair activities and to get involved to a larger extent in social activities. Also, one could consider that the pre-retirement occupations of higher educated people could be more prone to stress that those of lower educated individuals. Leaving a higher-pressure career would ultimately be likely to induce a stronger positive effect of retirement on psychological health, and thus, a larger effect on memory score. Nonetheless, those perspectives remain hypothesis that are open to investigation.

Gender

Further, I run the estimations separately for men and women. While the effect of retirement is not significant for men, the effect on women is very significant and its coefficient is larger than those found before. Additionally, when discriminating women according to their level of education, I still find positive and highly significant effects. Congruently with the results highlighted before, retirement has a larger effect on cognition on higher educated women. Again, those heterogenous reactions to the lifestyle changes triggered by retirement would need further analysis. These results highlight the crucial role of the gender selection process detailed in section 1.1.

Region

Finally, I fit separate models for Southern, Northern, Eastern and Western European countries, and thereby, test if the association between retirement and cognitive decline is markedly different in these heterogenous contexts. No significant effects are found for Southern Europe and former Eastern bloc countries[2].

However, I find significant positive effects for the Nordic countries. The effect of retirement on cognition is much larger than in the global average, reaching a coefficient corresponding to an increase caused by retirement of almost one word in memory score. The estimates for Denmark and Sweden evaluated separately are statistically significant and confirm this result.

[2]In this section, I estimated the model with two regressors across subsamples of the population. Table 2.5 displays the corresponding results. However, as a robustness check, I ran the subsample regressions on the model with one regressor. While all the results approximately coincide, only those of Southern Europe and former Eastern bloc countries widely differ between the two models. I thus conclude that no trend can be identified from these non significant estimates.

Then, the table shows that the estimates for Western European countries are precisely estimated and in line with the overall results found in previous sections. Nonetheless, Western Europe does not include Belgium, which is estimated separately. Indeed, when analysing countries of this group separately, I found that they all share a similar trend that is reflected in the Western Europe estimation. In contrast, the effect of retirement on Belgian respondents, when estimated alone, is negative and statistically significant. While this study does not aim to explain the wide variety of schemes and the plurality of contexts among regions and countries, the estimation results lead to believe that they are largely influenced by the national environment and history of the different regions of Europe.

Policy discussion and conclusion

In this thesis, I estimate the causal effect of retirement on cognitive abilities using data from 13 European countries. I exploit the panel dimension of SHARE to control for time-invariant individual heterogeneity that can affect both the retirement decision and the cognitive abilities. Additionally, I use legal retirement ages as instrumental variables to control for the endogeneity of the retirement decision. The approach is thus unlikely to suffer from reverse causality or self-selection issues.

My results suggest a significant positive effect of retirement on cognitive functioning. Retirement is found to be favorable to episodic memory, measured by word recall. While retirement could, *prima facie*, be supposed to have a positive effect resulting from the so-called honeymoon phase, I show that the estimated effect remains positive when controlling for this short-term upturn, by introducing a one-year delay in the model. Further, I explore different sources of heterogeneity among the sample and show that retirement does not affect everyone to the same extent.

Those findings have important implications in a context of an ageing population. Driven by low fertility rates and increasing life expectancy, population ageing has been one of the main driving forces behind the wave of pension reforms in recent years (OECD, 2017). As a matter of fact, Eurostat (2018) projects the old age dependency ratio to increase from 32 in 2020 to 50 in 2050, in average, for the countries analysed in this study (except Switzerland). This means that for every person aged 65 and over, only 2 persons will be of working age in 2050 (figure A.2 displays the evolution up to 2080 of this indicator by region). This sharp increase in the old age dependency ratio is expected to place additional burdens on the working age population to finance pensions and health care for older people. As a result, many countries have increased or plan to increase their pension benefit withdrawal ages (OECD, 2017).

The results of this analysis suggest that increasing the legal age of retirement has detrimental effects on cognition. Given that memory decline is associated with a higher probability of developing mental diseases and cognitive impairment, the recent reforms may produce negative health externalities and additional expenditures related to long-term care.

While these findings are significant and provide new insights to the question at stake, further research is needed to characterise the mechanisms driving the heterogeneity found in the impact of retirement on cognitive functioning.

References

Adam, S., Bonsang, E., Germain, S., & Perelman, S. (2007). *Retirement and cognitive reserve: A stochastic frontier approach applied to survey data*. CREPP.

Anderson, N. D. & Craik, F. I. (2000). Memory in the aging brain. *The Oxford handbook of memory*, 411–425.

Atchley, R. C. (1976). *The sociology of retirement*. Halsted Press.

Banks, J. & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: Evidence from a regression discontinuity design. *The Economic Journal*, *122*(560), 418–448.

Baum, C., Schaffer, M., & Stillman, S. (2010). Ivreg2: Stata module for extended instrumental variables/2sls, gmm and ac/hac, liml and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html.

Bianchini, L. & Borella, M. (2016). Retirement and memory in europe. *Ageing & Society*, *36*(7), 1434–1458.

Bingley, P. & Martinello, A. (2013). Mental retirement and schooling. *European Economic Review*, *63*, 292–298.

Bingöl, B., Crespo, L., & Mira, P. (2016). *Retirement and cognitive decline: A panel data approach using share*.

Bischof, D. (2017). New graphic schemes for stata: Plotplain and plottig. *Stata Journal*, *17*(3), 748–759.

Bonsang, E., Adam, S., & Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of health economics*, *31*(3), 490–501.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Cavapozzi, D., Garrouste, C., & Paccagnella, O. (2011). Childhood, schooling and income inequality. In *The individual and the welfare state* (pp. 31–43). Springer.

Celidoni, M., Dal Bianco, C., & Weber, G. (2017). Retirement and cognitive decline. a longitudinal analysis using share data. *Journal of health economics*, *56*, 113–125.

Coe, N. B., von Gaudecker, H.-M., Lindeboom, M., & Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health economics*, *21*(8), 913–927.

Coe, N. B. & Zamarro, G. (2011). Retirement effects on health in europe. *Journal of health economics*, *30*(1), 77–86.

Dal Bianco, C., Garrouste, C., & Paccagnella, O. (2013). Early-life circumstances and cognitive functioning dynamics in later life. In *Active ageing and solidarity between generations in europe: First results from share after the economic crisis* (Chap. 18). Walter de Gruyter.

Eurostat. (2016). International standard classification of education (isced). Statistics Explained : http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED).

Eurostat. (2018). Projected old-age dependency ratio. http://ec.europa.eu/eurostat/web/products-datasets/-/tps00200.

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging*, *19*(2), 243.

Fonseca, R., Kapteyn, A., & Zamarro, G. (2016). *Retirement and Cognitive Functioning: International Evidence* (tech. rep. No. 1610). Chaire de recherche Industrielle Alliance sur les enjeux économiques des changements démographiques.

Gall, T. L., Evans, D. R., & Howard, J. (1997). The retirement adjustment process: Changes in the well-being of male retirees across time. *The Journals of Gerontology: Series B*, *52B*(3), P110–P117.

Gruber, S., Hunkler, C., & Stuck, S. (2014). Generating easyshare: Guidelines, structure, content and programming. *SHARE Working Paper Series 17-2014*.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Psychological science in the public interest*, *9*(1), 1–65.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126.

Malter, F. & Börsch-Supan, A. (2013). *Share wave 4: Innovations & methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Mazzonna, F. & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, *56*(4), 691–710.

Mazzonna, F. & Peracchi, F. (2014). Unhealthy retirement? *EIEF Working Paper*.

Meng, A., Nexø, M. A., & Borg, V. (2017). The impact of retirement on age related cognitive decline–a systematic review. *BMC geriatrics*, *17*(1), 160.

Morris, M. C., Evans, D. A., Hebert, L. E., & Bienias, J. L. (1999). Methodological issues in the study of cognitive decline. *American journal of epidemiology*, *149*(9), 789–793.

OECD. (2017). *Pensions at a glance 2017: Oecd and g20 indicators*. OECD Publishing, Paris.

Prull, M. W., Gabrieli, J. D., & Bunge, S. A. (2000). Age-related changes in memory: A cognitive neuroscience perspective.

Rohwedder, S. & Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, *24*(1), 119–38.

Schaffer, M. (2010). Xtivreg2: Stata module to perform extended iv/2sls, gmm and ac/hac, liml and k-class regression for panel data models. http://ideas.repec.org/c/boc/bocode/s456501.html.

Schmitz-Scherzer, R. & Thomae, H. (1983). Constancy and change of behavior in old age: Findings from the bonn longitudinal study on aging. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 191–221). The Guilford Press New York.

Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2012). Does schooling improve cognitive functioning at older ages? *IZA Discussion Paper 6958*.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, *14*(2), 299.

Tosi, M. & Grundy, E. (2018). Returns home by children and changes in parents' well-being in europe. *Social Science & Medicine*.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*(2), 133–138.

Appendices

APPENDIX A

Additional tables and figures

Table A.1
Gender distribution by country in original sample (in %)

	Men	Women	Diff.
Austria	42.3	57.7	15.4
Germany	47.0	53.0	6
Sweden	46.2	53.8	7.6
Spain	44.8	55.2	10.4
Italy	44.9	55.1	10.2
France	42.3	57.7	15.4
Denmark	45.6	54.4	8.8
Switzerland	44.8	55.2	10.4
Belgium	44.2	55.8	11.6
Czech Republic	41.0	59.0	18
Slovenia	42.1	57.9	15.8
Estonia	39.5	60.5	21

Table A.2
Estimations of the effect of retirement on cognition

	Fixed Effects	2SLS–FE	
		1st stage	2nd stage
Endogenous regressor:			
Retired	0.090	–	0.419**
	[0.061]		[0.191]
Instrumental variables:			
Above early retirement age	–	0.142***	–
		[0.024]	
Above statutory retirement age	–	0.196***	–
		[0.018]	
Age polynomial:			
Age	0.545***	0.036***	0.514***
	[0.068]	[0.006]	[0.048]
$Age^2/10$	-0.421***	-0.012***	-0.406***
	[0.052]	[0.004]	[0.035]
Control variables:			
Already interviewed before	0.226***	-0.013**	0.233***
	[0.032]	[0.006]	[0.021]
Living with spouse/partner	-0.059	0.025***	-0.068
	[0.085]	[0.007]	[0.065]
Number of observations	65120	65120	65120
Number of individuals		25031	25031
Under-identification			2.815
p-Value			0.245
Weak identification			103.9
Over-identification			0.465
p-Value			0.495

This table is the complete version of table 2.1 (page 21). Refer to section 2.2 for interpretation.
Robust standard errors are reported between brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3
Estimations of the *delayed* effect of retirement on cognition

	Fixed Effects	2SLS–FE	
		1st stage	2nd stage
Endogenous regressor:			
Retired for at least one year	0.094*	–	0.406**
	[0.056]		[0.172]
Instrumental variables:			
Above early retirement age +1	–	0.159***	–
		[0.025]	
Above statutory retirement age +1	–	0.236***	–
		[0.013]	
Age polynomial:			
Age	0.547***	0.026***	0.528***
	[0.068]	[0.004]	[0.045]
$Age^2/10$	-0.423***	-0.005	-0.418***
	[0.052]	[0.003]	[0.033]
Control variables:			
Already interviewed before	0.225***	-0.001	0.227***
	[0.032]	[0.007]	[0.021]
Living with spouse/partner	-0.058	0.011*	-0.062
	[0.085]	[0.006]	[0.064]
Number of observations	65120	65120	65120
Number of individuals		25031	25031
Under-identification			2.883
p-Value			0.237
Weak identification			267.3
Over-identification			1.368
p-Value			0.242

This table is the complete version of table 2.2 (page 23). Refer to section 2.4 for interpretation.
Robust standard errors are reported between brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4
Estimation of the *immediate* and *delayed* effects of retirement on cognition

	1st stage		2nd stage
	less than 1y	at least 1y	
Endogenous regressor:			
Retired for less than one year	–	–	0.220
			[0.226]
Retired for at least one year	–	–	0.450**
			[0.204]
Above early retirement age +1	0.040***	0.172***	–
	[0.009]	[0.021]	
Above statutory retirement age +1	-0.057***	0.258***	–
	[0.006]	[0.013]	
Exactly on early retirement age	0.041***	0.041***	–
	[0.014]	[0.009]	
Exactly on statutory retirement age	0.135***	0.040***	–
	[0.011]	[0.007]	
Age polynomial:			
Age	0.032***	0.018***	0.513***
	[0.005]	[0.004]	[0.050]
$Age^2/10$	-0.027***	0.001	-0.407***
	[0.003]	[0.003]	[0.036]
Control variables:			
Already interviewed before	-0.011***	-0.000	0.231***
	[0.003]	[0.007]	[0.021]
Living with spouse/partner	0.007	0.010*	-0.064
	[0.007]	[0.006]	[0.065]
Number of observations	65120	65120	65120
Number of individuals	25031	25031	25031
Under-identification			2.978
p-Value			0.395
Weak identification			86.9
Over-identification			1.575
p-Value			0.455

This table is the complete version of table 2.3 (page 25). Refer to section 2.4 for interpretation.
Robust standard errors are reported between brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1

Memory score averages by gender and country

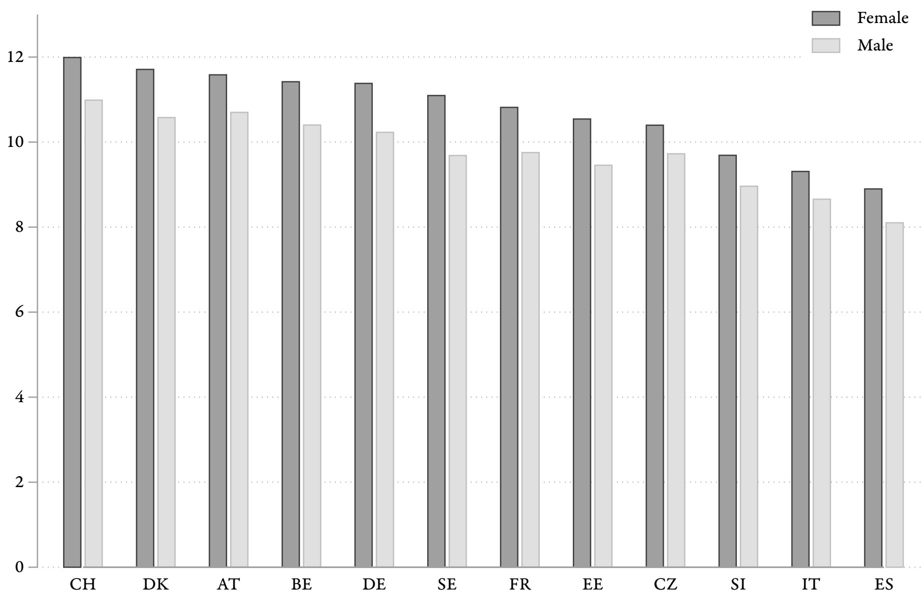
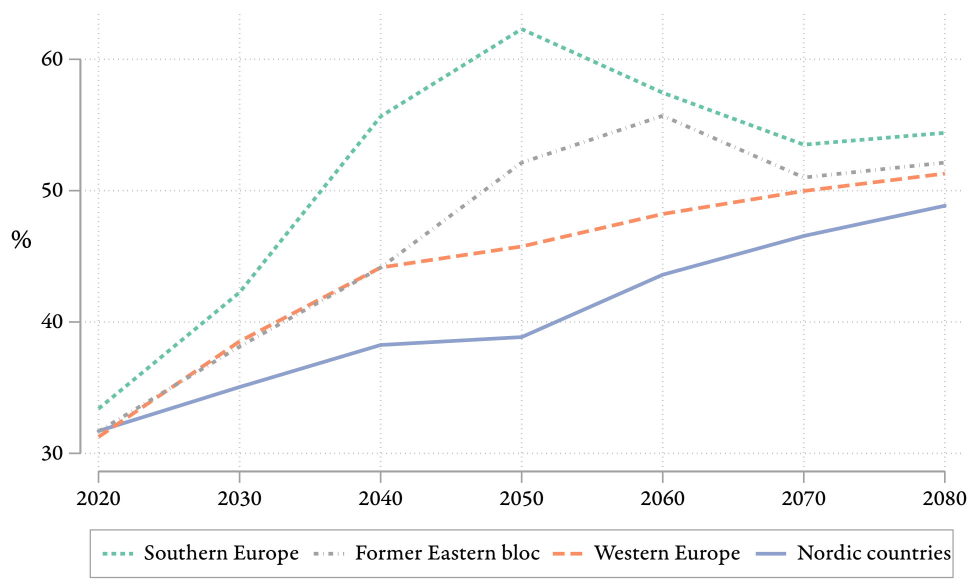


Figure A.2

Old age dependency ratio projections - Population 65 and over to population 15 to 64 years



Notes: The indicator is the ratio between the projected number of persons aged 65 and over (age when they are generally economically inactive) and the projected number of persons aged between 15 and 64. The value is expressed per 100 persons of working age (15-64). Region averages represent simple means of each country of the region used in this study (except Switzerland). Source: Eurostat (2018) [tps00200]

APPENDIX B

Retirement eligibility ages in Europe

The initial sources of information about early and normal retirement eligibility criteria are Gruber and Wise, 2000; Gruber, Milligan, and Wise, 2009 and Wise, 2012. Celidoni, Dal Bianco, and Weber, 2017 have compiled other country specific auxiliary data of which sources are reported below. Further updates are mainly from OECD and Mutual Information System on Social Protection (MISSOC).

Imputation to the individuals from the sample is made differentiating for countries, gender and cohort, when not excessively cumbersome. However, given the lack of reliable information concerning the contribution years, this criteria has been ignored. Methodological details can be consulted in the STATA code, in appendix C.

Austria

Source : Angelini, Brugiavini, and Weber, 2009 and MISSOC

Statutory age From 1961 onwards, 65 for men and 60 for women.

Early age From 1961 to 2001, 60 for men and 55 for women. From 2002 to 2004, 61 for men and 56 for women. From 2005 onwards, 62 for men and women.

Belgium

Source : Jousten, Lefèbvre, Perelman, and Pestieau, 2010, Angelini et al., 2009

Statutory age 65 for men. For women 60 until 1996, 61 from 1997 to 1999, 62 from 2000 to 2002, 63 from 2003 to 2005, 64 from 2006 to 2008, 65 from 2009.

Early age From 1961 to 1966, no retirement age. From 1967 to 1986, 60 for men and 55 for women. From 1987 to 1997, 60 for men and 60 for women. From 1998 to 2007, 60 conditional to years of contributions both for men and women. Then, for both men and women, in 2013, 60,5 and 38 years of contributions, in 2014, 61 and 39 years of contributions and in 2015, 61.5 and 40 years of contributions.

Czech Republic

Source : SSA and MISSOC

Statutory age For men born before 1936, 60. For women born before 1936, 57 if no children, 56 if one child, 55 if two children, 54 if three or four children and 53 if five or more children. For people born between 1936 and 1977, the retirement age threshold rises steadily but differs for every category. I thus take the current age, 63 for men and 62 for women.

Early age 60 and 33 years of contribution for men. 59 and 33 years of contribution for women.

Denmark

Source : Angelini et al., 2009

Statutory age From 1961 to 2003, 67 both for men and women. From 2004 to 2019, 65 both for men and women.

Early age From 1961 to 1975, no early retirement. From 1976 to 1978, 60 for both men and women. From 1979 onwards, 60 with 25 years out of the last 30 years of contributions.

Estonia

Source : SSA

Statutory age Age 63 with at least 15 years of contributions

Early age Age 60 with at least 15 years of contributions

France

Source : Angelini et al., 2009, Hamblin, 2013

Statutory age Until 1982, 65 both for men and women. From 1983 to 2010, 60 both for men and women. From 2011, 60 for those born until 1952, 61 if born in 1953 or 1954 and 62 for those born since 1955.

Early age No early retirement until 1963. 60 from 1963 to 1980. 57 from 1981 onwards.

Germany

Source : Angelini et al., 2009 and Boersch-Supan and Jürges, 2011

Statutory age 65 both for men and women until 2012. Then gradually rising by one month a year until 2024 and by two months a year until reaching age 67 in 2029.

Early age For men, no early retirement until 1972, 60 with 15 years of contributions from 1973 until 2003, 63 from 2004 onwards. For women, no early retirement in 1961, 60 with 15 years of contributions from 1962 until 2003, 62 from 2004 until 2005, 63 from 2006.

Italy

Source : Celidoni et al., 2017

Statutory age from 1961 to 1993, 60 (65 in the public sector) for men and 55 (60 in the public sector) for women; in 1994, 61 for men and 56 for women; in 1995, 61.5 for men and 56.5 for women; in 1996, 62 for men and 57 for women; in 1997, 63 for men and 58 for women; in 1998, 63.5 for men and 58.5 for women; in 1999, 64 for men and 59 for women; from 2000 to 2011, 65 for men and 60 for women (both private and public sector). In 2013, 65 years and 1 month, then rising by a month every year.

Early age from 1965 to 1995, early retirement was possible at any age with 35 years of contributions (25 in the public sector) for both men and women; from 1996 it was stepwise increased up to 57 for both the private and public sector. It is 63 in 2017.

Slovenia

Source : MISSOC and OECD

Statutory age Until 2013, for women, 63 with 15 years of contributions and for men, 65 with 15 years of contributions. Then, 65 for both men and women.

Early age For women, 61 with 20 years of contributions and 58 after 38 years of contributions. For men, 63 with 20 years of contributions and 58 after 40 years of contributions.

Spain

Source : Angelini et al., 2009

Statutory age 65 both for men and women.

Early age 64 until 1982, 60 from 1983 to 1993, 61 from 1994 onwards, for both men and women. From 2002 for both men and women 61 with 30 years of contributions.

Sweden

Source : Mazzonna and Peracchi, 2014

Statutory age 67 for both men and women until 1994, 65 from 1995 onward.

Early age No early retirement until 1962, 60 from 1963 to 1997, 61 from 1998 onwards.

Switzerland

Source : Dorn and Sousa-Poza, 2003 and Angelini et al., 2009

Statutory age 65 for men. For women 63 until 1963, 62 from 1964 until 2000, 63 from 2001 to 2004, 64 from 2005.

Early age No early retirement until 1996 for men and until 2000 for women. Then, 64 for men from 1997 until 2000 and 63 from 2001, for women 62 from 2001.

Additional references for retirement ages

Angelini, V., Brugiavini, A., & Weber, G. (2009). Ageing and unused capacity in europe: Is there an early retirement trap? *Economic Policy*, *24*(59), 463–508.

Boersch-Supan, A. H. & Jürges, H. (2011). *Disability, pension reform and early retirement in germany*. National Bureau of Economic Research.

Celidoni, M., Dal Bianco, C., & Weber, G. (2017). Retirement and cognitive decline. a longitudinal analysis using share data. *Journal of health economics*, *56*, 113–125.

Dorn, D. & Sousa-Poza, A. (2003). Why is the employment rate of older swiss so high? an analysis of the social security system. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *28*(4), 652–672.

Gruber, J., Milligan, K., & Wise, D. (2009). *Social security programs and retirement around the world: The relationship to youth employment, introduction and summary*. National Bureau of Economic Research.

Gruber, J. & Wise, D. (2000). Social security programs and retirement around the world. In *Research in labor economics*. University of Chicago Press.

Hamblin, K. (2013). *Active ageing in the european union: Policy convergence and divergence*. Springer.

Jousten, A., Lefèbvre, M., Perelman, S., & Pestieau, P. (2010). The effects of early retirement on youth unemployment: The case of belgium. In *Gruber, j. and wise, d., social security programs and retirement around the world. the relationship to youth employment. chicago*. University of Chicago Press.

Mazzonna, F. & Peracchi, F. (2014). Unhealthy retirement? *EIEF Working Paper*.

Mutual Information System on Social Protection. (2017). Comparative tables database. http://www.missoc.org/.

OECD. (2017). *Pensions at a glance 2017: Oecd and g20 indicators*. OECD Publishing, Paris.

Wise, D. A. (2012). *Social security programs and retirement around the world: Historical trends in mortality and health, employment, and disability insurance participation and reforms*. University of Chicago Press.

APPENDIX C

Stata code

This Stata program:

— extracts the data from SHARE modules and merges it to easySHARE database,

— selects the population and cleans the data,

— generates the variables and information used in the study,

— creates tables and graphs,

— runs the regressions and outputs their results.

The source of the data is the SHARE Project, release 6.0.0.

```
/*****************************************************************************

This code is related to the master thesis:

                     RETIREMENT AND COGNITIVE FUNCTIONING:
                            A LONGITUDINAL ANALYSIS

This Stata program:

    - Extracts the data from SHARE modules and merges it to easySHARE database
    - Selects the population and cleans the data
    - Generates the variables and information used in the study
    - Creates tables and graphs
    - Runs the regressions and outputs their results

The source of the data is the SHARE Project, release 6.0.0
see http://www.share-project.org


Author:  Martin Habets
Date:    June 2018


*****************************************************************************/


*----[ O v e r v i e w   o f   C o n t e n t s ]--------------------------line-

*----[ 0. Stata Version & Settings]---------------------------------------- 75 -
*----[ 1. Define paths and open log file]---------------------------------- 86 -

*----[ 2. Select easySHARE data]------------------------------------------- 120-
*----[ 3. Extract variables from different modules]------------------------ 138-
*----[ 4. Merge modules per wave]------------------------------------------ 194-
*----[ 5. Append waves to form a panel (long format)]---------------------- 213-
*----[ 6. Merge easySHARE and modules data]-------------------------------- 228-
*----[ 7. Modification of the wave participation pattern]------------------ 240-
*----[ 8. Recode, format and label variables]------------------------------ 259-

*----[ 9. Save the main database]------------------------------------------ 317-

*----[10. Select information collected in wave 4, 5 and 6 only]------------ 325-
*----[11. Reshape the panel to wide]---------------------------------------- 338-
*----[12. Select the population]-------------------------------------------- 363-
*----[13. Retirement-related variables and selection]---------------------- 450-
*----[14. Treat additional invariant variables]---------------------------- 547-
*----[15. Reshape database from wide to long]------------------------------ 614-

*----[16. Generate the endogenous regressors]------------------------------ 641-
*----[17. Integrate legal retirement ages]---------------------------------- 687-
*----[18. Generate the instrumental variables]----------------------------- 844-
*----[19. Treat missing variables]----------------------------------------- 919-

*----[20. Save final database]--------------------------------------------- 953-

*----[21. Descriptive tables]----------------------------------------------- 961-

*----[22. Graphs]-----------------------------------------------------------1178-

*----[23. Regressions - Setting up global variables]-----------------------1410-
*----[24. Regressions - Preliminary model]---------------------------------1429-
*----[25. Regressions - Alternative model - one regressor]-----------------1448-
*----[26. Regressions - Retired at least 1--24 months]---------------------1467-
*----[27. Regressions - Alternative model - two regressors]----------------1516-
*----[28. Regressions - Sensitivity to age polynomials]--------------------1533-
*----[29. Regressions - Heterogeneity across subsamples]-------------------1555-

*----[30. Exit Stata & Close Log File]-------------------------------------1600-
```

```

*******************************************************************************/

*------------------------------------------------------------------------------

*----[ 0. Stata Version & Settings]--------------------------------------------

version 14.1
clear
clear matrix
set more off

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 1. Define paths and open log file]--------------------------------------

*>> Define location of data and working directory

    global SHARE_data "/Users/Martin/Documents/SHARE_data/"
    global easy "/Users/Martin/Documents/SHARE_data/easySHARE_rel6-0-0.dta"
    global wd "/Users/`c(username)'/Dropbox/University/THESIS/SHARE/STATA"
*>> Define hierarchy of subfolders within the working directory

    global modules "$wd/modules"
    global temp "$wd/temp"
*>> Generate data folders within the pre-existing working directory:

    cd $wd                          // change to directory stored above
    capture mkdir "log"             // log  folder
    capture mkdir "modules"         // modules folder
    capture mkdir "temp"            // to store temporary data versions
    capture mkdir "graphs"          // graphs folder
    capture mkdir "tables"          // tables folder
*>> Name and open a log file and store in the log directory

    capture log close
    local h =  substr("`c(current_time)'",1,2) // These commands
    local m =  substr("`c(current_time)'",4,2) // are only to
    local s =  substr("`c(current_time)'",7,2) // automatically
    local d = "`c(current_date)'"              // generate the
    local u = "`c(username)'"                  // name of the log file

    log using "$wd/log/LOG_`u'__`d'_`h'-`m'-`s'.log", replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 2. Select easySHARE data]-----------------------------------------------

*>> Keep only used variables

use $easy, clear

    drop if wave==3

    keep mergeid hhid wave wavepart int_year int_month country      ///
         female dn002_mod dn003_mod age isced1997_r partnerinhh ///
         recall_1 recall_2 orienti numeracy_1 numeracy_2        ///
         ep005_ eurod ch007_km
save $temp/easyclean, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 3. Extract variables from different modules]----------------------------
```

```
*>> module CV_R:

foreach i in 1 2 4 5 6 {

    use "$SHARE_data/sharew`i'_cv_r.dta", clear

        gen wave=`i'
        drop if age_int=="Not applicable":age
        //drops people that were not interviewed but only helped
        rename yrbirth birth_year
        keep mergeid wave firstwave birth_year mobirth

    save $modules/cv_r_w`i', replace
}
*>> module EP:

foreach i in 1 2 4 5 6 {

    use "$SHARE_data/sharew`i'_ep.dta", clear

        isvar ep006 ep049 ep050 ep210 ep328 ep329 ep649
        keep mergeid `r(varlist)'
        ds
        capture rename ep649 ep049
        tab ep049, m

    save $modules/ep_w`i'.dta, replace
}
*>> module CF:

foreach i in 1 2 4 5 6 {

    use "$SHARE_data/sharew`i'_cf.dta", clear

        keep mergeid cf018d1 cf010_

    save $modules/cf_w`i'.dta, replace
}
*>> module IV:

foreach i in 1 2 4 5 6 {

    use "$SHARE_data/sharew`i'_iv.dta", clear

        keep mergeid iv002d1 iv003_

    save $modules/iv_w`i'.dta, replace
}
*>> module gv_weights:


*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------

*----[ 4. Merge modules per wave]--------------------------------------------

*>> We use the CV_R modules as master and then merge the other modules.

foreach i in 1 2 4 5 6 {

    use $modules/cv_r_w`i', clear

        merge 1:1 mergeid using $modules/ep_w`i', assert(3) nogen
        merge 1:1 mergeid using $modules/cf_w`i', assert(3) nogen
        merge 1:1 mergeid using $modules/iv_w`i', assert(3) nogen

    save $temp/wave`i'_merged, replace
```

```
}
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 5. Append waves to form a panel (long format)]--------------------------

*>> Append single wave files to one long file:

    use       $temp/wave1_merged.dta, clear
    app using $temp/wave2_merged.dta
    app using $temp/wave4_merged.dta
    app using $temp/wave5_merged.dta
    app using $temp/wave6_merged.dta
    save $temp/imported_all, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 6. Merge easySHARE and modules data]------------------------------------

*>> Merge 1:1
    use $temp/easyclean, clear
    summ
    merge 1:1 mergeid wave using $temp/imported_all, assert(3) nogen
    summ

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 7. Modification of the wave participation pattern]----------------------

*>> Modify to keep into account only waves 4, 5 and 6

    clonevar origwavepart = wavepart
    tostring wavepart, replace
    replace wavepart = subinstr(wavepart, "1", "",.)
    replace wavepart = subinstr(wavepart, "2", "",.)
    replace wavepart = subinstr(wavepart, "3", "",.)
    destring wavepart, replace
    la var wavepart "Wave participation pattern for waves 4,5 and 6 only"

    clonevar origfirstwave = firstwave
    replace firstwave=4
    replace firstwave=5 if wavepart==5 | wavepart==56
    replace firstwave=6 if wavepart==6

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 8. Recode, format and label variables]----------------------------------

*>> Recode and format variables
* partnerinhh

    recode partnerinhh (1 = 1) (3 = 0)

* retest effect

    mark retest if firstwave>origfirstwave
    replace retest=1 if wave!=firstwave
    la var retest "Already interviewed before"

* age functional forms

    gen agesq = ((age)/10)^2
```

```
    la var agesq "Age$^2/10$"
    gen agethree = ((age)/10)^3
    la var agethree "Age$^3/100$"
    gen agefour = ((age)/10)^4
    la var agefour "Age$^4/1000$"

    gen ageround = round(age)
    la var ageround "Age round"

* ch007_km

    recode ch007_km (1 = 1) (-9 5 = 0)

*>> birth_month
*   same procedure as for ret_month
    egen birth_month = mode(mobirth), by(mergeid) maxmode
    drop mobirth
    replace birth_month=dn002_mod if birth_month==.

*>> ret_year
    egen ret_year = mode(ep329), by(mergeid) maxmode

*>> ret_month
*   Generate new month of birth variable, taking the mode of all info
*   I do not set the self-report to missing if it deviates between
*   waves, instead I take the minimum modus answer
    egen ret_month = mode(ep328), by(mergeid) minmode
    drop ep328*


*>> Labels
    la val birth_month month3
    la val ret_month month3
    la var age "Age"
    la var ch007_km "A child lives less than 1km away"

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[ 9. Save the main database]----------------------------------------------

*>> Create complete_long
    sort  mergeid wave
    order mergeid wave firstwave wavepart
    save temp/complete_long, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[10. Select information collected in wave 4, 5 and 6 only]----------------

*>> Create selected_long
    keep if wave==4 | wave==5 | wave==6
    save temp/selected_long, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[11. Reshape the panel to wide]-------------------------------------------

*>> Reshape from long to wide

reshape wide                                                        ///
        hhid int_year int_month age*                                ///
        isced1997_r partnerinhh                                     ///
        recall_1 recall_2 orienti numeracy_1 numeracy_2             ///
```

```
        ep* cf* iv* eurod retest ch007_km                           ///
        , i(mergeid) j(wave)

order   mergeid wave firstwave wavepart                             ///
        country female birth_* age* int_*                           ///
        dn002_mod dn003_mod                                         ///
        isced1997_r* partnerinhh*                                   ///
        recall_1* recall_2* orienti* numeracy_1* numeracy_2*        ///
        ep005_* ep049_* ep329_* ep*                                 ///
        cf018* iv* eurod* ch007_km*

        tab wavepart firstwave

*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------

*----[12. Select the population]----------------------------------------------

*   We have 62,388 respondents.
*   In the study, we restrict the sample according to some criterias.

*>> Country participated in all three waves
*   http://www.share-project.org/data-documentation/waves-overview.html

    tab country, p m
    drop if   country == "14. Netherlands":country      ///
            | country == "19. Greece":country           ///
            | country == "25. Israel":country           ///
            | country == "29. Poland":country           ///
            | country == "31. Luxembourg":country       ///
            | country == "32. Hungary":country          ///
            | country == "33. Portugal":country         ///
            | country == "47. Croatia":country

*>> Individual participated to at least two of the waves 4, 5 and 6

    drop if wavepart==4 | wavepart==5 | wavepart==6
    tab wavepart


*>> Individual participated in wave 6

    drop if wavepart==45
    tab wavepart


*>> Individual is at least 50 at first itw and less than 75 at last itw

    drop if age4<50 & firstwave==4
    drop if age5<50 & firstwave==5
    drop if age6>=76
    tab ageround6

// At this point, 34,597 individuals are included the sample.

*>> Employment status
*   We drop individuals who are neither retired nor employed/self-employed
//  or who do not provide information about their employment status.

    foreach i in 4 5 6 {
        tab ep005_`i', m
        drop if  ep005_`i'==-15     ///     // No information
                |ep005_`i'==-12     ///     // Don't know / Refusal
                |ep005_`i'== 3      ///     // Unemployed
                |ep005_`i'== 4      ///     // Permanently sick or disabled
                |ep005_`i'== 5      ///     // Homemaker
                |ep005_`i'== 97             // Other
        tab ep005_`i', m
```

```
    }
    // we drop 3,925 + 2,767 + 1,271 observations


*   We exclude individuals who return to the labor force after being retired

    tab     ep005_4 ep005_5
    drop if ep005_4==1 & ep005_5==2
    //we drop 134 obs

    tab     ep005_5 ep005_6
    drop if ep005_5==1 & ep005_6==2
    //we drop 215 obs

    tab     ep005_4 ep005_6
    drop if ep005_4==1 & ep005_6==2
    //we drop 12 obs

*>> Availability of cognitive tests

    foreach k in 1 2 {
        foreach i in 4 5 6 {
            tab recall_`k'`i'
            drop if recall_`k'`i'==-15
        }
    }
    //we drop 176 + 340 + 457 + 4 + 21 + 11


// At this point, 25,264 individuals are included the sample.

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[13. Retirement-related variables and selection]--------------------------

*>> Employment status

    gen  es_fromWtoR_4=.
    mark es_fromWtoR_5 if ep005_4==2 & ep005_5==1
    mark es_fromWtoR_6 if ep005_5==2 & ep005_6==1
    replace es_fromWtoR_6=1 if ep005_4==2 & ep005_6==1 & wavepart==46

    gen  es_fromWtoR_overall = es_fromWtoR_5 + es_fromWtoR_6
    la var es_fromWtoR_overall "Transited during the sample period"

    foreach i in 4 5 6 {
        recode ep005_`i' (1=1) (2=0), gen(es_retired_`i')
        recode ep005_`i' (1=0) (2=1), gen(es_working_`i')
        la var es_retired_`i' "Retired"
    }

    gen     es_evol = 1
    // "everyone" retired
    replace es_evol = 2 if ep005_6==2
    // if working in 6, they could only be working before (selection)
    replace es_evol = 3 if es_fromWtoR_overall==1
    la var es_evol "Employment status evolution over waves 4, 5 and 6"
    lab def es_evol 1 "Retired" 2 "Working" 3 "From Working to Retired"
    lab val es_evol es_evol
    mark es_evol_retired if es_evol==1
    mark es_evol_working if es_evol==2
    mark es_evol_fromwtor if es_evol==3


*>> Retirement age

*   I generate new retirement year variable from
```

```
*   variable ep050 : Year last job ended
*   variable ep329 : Retirement year

*   check for consistency across waves
    mark check1 if ep329_6!=ep329_5 & ep329_6!=. & ep329_5!=.
    mark check2 if ep329_6!=ep329_4 & ep329_6!=. & ep329_4!=.
    mark check3 if ep329_5!=ep329_4 & ep329_5!=. & ep329_4!=.
    mark check if check1==1 | check2==1 | check3==1

    tab check   //almost 35pc are inconsistent across waves
    drop check*

*   retirement year
    tab ret_year if es_evol!=2, m

*   I drop individuals for whom we do not know the retirement year
    drop if ret_year==. & es_evol!=2
    //we drop 859 observations


*   I compute the age of retirement. For the respondents with available year
*   pf ret but missing month of birth, I assume they are retired in June

    tab ret_month if ret_year!=. , m

    gen     ret_age= ( (ret_year * 12 + ret_month) -                ///
                    (birth_year * 12 + birth_month) ) / 12 if ret_month!= .

    replace ret_age = ( (ret_year * 12 +      6    ) -              ///
                    (birth_year * 12 + birth_month) ) / 12 if ret_month==.

    gen ret_age_round=round(ret_age)

    drop if ret_age<45
    //we drop 573 obs


*>> ret_length

    tab ret_year  if es_evol!=2, m
    tab ret_month if es_evol!=2, m

    foreach i in 4 5 6 {
    gen     ret_length`i'= ((int_year`i'* 12 + int_month`i') -               ///
                            (ret_year   * 12 + ret_month)) /12 if ret_month!= .

    replace ret_length`i'= ((int_year`i'* 12 + int_month`i') -            ///
                           (ret_year    * 12 +     6     )) /12 if ret_month== .
    }

    replace ret_length4=. if ret_length4<0
    replace ret_length5=. if ret_length5<0
    replace ret_length6=0 if ret_length6<0
    //if the respondent retired after the interview date, it is set to missing
    //for wave 6, due to imprecisions from missing ret_month, it is set to 0


// Final sample includes 25,031 individuals

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[14. Treat additional invariant variables]--------------------------------

*>> low cognition
*   dummy that takes the value one if the baseline memory score is lower than
*   the median value by wave, country and gender

```

```
*   memory_score

    foreach i in 4 5 6 {
        gen mem_sc_`i' = recall_1`i' + recall_2`i'
        tab mem_sc_`i'
        la var mem_sc_`i' "Memory score total (/20)"
    }

    statsby mediancog=r(p50), ///
    by(country female firstwave)  sa(temp/mediancog, replace) : ///
        summ mem_sc_4 mem_sc_5, detail

    merge m:1 country female firstwave using temp/mediancog, nogen
    la var mediancog "Median value of baseline mem_sc by country and gender"
    sort mergeid
    order mediancog, b(mem_sc_4)

    gen lowcog=0
    replace lowcog=1 if mem_sc_4<mediancog & firstwave==4
    replace lowcog=1 if mem_sc_5<mediancog & firstwave==5
    la var lowcog "baseline mem_sc lower than median by country and gender"
*   la val lowcog dummi

    order lowcog*, a(mediancog)

*>> less repetitions
*   identifies individuals who performed the test only twice because
*   they enter the sample in wave 5
*   or they were interviewed in wave 4 and wave 6 but not in wave 5

    mark lessrepet if wavepart!=456


*>> educ

    clonevar educ = isced1997_r6 // r6 has no missing and is most up to date
    drop isced1997*
    la var educ "Education of respondent in ISCED-97 Coding"
    mark highschool if educ==3 | educ==4
    mark college    if educ==5 | educ==6

*>> degree
    gen degree=educ
    recode degree (-15 -12 95 96 97 = .) (0 1 2 = 0) (3 4 = 1) (5 6 = 2)
    la var degree "Highest degree obtained based on ISCED"
    la def degree 0 "0. Low education" 1 "1. Medium education" 2 "2. High education"
    la val degree degree

*>> regions

    gen region=country
    // macro regions
    // 1. Nordic group = Denmark Sweden
    // 2. Western European group = Austria Belgium France Germany Switzerland
    // 3. Southern European group = Italy Spain
    // 4. former Eastern bloc cluster =  Czech_Republic Slovenia Estonia

    recode region (13 18 = 1) (11 12 17 20 23 = 2) (15 16 = 3) ( 28 34 35 = 4)
    la def region 1 "1. Nordic group"  2 "2. Western Europe"             ///
                  3 "3. Southern Europe" 4 "4. Former Eastern bloc"
    la val region region


*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------

*----[15. Reshape database from wide to long]-----------------------------------
```

```

*>> Reshape final_wide to and save final_long
save final_wide, replace

    unab vars : *4
    local stubs : subinstr local vars "4" "", all
    reshape long `stubs'         , i(mergeid) j(wave)

    drop if int_year==.          //drop artificially created observations

    encode mergeid, gen(id)
    encode hhid, gen(hholdid)

    xtset id wave


*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------

*----[16. Generate the endogenous regressors]-----------------------------------

*>> Retired

    tab es_retired_
    lab var es_retired_ Retired


*>> Retired at least one year

    mark es_ret_1year if ret_length>=1 // & es_retired==1
    replace es_ret_1year=0 if ret_length==.
    la var es_ret_1year "Retired for at least one year"


*>> Retired less than one year

    mark es_ret_less1y if ret_length<=1 //es_ret_1year+es_retired==1
    la var es_ret_less1y "Retired for less than one year"


*>> Retired at least/less than ... months

tab ret_length if ret_length<=2

forval m = 0/24 {

    mark es_ret_al`m'm if ret_length>=(`m'/12)-0.01
    replace es_ret_al`m'm=0 if ret_length==.
    la var es_ret_al`m'm "Retired for at least `m' months"

    mark es_ret_lt`m'm if ret_length<=(`m'/12)-0.01
    la var es_ret_lt`m'm "Retired less than `m' months"
    }


*>> Retirement duration

    gen ret_duration=0
    replace ret_duration = age-ret_age if ret_age < age


*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------

*----[17. Integrate legal retirement ages]--------------------------------------

// Early and statutory retirement ages differ among countries, gender and
```

```
// cohorts. This section assigns each observation their corresponding tresholds

*>> Generate the variables ret_sr and ret_er
    gen ret_sr=.
    la var ret_sr "Legal statutory retirement age"

    gen ret_er=.
    la var ret_er "Legal early retirement age"


*>> 11. Austria

    replace ret_sr=65 if country==11 & female==0
    replace ret_sr=60 if country==11 & female==1

    replace ret_er=60 if country==11 & female==0
    replace ret_er=55 if country==11 & female==1
    replace ret_er=61 if country==11 & ret_year>=2002 & female==0
    replace ret_er=56 if country==11 & ret_year>=2002 & female==1
    replace ret_er=62 if country==11 & ret_year>=2005
*>> 23. Belgium

    replace ret_sr=65 if country==23 & female==0
    replace ret_sr=60 if country==23 & female==1
    replace ret_sr=61 if country==23 & female==1 & ret_year>=1997
    replace ret_sr=62 if country==23 & female==1 & ret_year>=2000
    replace ret_sr=63 if country==23 & female==1 & ret_year>=2003
    replace ret_sr=64 if country==23 & female==1 & ret_year>=2006
    replace ret_sr=65 if country==23 & female==1 & ret_year>=2009

    replace ret_er=60 if country==23 & female==0
    replace ret_er=55 if country==23 & female==1
    replace ret_er=60 if country==23 & ret_year>=1987
    replace ret_er=60.5 if country==23 & ret_year>=2013
    replace ret_er=61 if country==23 & ret_year>=2014
    replace ret_er=60.5 if country==23 & ret_year>=2015
*>> 28. Czech Republic

    replace ret_sr=60 if country==28 & female==0
    replace ret_sr=55 if country==28 & female==1
    replace ret_sr=63 if country==28 & female==0 & birth_year>1936
    replace ret_sr=62 if country==28 & female==1 & birth_year>1936

    replace ret_er=60 if country==28 & female==0 & birth_year>1936
    replace ret_er=59 if country==28 & female==1 & birth_year>1936
*>> 18. Denmark

    replace ret_sr=67 if country==18
    replace ret_sr=65 if country==18 & ret_year>=2004

    replace ret_er=60 if country==18 & ret_year>=1976
    replace ret_er=60 if country==18 & ret_year>=1979

*>> 35. Estonia

    replace ret_sr=63 if country==35

    replace ret_er=60 if country==35


*>> 17. France

    replace ret_sr=65 if country==17
    replace ret_sr=60 if country==17 & ret_year>=1983
    replace ret_sr=61 if country==17 & ret_year>=2011 & birth_year>=1953
    replace ret_sr=62 if country==17 & ret_year>=2011 & birth_year>=1955

    replace ret_er=60 if country==17
    replace ret_er=58 if ret_year>=1981 & country==17
```

```
*>> 12. Germany

    replace ret_sr=65 if country==12
    replace ret_sr=65.08 if country==12 & ret_year==2013
    replace ret_sr=65.17 if country==12 & ret_year==2014
    replace ret_sr=65.25 if country==12 & ret_year==2015

    replace ret_er=60 if country==12 & female==0 & ret_year>=1973
    replace ret_er=63 if country==12 & female==0 & ret_year>=2004

    replace ret_er=60 if country==12 & female==1 & ret_year>=1962
    replace ret_er=62 if country==12 & female==1 & ret_year>=2004
    replace ret_er=63 if country==12 & female==1 & ret_year>=2006


*>> 16. Italy

    replace ret_sr=60 if country==16 & female==0
    replace ret_sr=55 if country==16 & female==1
    replace ret_sr=61 if country==16 & female==0 & ret_year==1994
    replace ret_sr=56 if country==16 & female==1 & ret_year==1994
    replace ret_sr=61.5 if country==16 & female==0 & ret_year==1995
    replace ret_sr=56.5 if country==16 & female==1 & ret_year==1995
    replace ret_sr=62 if country==16 & female==0 & ret_year==1996
    replace ret_sr=57 if country==16 & female==1 & ret_year==1996
    replace ret_sr=63 if country==16 & female==0 & ret_year==1997
    replace ret_sr=58 if country==16 & female==1 & ret_year==1997
    replace ret_sr=63.5 if country==16 & female==0 & ret_year==1998
    replace ret_sr=58.5 if country==16 & female==1 & ret_year==1998
    replace ret_sr=64 if country==16 & female==0 & ret_year==1999
    replace ret_sr=59 if country==16 & female==1 & ret_year==1999
    replace ret_sr=65 if country==16 & female==0 & ret_year>=2000
    replace ret_sr=60 if country==16 & female==1 & ret_year>=2000
    replace ret_sr=65.08 if country==16 & female==0 & ret_year>=2013
    replace ret_sr=60.08 if country==16 & female==1 & ret_year>=2013
    replace ret_sr=65.17 if country==16 & female==0 & ret_year>=2014
    replace ret_sr=60.17 if country==16 & female==1 & ret_year>=2014
    replace ret_sr=65.25 if country==16 & female==0 & ret_year>=2015
    replace ret_sr=60.25 if country==16 & female==1 & ret_year>=2015

    replace ret_er=57 if country==16 & ret_year>=1996

*>> 34. Slovenia

    replace ret_sr=63 if country==34 & female==1
    replace ret_sr=65 if country==34 & female==0
    replace ret_sr=65 if country==34 & ret_year>=2013

    replace ret_er=58 if country==34 & female==1
    replace ret_er=58 if country==34 & female==0

*>> 15. Spain

    replace ret_sr=65 if country==15

    replace ret_er=64 if country==15
    replace ret_er=60 if country==15 & ret_year>=1983
    replace ret_er=61 if country==15 & ret_year>=1994

*>> 13. Sweden

    replace ret_sr=67 if country==13 & ret_year<=1994
    replace ret_sr=65 if country==13 & ret_year>=1995


    replace ret_er=60 if country==13
    replace ret_er=63 if country==13 & ret_year>=1998

*>> 20. Switzerland
```

```
    replace ret_sr=62 if country==20 & ret_year<=2000
    replace ret_sr=63 if country==20 & ret_year>=2001
    replace ret_sr=64 if country==20 & ret_year>=2005

    replace ret_er=64 if country==20 & ret_year>=1997 & female==0
    replace ret_er=63 if country==20 & ret_year>=2001 & female==0

    replace ret_er=62 if country==20 & ret_year>=2001 & female==1



*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[18. Generate the instrumental variables]---------------------------------

*>> IV - Retired
*   dummy variables taking value zero if the individual's age is less than the
statutory age for either early or regular retirement.

    gen iv_retired_er = 0
    replace iv_retired_er = 1 if age>ret_er
    la var iv_retired_er "Above early retirement age"

    gen iv_retired_sr = 0
    replace iv_retired_sr = 1 if age>ret_sr
    la var iv_retired_sr "Above statutory retirement age"



*>> IV - Retired at least one year
*   dummy variables taking value zero if the individual's age is less than the
statutory age for either early or regular retirement.

    gen iv_atleast1_er = 0
    replace iv_atleast1_er = 1 if age > ret_er +1
    la var iv_atleast1_er "Above early retirement age +1"

    gen iv_atleast1_sr = 0
    replace iv_atleast1_sr = 1 if age > ret_sr +1
    la var iv_atleast1_sr "Above statutory retirement age +1"



*>> IV - Retired less than one year
*   dummy variables taking value one if the individual's age is exaclty the
statutory age for either early or regular retirement.

// first we generate the age, not rounded to the nearest integer
// but floored to the integer n such that n < x < n + 1
    gen agefloor=floor(age)

    gen iv_lessthan1_er = 0
    replace iv_lessthan1_er = 1 if agefloor==ret_er
    la var iv_lessthan1_er "Exactly on early retirement age"

    gen iv_lessthan1_sr = 0
    replace iv_lessthan1_sr = 1 if agefloor==ret_sr
    la var iv_lessthan1_sr "Exactly on statutory retirement age"



*>> IV - Retired at least/less than ... months


forval m = 0/24 {

```

```
    gen iv_al`m'm_er = 0
    replace iv_al`m'm_er = 1 if age>ret_er + `m'/12
    la var iv_al`m'm_er "Above early retirement age +`m' months"

    gen iv_al`m'm_sr = 0
    replace iv_al`m'm_sr = 1 if age>ret_sr + `m'/12
    la var iv_al`m'm_sr "Above statutory retirement age +`m' months"


    gen agefloor`m'months=floor(age + max( 0 , (`m'/12)-1 ) )

    gen iv_lt`m'm_er = 0
    replace iv_lt`m'm_er = 1 if agefloor`m'months==ret_er
    la var iv_lt`m'm_er "Exactly on early retirement age"

    gen iv_lt`m'm_sr = 0
    replace iv_lt`m'm_sr = 1 if agefloor`m'months==ret_sr
    la var iv_lt`m'm_sr "Exactly on statutory retirement age"

    }


*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[19. Treat missing variables]---------------------------------------------


*>> Missing values treatment
/*
SHARE : general missing codes
-1: “Don’t know”
-2: “Refusal”
-3: “Implausible value/suspected wrong”
-4: “Not codable”
-5: “Not answered”
-7: “Not yet coded”
-9: “Not applicable”

EASYSHARE : Compared to the SHARE main release we recoded easySHARE to one of
the following (partially new) codes:
-3: “implausible value/suspected wrong”
-7: “not yet coded”
-9: “not applicable filtered”
-12: “don’t know / refusal”
-13: “not asked in this wave”
-14: “not asked in this country”
-15: “no information”
-16: “no drop-off (information in drop-off in this wave)”
*/

    summ
    mvdecode _all , mv(-1=. \ -2=. \ -12=. \ -15=.)

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[20. Save final database]-------------------------------------------------

save final_long, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[21. Descriptive tables]--------------------------------------------------

```

```
*>> Gender distribution of selected individuals by country

    use final_long, clear

    keep mergeid country female birth_year

    bysort mergeid: gen dup = cond(_N==1,0,_n)
    keep if dup==1

    tab country female

    collapse (count) totalw=dup, by (country female)

    reshape wide totalw, i(country ) j(female)
    rename totalw0 male
    rename totalw1 female

    gen total=male+female
    gen pcmale=round((male/total)*100,.1)
    gen pcfemale=round((female/total)*100,.1)

    decode country, gen(country_name)
    drop country
    gen newvarname = substr(country_name, 4, .)
    rename newvarname country
    drop country_name

    save ./temp/temp.dta, replace
    collapse (sum) male female total
    gen country="All countries"
    app using ./temp/temp
    sort country
    order country

    order country male pcmale female pcfemale total
    texsave * using "./tables/temp/sample.tex" , replace hlines(-1)


*>> Selection process: Percentages of selected men and women from the original sample, by country

*** selected individuals database

    use final_long, clear

    bysort mergeid: gen dup = cond(_N==1,0,_n)
    keep if dup==1

    keep mergeid wave firstwave country female

    save ./temp/baseline_select.dta, replace

    //25,031 obs

*** complete database (waves 4,5 and 6)

    use temp/selected_long, clear

    bysort mergeid: gen dup = cond(_N==1,0,_n)
    keep if dup==1

    keep mergeid wave firstwave country female

        drop if    country == "14. Netherlands":country      ///
             | country == "19. Greece":country          ///
             | country == "25. Israel":country          ///
             | country == "29. Poland":country          ///
             | country == "31. Luxembourg":country      ///
             | country == "32. Hungary":country         ///
```

```
            | country == "33. Portugal":country         ///
            | country == "47. Croatia":country

    save ./temp/baseline_complete.dta, replace

    //54.389 obs total

*** Merge the to two databases to be able to compare

    use ./temp/baseline_select.dta, clear

    merge 1:1 mergeid using ./temp/baseline_complete

    mark selected if _merge==3

*** Create the table that illustrates the selection process

    collapse (count) totalcount=wave, by (country female selected )
    reshape wide totalcount , i(country female) j(selected)

    rename totalcount0 unselected
    rename totalcount1 selected

    gen total= unselected + selected

    replace unselected=round((unselected/total)*100)
    replace selected=round((selected/total)*100)
    drop total

    reshape wide unselected selected, i(country) j(female)

    decode country, gen(country_name)
    drop country
    gen newvarname = substr(country_name, 4, .)
    rename newvarname country
    drop country_name
    order country

    la var unselected0 "Men unselected"
    la var selected0 "Men selected"
    la var unselected1 "Women unselected"
    la var selected1 "Women unselected"

    gen diff=round(unselected1-unselected0,.1)
    la var diff "Gender difference in selection"


    save ./temp/temp.dta, replace
    collapse (mean) un* sel* diff
    gen country="All countries"
    app using ./temp/temp
    sort country
    order country unselected0 selected0 unselected1 selected1

    texsave * using "./tables/temp/labor_force.tex" , replace hlines(-1)


*>> Gender distribution by country in original sample
    use ./temp/baseline_complete.dta, clear

    tab country female
    collapse (count) totalcount=wave, by (country female )
    reshape wide totalcount, i(country) j(female)

    la var totalcount0 "Men"
    la var totalcount1 "Women"

    decode country, gen(country_name)
    drop country
```

```
    gen newvarname = substr(country_name, 4, .)
    rename newvarname country
    drop country_name
    order country

    gen total=totalcount0+totalcount1
    replace totalcount0=round((totalcount0/total)*100,.1)
    replace totalcount1=round((totalcount1/total)*100,.1)
    drop total
    gen diff=totalcount1-totalcount0
    summ diff
    label var diff "Diff."

    format totalcount0 %9.1f
    format totalcount1 %9.1f

    texsave * using "./tables/genderdis.tex" , size(small) ///
    width(1\textwidth) title("Gender distribution by country in original sample (in\%)") ///
    replace varlabels frag location(ht) nofix marker(genderdis) align(@{\extracolsep{\fill} } l r r r)


*>> Retirement related summary statistics by country

use final_wide, clear

    collapse (mean) ret_age es_evol_working es_evol_fromwtor es_evol_retired, by(country)

    replace ret_age=round(ret_age,0.1)
    replace es_evol_retired=round(100*es_evol_retired)
    replace es_evol_fromwtor=round(100*es_evol_fromwtor)
    replace es_evol_working=round(100*es_evol_working)

    decode country, gen(country_name)
    drop country
    gen newvarname = substr(country_name, 4, .)
    rename newvarname country
    drop country_name
    order country

    la var ret_age "Retirement age"
    la var es_evol_retired "Retired during the entire sample period"
    la var es_evol_fromwtor "Transitions into retirement"
    la var es_evol_working "Working during the entire sample period"

    save ./temp/temp.dta, replace
    collapse (mean) ret_age es*
    gen country="All countries"
    app using ./temp/temp
    sort country
    order country

    texsave * using "./tables/temp/ret_stat.tex" , replace hlines(-1)


*>> Descriptive statistics

use final_long, clear

*cognitive
    summ numeracy_2 orienti cf010_ mem_sc_ recall_1 recall_2

*retirement
    summ es_retired_ es_evol_* ret_age ret_length

*control and descriptive
    summ female age partnerinhh retest
```

```
    tab degree

*country
    tab country
    tab region
    tab country region


*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------

*----[22. Graphs]---------------------------------------------------------------

*>> Preliminary settings

query graphics

graph query, schemes

*ssc install blindschemes
*set scheme plotplainblind, perm

*graph set window fontface default //change font
graph set window fontface "Garamond Premier Pro Caption" //change font

*** Colors
//  Define global variables with the color coding in RGB

graph query colorstyle

*blue
global fill1     ""203 213 232""
global line1     ""141 160 203""

*red
global fill2     ""253 205 172""
global line2     ""252 141 98""

*green
global fill3     ""179 226 205""
global line3     ""102 194 165""

*grey
global fill4     "gs10*0.5"
global line4     "gs10"


*>> Age profiles - gender
foreach test in Immediate_recall Delayed_recall Fluency {

    use final_long, clear
    keep if ageround>=60 & ageround<=75

    rename (orienti recall_1 recall_2 cf010_ numeracy_2) ///
           (Orientation Immediate_recall Delayed_recall Fluency Numeracy)

    statsby score=r(mean) upper=r(ub) lower=r(lb)   ,               ///
        by(ageround female) clear                                   ///
        : ci mean (`test')

#delimit ;
    twoway
        (rarea upper lower ageround if female==0, color($fill1))
        (rarea upper lower ageround if female==1, color($fill2))
        (line score ageround if female==0, color($line1) lp(solid) lw(medium))
        (line score ageround if female==1, color($line2) lp(solid) lw(medium))
        , leg( order(2 "Women" 1 "Men") r(1) colgap(*2) region(lstyle(refline ))
si(medium) )   scale(1.5)
```

```
          title("`test'") name(`test', replace) xtitle("");
#delimit cr
    }

grc1leg2   Immediate_recall Delayed_recall Fluency ,rows(1) ring(1) title("") xtitle("")
graph di , xsize(9) ysize(5) margin(zero)
graph export graphs/tests/gender.pdf, replace


*>> Age profiles - education level
foreach test in Immediate_recall Delayed_recall Fluency  {

    use final_long, clear
    rename (orienti recall_1 recall_2 cf010_ numeracy_2) ///
    (Orientation Immediate_recall Delayed_recall Fluency Numeracy)
    keep if ageround>=60 & ageround<=75

    statsby score=r(mean) upper=r(ub) lower=r(lb)   ,               ///
        by(ageround degree) clear                                   ///
        : ci mean (`test')

#delimit ;
    twoway
        (rarea upper lower ageround if degree==0, color($fill1)  )
        (rarea upper lower ageround if degree==1, color($fill2) )
        (rarea upper lower ageround if degree==2, color($fill3) )
        (line score ageround if degree==0, color($line1) lp(solid) lw(medium))
        (line score ageround if degree==1, color($line2) lp(solid) lw(medium))
        (line score ageround if degree==2, color($line3) lp(solid) lw(medium))
        , leg( order( 1 "Low" 2 "Medium" 3 "High" ) r(1) colgap(*2) region(lstyle(refline )) si(medium))   scale(1.5)
        title("`test'") name(`test', replace) xtitle("") ;
#delimit cr
    }

grc1leg2  Immediate_recall Delayed_recall Fluency, rows(1)  ring(1) title("")
graph di , xsize(9) ysize(5) margin(zero)
graph export graphs/tests/hsdegree.pdf, replace

*>> Age profiles - retirement status
foreach test in  Immediate_recall Delayed_recall Fluency  {

    use final_long, clear
    rename (orienti recall_1 recall_2 cf010_ numeracy_2) ///
    (Orientation Immediate_recall Delayed_recall Fluency Numeracy)

    recode ep005_ (-15 -12 = .) (1 = 0) (2 = 1) (3 4 5 97 = .) // employment status

    statsby score=r(mean) upper=r(ub) lower=r(lb), by(ageround ep005_) clear ///
        : ci mean (`test')

#delimit ;
    twoway
        (rarea upper lower ageround if ep005_==0 & ageround>=55, color($fill1))
        (rarea upper lower ageround if ep005_==1 & ageround<=65, color($fill2) )
        (line score ageround if ep005_==0 & ageround>=55, color($line1 ) lp(solid) lw(medium))
        (line score ageround if ep005_==1 & ageround<=65, color($line2) lp(solid) lw(medium))
        , leg( order(2 "Employed" 1 "Retired" ) r(1) colgap(*2) region(lstyle(refline )) si(medium))   scale(1.5)
          name(`test', replace) xtitle("") title("`test'") ;
#delimit cr
    }

grc1leg2 Immediate_recall Delayed_recall Fluency, rows(1) ring(1) title("")
graph di , xsize(9) ysize(5) margin(zero)
```

```
graph export graphs/tests/retirement.pdf, replace

*>> Age profiles - macro region
foreach test in  Immediate_recall Delayed_recall Fluency  {

    use final_long, clear
    rename (orienti recall_1 recall_2 cf010_ numeracy_2) ///
           (Orientation Immediate_recall Delayed_recall Fluency Numeracy)
    keep if ageround>=60 & ageround<=75

    statsby score=r(mean) upper=r(ub) lower=r(lb)   ,               ///
        by(ageround region) clear                                   ///
        : ci mean (`test')

#delimit ;
    twoway
        (rarea upper lower ageround if region==1, color($fill1) )
        (line score ageround if region==1, color($line1) lp(solid) lw(medium))
        (rarea upper lower ageround if region==2, color($fill2) )
        (line score ageround if region==2, color($line2) lp(solid) lw(medium))
        (rarea upper lower ageround if region==4, color($fill4) )
        (line score ageround if region==4, color($line4) lp(solid) lw(medium))
        (rarea upper lower ageround if region==3, color($fill3) )
        (line score ageround if region==3, color($line3) lp(solid) lw(medium))

        , title("`test'")
          leg(order(1 "Nordic countries" 3 "Western Europe" 7 "Southern Europe" 5 "Former Eastern bloc" ) r(1) colgap(*1) region(lstyle(refline )) si(medium) ) scale(1.5)
          name(`test', replace)
          xtitle("")
          ;
#delimit cr
    }

grc1leg2   Immediate_recall Delayed_recall Fluency, rows(1) ring(1) title("")
graph di , xsize(9) ysize(5) margin(zero)
graph export graphs/tests/region.pdf, replace

*>> Memory score averages by gender and country
    use final_long, clear

    statsby score=r(mean), by (country female) clear : summ mem_sc

    la def country_abbr 11 "AT" 12 "DE" 13 "SE" 15 "ES" 16 "IT" 17 "FR" ///
                        18 "DK" 20 "CH" 23 "BE" 28 "CZ" 34 "SI" 35 "EE"
    la val country country_abbr

    graph bar score,    over(female, descending)                        ///
                        over(country, gap(*3) sort(2) descending )      ///
                        asyvars ytitle("") ylabel(0(2)13) bargap(25)    ///
                        leg(label(2 Female) label(1 Male) order(2 1)ring(0) position (1))   ///
                        bar(1, bcolor(gs12)) bar(2, bcolor(gs4)) ///
                        yscale(lcolor(gs10))

    graph export graphs/countries_recall.pdf, replace

*>> Average memory score by age

    use final_long, clear

    egen ageroundbis = cut(age), at(55(.5)75.5)

    statsby mean=r(mean) upper=r(ub) lower=r(lb),       ///
        by(ageroundbis) clear                               ///
        : ci mean (mem_sc_)

    #delimit ;
```

```
    twoway
        (rarea upper lower ageroundbis, color($fill2))
        (line mean ageroundbis, color($line2) lp(solid) lw(medium))
        , title() leg(off)
        xlabel(55(5)75) xtitle("") scale(2)
        ;
    #delimit cr

    graph di , xsize(9) ysize(4) margin(zero)
    graph export graphs/avg_score.pdf, replace

*>> Retirement age distribution

*all countries
    use final_long, clear
    keep ret_age_round

    drop if ret_age_round<50 | ret_age_round>70.00
    tab ret_age_round

    hist ret_age_round, percent d norm gap(15) bfcolor($fill2) blcolor($line2) ///
    leg(off) xtitle("") ytitle("") mlabsize(huge) scale(2.5)

    graph di , xsize(3) ysize(1)
    graph export graphs/ret_age_all.pdf, replace

*country by country
    use final_long, clear
    decode country, gen(country_name)
    drop country
    gen newvarname = substr(country_name, 4, .)
    rename newvarname country

    drop if ret_age_round<54 | ret_age_round>70.00

    hist ret_age_round , percent d norm gap(15) bfcolor($fill2) blcolor($line2) ///
    by( country, col(3) iscale(*1.15) leg(off) note("") im(0 0 0 7 0) ) ///
        subtitle(, ring(0) pos(12) nobexpand) xtitle("") ytitle("") mlabsize(huge) leg(off) ///
        xlabel(55(5)70) xsc(r(54 70))

    graph di , ysize(12) xsize(9) margin(0 0 0 -6)
    graph export graphs/ret_age_country.pdf, replace

*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[23. Regressions - Setting up global variables]---------------------------

*>> Global variables used in the following regressions

use final_long, clear

    global iv_retired          "iv_retired_er iv_retired_sr"
    global iv_atleastoneyear   "iv_atleast1_er iv_atleast1_sr"
    global iv_lessthanoneyear  "iv_lessthan1_er iv_lessthan1_sr"

    global agef                "age agesq"
    global controls            "retest partnerinhh"



*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[24. Regressions - Preliminary model]-------------------------------------
```

```
*>> Fixed effects

eststo retired1 : ///
xtreg mem_sc_ es_retired  $agef  $controls, fe cluster(id) robust


*>> 2SLS FE

eststo retired2 : ///
xtivreg2 mem_sc_ (es_retired = $iv_retired) $agef  $controls , ///
    fe cluster(id wave) first savefprefix(fretired2) robust


*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[25. Regressions - Alternative model - one regressor]---------------------

*>> Fixed effects

eststo atleastoneyear1 : ///
 xtreg mem_sc_ es_ret_1year  $agef $controls, fe cluster(id) robust


*>> 2SLS FE

eststo atleastoneyear2 : ///
xtivreg2 mem_sc_ (es_ret_1year = $iv_atleastoneyear) $agef  $controls , ///
    fe cluster(id wave) first savefprefix(fatleast) robust endog(es_ret_1year)


*------------------------------------------------------------------------------
*------------------------------------------------------------------------------
*------------------------------------------------------------------------------

*----[26. Regressions - Retired at least 1--24 months]------------------------

*>> 2SLS FE - one endogenous regressor

forval m = 0/24 {
    preserve

    eststo atleast`m' : ///
    xtivreg2 mem_sc_ (es_ret_al`m'm = iv_al`m'm_er iv_al`m'm_sr) ///
        $agef  $controls , ///
        fe cluster(id wave) robust

    regsave es_ret_al`m'm using temp/months/month`m', ci replace

    restore
    }


*>> dataset and graph
preserve

use temp/months/month0, clear

forvalues i=1/24 {
    appen using temp/months/month`i'
    }
save temp/months/allmonths.dta, replace

use temp/months/allmonths.dta, clear

    gen month = _n-1
    keep month coef ci*
```

```
    mark honeymoon if month>12
    egen mean = mean(coef), by(honeymoon)

    twoway                                                  ///
        (lfit mean month if honeymoon==0, lc($line2) lp(solid)) ///
        (lfit mean month if honeymoon==1, lc($line2) lp(solid)) ///
        (rcap ci_upper ci_lower month, lcolor(gs12) ) ///
        (scatter coef month, m(square) mc(gs5)) ///
        , leg(off) xlabel(0(1)24) ylabel(-.2(.2)1) ///
          xtitle("Delay after retirement, in months") ///
          ytitle("Coefficient estimate")

    graph di , xsize(6) ysize(4) margin(zero)
    graph export graphs/months_one.pdf, replace

restore


*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------

*----[27. Regressions - Alternative model - two regressors]--------------------

*>> 2SLS FE

eststo lessthanoneyear2 : ///
xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  = $iv_atleastoneyear
$iv_lessthanoneyear) $agef  $controls , ///
    fe cluster(id wave) first savefprefix(flessthan) robust endog(es_ret_1year)

*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------

*----[28. Regressions - Sensitivity to age polynomials]------------------------

*>> Log age, and polynomials up to degree 4

eststo contagelog : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) agelog  $controls , ///
                 fe cluster(id wave) robust

eststo contage1 : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) age  $controls , ///
                 fe cluster(id wave) robust

eststo contage2 : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) age agesq  $controls , ///
                 fe cluster(id wave) robust

eststo contage3 : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) age agesq agethree  $controls , ///
                 fe cluster(id wave) robust

eststo contage4 : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) age agesq agethree agefour  $controls ,
///
                 fe cluster(id wave) robust


*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------
*-----------------------------------------------------------------------------

*----[29. Regressions - Heterogeneity across subsamples]-----------------------

*>> Education

```

```
tab degree educ, m
mark educalt if degree==2
replace educalt=. if degree==.

bysort educalt : xtivreg2 mem_sc_ (es_ret_1year = $iv_atleastoneyear) $age
$controls , ///
        fe cluster(id wave) robust

bysort educalt : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  =
$iv_atleastoneyear $iv_lessthanoneyear) $agef  $controls , ///
    fe cluster(id wave) robust


*>> Gender

tab female, m

bysort female : xtivreg2 mem_sc_ (es_ret_1year = $iv_atleastoneyear) $age
$controls , ///
        fe cluster(id wave) robust

bysort female : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  = $iv_atleastoneyear
$iv_lessthanoneyear) $agef  $controls , ///
    fe cluster(id wave) robust


*>> Macro region

preserve
tab region country
replace region=5 if country==23 // to estimate BE separately

bysort region : xtivreg2 mem_sc_ (es_ret_1year = $iv_atleastoneyear) $age
$controls , ///
        fe cluster(id wave) robust

bysort region : xtivreg2 mem_sc_ (es_ret_less1y es_ret_1year  = $iv_atleastoneyear
$iv_lessthanoneyear) $agef  $controls , ///
    fe cluster(id wave) robust
restore

*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------
*-------------------------------------------------------------------------------

*----[30. Exit Stata & Close Log File]------------------------------------------

capture log close
exit

*-------------------------------------------------------------------------------
```