# Effects of working part-time and full-time on physical and mental health in old age in Europe

Tunga Kantarcı and Ingo Kolodziej

Tilburg University and Netspar
RWI - Leibniz Institute for Economic Research and Ruhr-University Bochum

4 October 2016

**Abstract**

We distinguish between part-time and full-time work activity and analyse their effects on the physical and mental health conditions of older workers in Europe. We use statutory eligibility ages for receiving retirement benefits as instruments for part-time and full-time work decisions to avoid the potential bias that deteriorating health conditions can cause employees to work fewer hours or not at all. We also control for unobserved heterogeneity across individuals. We find that working full-time deteriorates general health, reduces body weight, increases depression symptoms, has a negative effect on word recall score and a positive effect on numeracy score. Working part-time has the opposite effects that are larger in magnitude. A comparison of the results obtained in Europe with those obtained in the United States in an earlier study shows that health responses to working part-time and full-time in old age differ across Europe and the United States.

## 1 Introduction

Older workers today spend more years in the labor market, mainly as a result of the rising retirement ages, in Europe or in the United States. Moreover, the fraction of the older workers in the labor market is increasing due to increasing life expectancy. This means that working is becoming more common than ever before among older people. Since a main determinant of the general well-being in old age is health, it is important to understand how working or working conditions affect health status in old age. This explains why there is a growing body of empirical studies analyzing the causal effect of retirement on physical and mental health (Bonsang et al., 2012; Charles, 2004; Coe and Zamarro, 2011; Mazzonna and Peracchi, 2012; Neuman, 2008; Rohwedder and Willis, 2010). These studies have analyzed the effect of retirement against work activity, but did not distinguish between part-time and full-time work activity. Kantarcı (2016) has analyzed how the amount of work hours affects the physical and mental health conditions of US residents between 50 and 75 years of age in the Health and Retirement Study (HRS). He finds that self-perceived general health status, self-perceived memory skills, and body weight respond to working part-time much more than they respond to working full-time, suggesting that the effect of the number of hours worked on health outcomes is not linear. In this study we closely follow the empirical approach and adopt similar health indicators used by Kantarcı, but conduct our analysis among 12 European countries. These countries widely differ from each other in terms of the pension reforms that have changed the long-standing public pension eligibility ages in the last ten years, allowing us to use the rich source of variation in the retirement eligibility ages as determinants of working part-time and full-time, to identify the effects of working part-time and full-time on health in an instrumental variables framework. We also account for time-invariant individual specific unobserved heterogeneity among survey respondents.

Our results show that working full-time deteriorates general health but working part-time tends to improve it. On the other hand, working full-time reduces body weight while working part-time increases it. Working full-time increases depression symptoms while working part-time has a large and positive effect. With respect to cognition, we find that working part-time has a large and positive effect on word recall score, while working full-time has a relatively smaller and negative effect. A comparison of the results obtained in Europe with those obtained in the United States in an earlier study shows that health responses to working part-time and full-time in old age differ across Europe and the United States.

This paper proceeds as follows. Section 2 discusses the empirical model. Section 3 describes the data and the health and work effort indicators, and Section 4 explores the data using graphical analysis. Section 5 presents the results and robustness checks. Section Section 6 concludes.

## 2 Empirical approach

### 2.1 Controlling for heterogeneity

Our aim is to determine the effects of working part-time and full-time on health. The first attempt could be to estimate the parameter of interest by ordinary least squares in the following equation:

$$Y_{it} = \alpha + f(S_{it}) + D_{it}\boldsymbol{\beta} + u_{it}. \tag{2.1}$$

$Y_{it}$ is a measure of health, for example the self-perceived health or body mass index. $S_{it}$ is the age of the individual. $f(S_{it})$ is a continuous function of age that controls for changes in the health outcome with age. $D_{it}$ is a vector of two dummy variables indicating part-time and full-time work. The parameter of interest is the vector $\boldsymbol{\beta}$, which measures the responses of the health outcome to working part-time and full-time.

OLS on Equation (2.1) leads to a consistent estimator for $\boldsymbol{\beta}$ only if $D_{it}$ is not correlated with the error term $u_{it}$. One reason why this assumption may not be satisfied is that individuals might differ from each other because of time-invariant idiosyncratic characteristics that are correlated with the health outcome as well as the retirement behavior. We follow a fixed effects approach to allow for this, augmenting Equation (2.1) as follows:

$$Y_{it} = \alpha + f(S_{it}) + D_{it}\boldsymbol{\beta} + \mu_i + \nu_{it}. \tag{2.2}$$

$\mu_i$ is a time-invariant individual specific unobserved variable and it is potentially correlated with $D_{it}$ (and with $S_{it}$). The remaining error term $\nu_{it}$ is assumed to be uncorrelated with the control variables. The main parameters of interest, the effects of working part-time or full-time on the health measure considered, are contained in the vector $\boldsymbol{\beta}$. Note that we assume throughout that these 'treatment effects' are assumed to be homogeneous across the population. We will relax this assumption somewhat by estimating the model for specific demographic groups. Moreover, Murtazashvilia and Wooldridge (2008) have shown that under some additional assumptions the fixed effects instrumental variables estimator that we use remains consistent for the average treatment effect in the model with heterogeneous treatment effects. Following the main studies on this topic referred to above, however, we will not consider models with heterogeneous treatment effects.

Exploiting the panel structure of the data, $\mu_i$ is eliminated through the within group transformation:

$$\widetilde{Y}_{it} = \widetilde{f}(S_{it}) + \widetilde{D}_{it}\boldsymbol{\beta} + \widetilde{\nu}_{it}, \tag{2.3}$$

where $\widetilde{Y}_{it}$ represents $Y_{it} - \overline{Y}_i$, etc. The assumption that $\nu_{it}$ is uncorrelated with the control variables (strict exogeneity) implies that OLS on Equation (2.2) (the standard within group estimator for static linear panel data models with fixed effects) gives consistent estimates of $\boldsymbol{\beta}$.

### 2.2 Controlling for endogeneity

A potential problem in Equation (2.3) is that $\widetilde{D}_{it}$ may be correlated with the unobserved $\widetilde{\nu}_{it}$, making the fixed effects estimator for $\boldsymbol{\beta}$ inconsistent. This might happen because, for example, employees with a work-limiting health problem may opt for part-time work or full-time retirement (reverse causation). For example, examining the causal effect of health on labor market behavior, Gannon and Roberts (2011) find that, in the UK, people aged 50 and over with health problems are more likely to work part-time or to retire completely than to work full-time. Bound et al. (1999) show that, in the US, poor health is often followed by labor force exit. Mols et al. (2012) show that most of the patients who are diagnosed with cancer switched to part-time work or stopped working entirely in the Netherlands.

We follow an instrumental variables approach to solve the problem of potential endogeneity of hours worked, exploiting discontinuities in the probabilities to work part-time and full-time as a function of age at the eligibility ages, similar to Coe and Zamarro (2011) and Rohwedder and Willis (2010). The instrumental variables estimation consists of two stages. In the first stage, we estimate two equations explaining the dummies $D^j_{it}, j = p, f$ for part-time and full-time work:

$$D^j_{it} = f^j(S_{it}) + I(S_{it} \geq \bar{S})\boldsymbol{\gamma^j} + \eta^j_i + \epsilon^j_{it}. \tag{2.4}$$

$f^j(S_{it})$ are continuous functions of age. $\bar{S}$ is the vector of early and normal retirement eligibility ages for social security benefits, and the vector $I(S_{it} \geq \bar{S})$ indicates whether the individual is at least as old as each of these eligibility ages. $\boldsymbol{\gamma^j}$ measures the discontinuities in the probabilities of working part-time or full-time at the eligibility ages $\bar{S}$. Hence, this is essentially a regression discontinuity approach (Lee and Lemieux, 2010) in a fixed effects panel data model.[1] Since the elements of $D^j_{it}$ are binary indicators, Equation (2.4) is a linear probability model. The fixed effects $\eta^j_i$ are time-invariant, individual-specific unobserved variables, and they are potentially correlated with age. Exploiting the panel structure of the data, $\eta^j_i$ are eliminated through the within group transformation:

$$\widetilde{D}^j_{it} = \widetilde{f}^j(S_{it}) + \widetilde{I}(S_{it} \geq \bar{S})\boldsymbol{\gamma^j} + \widetilde{\epsilon}^j_{it}. \tag{2.5}$$

The predicted values from the first stage are used to estimate the main Equation (2.3) in the second stage:

$$\widetilde{Y}_{it} = \widetilde{f}(S_{it}) + \widehat{\widetilde{D}}_{it}\boldsymbol{\beta} + \widetilde{\upsilon}_{it}. \tag{2.6}$$

$\widehat{\widetilde{D}}_{it}$ represents the within group transformed part-time and full-time work probabilities predicted from Equation (2.5). To be valid instruments, retirement eligibility ages are required to be relevant predictors of the full-time and part-time work decisions and exogenous to the respondent's health status. It is well documented that the retirement ages are strong predictors of the retirement decision, and we will also check below that this is the case in our sample. Moreover, it

[1] In our baseline model, however, we supplement the retirement eligibility ages of the respondent with those of the partner while we do not allow for a continuous age polynomial for the partner. Hence, our baseline model does not follow a regression discontinuity design in the age of the partner.

is plausible to assume that health status does not change discontinuously at the institutionally determined eligibility ages. If the selected instruments are indeed valid, the causal effect of working part-time or full-time on health status, measured by $\boldsymbol{\beta}$, is consistently estimated using least squares on equation (2.6). The complete two-stage estimation procedure corresponds to the two-stage least squares estimation. The 2SLS estimates can then be interpreted as capturing the covariate-adjusted causal effects of each employment status.

# 3 Data

The data are taken from the Survey of Health, Ageing, and Retirement in Europe (SHARE). We use waves 1, 2, 4 and 5 of the survey which together cover the time period from 2004 until 2014.[2] SHARE is a nationally representative panel study of approximately 110,000 individuals aged 50 or older. The data includes extensive information on health and socio-economic status and makes cross-national comparison possible making it very well suited for our analysis.

The following sample restrictions are imposed. First, we dropped respondents who reported they never worked, or who said they worked, but with a tenure of less than five years on all jobs. Second, we dropped respondents who reported their last job ended before the age of 50 in all survey years, or who reported this in given survey years and this information was missing in other survey years, or if this information was missing in all survey years. Third, we dropped respondents who reported to be working, unemployed, permanently sick or disabled, or other (rentier, living off own property, student, doing voluntary work), after reporting retirement in a previous survey year, so that retirement is an absorbing state. Fourth, we dropped the respondents if they are unemployed, permanently sick or disabled, homemaker, or other in a given survey year. We dropped respondents who are unemployed because, like retired, they work 0 hours but they are probably more active since they would be searching for a part-time or full-time job. We dropped respondents who are disabled, homemaker or other because in these cases respondents are not working, do not report being retired, or they are not searching for a job. Fifth, we dropped the observations of respondents if they were younger than 50 years old or older than 75 years old in a given survey year. These sample restrictions lead to an unbalanced panel of 86,659 observations for 19,603 individuals (based on the information available on employment status). Finally, among the countries participating in the SHARE, we selected the countries where information is available in all survey waves: Austria, Belgium, Czech Republic, Denmark, France, Germany, Greece, Italy, Sweden, Netherlands, Spain, and Switzerland.

## 3.1 Measuring health

### Self-perceived health

We use self-perceived health as an assessment of one's own health status. In the survey respondents are asked to rate their health on a five-point scale: very good, good, fair, bad and very bad. Self-assessed health is a global index of health that captures physical and mental health in one simple survey measure. Analyzing self-reported health, however, may lead to biased conclusions about the effect of hours worked on health, since respondents may report an inferior health status to justify their labor market status (Bound, 1991). We therefore also consider several alternative indicators of physical and mental health, exploiting the rich health information in the SHARE.

[2] We do not use data from the third wave of SHARE (SHARELIFE) which is a special survey focusing on people's life histories and does not contain the same health information we use from other waves.

### Body mass index

We consider the body mass index (BMI) and also construct indexes of overweight and obesity based on the BMI. BMI is given by the weight (in kilograms) divided by the square of height of the respondent (in meters). Following the existing literature, overweight is defined as a BMI greater than 25 and less than or equal to 30; obesity is defined as a BMI greater than 30.

### Word recall score and numeracy

In SHARE, cognitive ability is measured through a number of cognitive function indices. We use the word recall score and numeracy score as objective measures of cognitive ability. Word recall test measures the memory performance of the respondents. In the test respondents are presented with a list of 10 words to memorize. They are then asked immediately to recall as many words as possible from the list in any order. After asking other survey questions for a considerable amount of time, they are asked for a second time to recall as many words as possible from the same list. Each immediate or delayed recall of a word is counted, giving a memory score ranging from 0 to 20.

Numeracy gives information on the mathematical performance of the respondents. It is based on four questions on percentage calculation summarized in a score that ranges from 1 (good) to 5 (bad). In waves 4 and 5, baseline respondents who already participated in one of panel waves are given a new test based on subtraction. Respondents who correctly answer the first question are asked a more difficult one, while those who make a mistake are asked an easier one.

### Depression score

We use the EURO-D symptom scale which measures the current depression and is constructed from several questions as a composite index of 12 items: depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, and tearfulness. The scale ranges from 0 'not depressed' to 12 'very depressed'.

### Health index

Following Coe and Zamarro (2011), we create an objective health index by predicting self-perceived health from objective physical and mental health measures. In particular, we estimate the following equation:

$$H_{it} = \alpha + L_{it}\boldsymbol{\beta} + \phi_i + \varepsilon_{it}. \tag{3.1}$$

$H_{it}$ is the self-perceived health status. $\phi_i$ is a time-invariant individual specific unobserved error that is potentially correlated with the control variables. $L_{it}$ is a vector of objective measures of health including the number of limitations in the activities of daily living (ADL), the number of limitations in the instrumental activities of daily living (IADL), the number of chronic diseases, a summary index of mobility, whether the respondent reports any overnight hospital stay within the last two years, overweight and obesity dummies, the scores of the word recall test discussed above, the score on a subtraction test for numerical skills, and the EURO-D score for depression.[3]

[3] ADL includes problems with bathing, dressing, eating, getting in/out of bed, and walking across a room. IADL includes problems with using the phone, managing money, taking medications, shopping for groceries, and preparing hot meals. Both variables take values from 0 (no problems) to 5 (many problems). The number of chronic diseases is a count of the diseases the respondent had according to a doctor. The diseases include high blood pressure, diabetes, cancer, lung disease, heart problems, stroke, psychiatric problems, and arthritis. The variable takes values from 0 (none of the conditions) to 8 (all conditions). The mobility index indicates problems with walking one block, walking several blocks, walking across a room, climbing one flight of stairs, and climbing

Equation (3.1) represents a fixed effects model. After the within group transformation, the predictions of the model, i.e., the estimates of $H_{it}$, creates a *health stock* variable that is less prone to reporting bias, as it aggregates objective measures of health, and at the same time reflects one's overall well-being, as measured by the self-assessed health status (Coe and Zamarro, 2011). The estimation results for this equation are presented in Table 1. A positive coefficient indicates that an increase in the particular health indicator leads to a self-report of worse health. Most of the coefficients are significant and their signs are plausible. Onsets of physical health problems are associated with reporting poorer health, and increasing depression symptoms (higher EURO-D score) also increase the odds of reporting poor health. A higher score on word recall is associated with reporting better health. On the other hand, the subtraction test result is not related to self-assessed health. Becoming obese leads to a significantly poorer self-assessment of health, while becoming overweight has a smaller and less significant effect, as we would expect.

several flights of stairs. The variable takes values from 0 to 5. Subtraction test asks the respondents to subtract 7 from 100 and continue subtracting 7 from each subsequent number for a total of five trials. Each correct subtraction is counted, yielding a score from 0 to 5.

Table 1: Results for FE model explaining self-perceived health

| | Self-perceived health | |
|---|---|---|
| | Coefficient | p-value |
| Number of ADL limitations | 0.057 | 0.007 |
| Number of IADL limitations | −0.024 | 0.384 |
| Number of mobility limitations | 0.122 | 0.000 |
| Number of difficulties in muscle use | 0.101 | 0.000 |
| Number of chronic diseases | 0.126 | 0.000 |
| Hospital stay | 0.217 | 0.000 |
| Body mass index | 0.012 | 0.001 |
| Word recall test | −0.002 | 0.404 |
| Numeracy | 0.016 | 0.007 |
| Depression | 0.060 | 0.000 |
| Fluency | −0.003 | 0.003 |
| Constant | 2.113 | 0.000 |
| F-test for overall significance | | 0.000 |
| N obs. | 56251 | |
| N ind. | 36668 | |

Notes: 1. Linear model with fixed effects. 2. Self-perceived health: 1 (Excellent), ..., 5 (poor). 3. Standard errors are robust to heteroskedasticity and clustering on panel groups.

### 3.2 Measuring work intensity

The aim of our analysis is to examine the effects of working part-time and full-time on health around retirement age. In the SHARE data, part-time or full-time work can be defined in a number of ways. Earnings, the number of hours worked per week, or the number of months worked per year are possible indicators of work effort. We define full-time work as working 35 or more hours per week for 8 months or more in a year. We define part-time work as working less than 35 hours a week for 8 months or more a year, or as working 35 or more hours a week but less than 8 months a year. We define retirement as working 0 hours a week. The hours and months from both the main and a possible second job are considered to determine whether the agent is working part-time or full-time.

### 3.3 Instruments

The instruments are based on the institutional variation in the retirement eligibility ages across selected European countries but also within these countries. We use two sets of instruments. Both sets aim to instrument for both working part-time and full-time. The first set includes two instruments indicating whether survey respondents are eligible for social security benefits. In particular, the indicators define whether the individual is between the early and normal retirement age, or above the normal retirement age. The age thresholds for early and full public retirement benefits are part of the public policies, and differ across the selected countries, but also often within each country by gender, birth cohort, and over time by up to 12 years. Figure 1 presents the early and normal retirement ages, and shows the rich source of variation in the eligibility ages we exploit to identify the part-time and full-time work probabilities.[4]

The literature on the effect of retirement on health shows that retirement ages are significant predictors of retirement behavior and are not likely to explain individual health status directly (Charles, 2004; Rohwedder and Willis, 2010; Coe and Zamarro, 2011; Bonsang et al., 2012; Mazzonna and Peracchi, 2012). Hence, as predictors of retirement behavior, dummies for reaching these institutional retirement ages present themselves as natural instruments. In our model, the instruments accommodate work preferences more explicitly as we distinguish between part-time and full-time work.

Following Neuman (2008), we also consider a second set of instruments which consists of the same two age indicators, but then for the married or unmarried partner. Whether the partner is eligible for social security benefits may explain the retirement behavior of an individual, whereas it has no direct effect on the health status of that individual. Indeed, several studies based on US data provide empirical evidence that couples coordinate their retirement timing (Blau, 1998; Gustman and Steinmeier, 2000, 2004). Note that the retirement eligibility ages vary between the gender groups in many of the sample countries in Figure 1, and therefore the retirement ages of the partner seem especially well suited to serve as valid instruments. We discuss the robustness of our results to restricting the instrument set to the eligibility ages of the individual in Section 5.3.

Table 2 presents the fraction of individuals in three employment states, based on reported hours of work, before the age at which they become eligible for social security, between the

[4] Eligibility ages are applicable to individuals who have reached the retirement eligibility ages at the calendar years presented in Figure 1. The eligibility ages chosen are those with the greatest incentives to retire and a minimum of group specific deviations apply. Earliest possible ages are presented, further variation apply in select countries. Italy: minimum ages to collect early retirement benefits differ by type of occupation; Denmark: normal retirement age is 65 and until 2008, normal retirement age is 67 for those born before 01.07.1939. France: as from 01.07.2011 normal retirement age increases by four months per birth year to reach 62 for individuals born in 1956 or later; as from 01.01.2012 normal retirement age increases by five months per birth year to reach 62 for persons born in 1955 or later.

early and normal retirement ages, and after the normal retirement age. The table also presents the fraction of the individuals in three employment states at the retirement eligibility ages of their partner. It appears that not only the fraction of those who work full-time, but also that of those who work part-time change substantially at the retirement eligibility ages. These figures suggest that retirement ages are relevant predictors of the number of hours worked in old age.

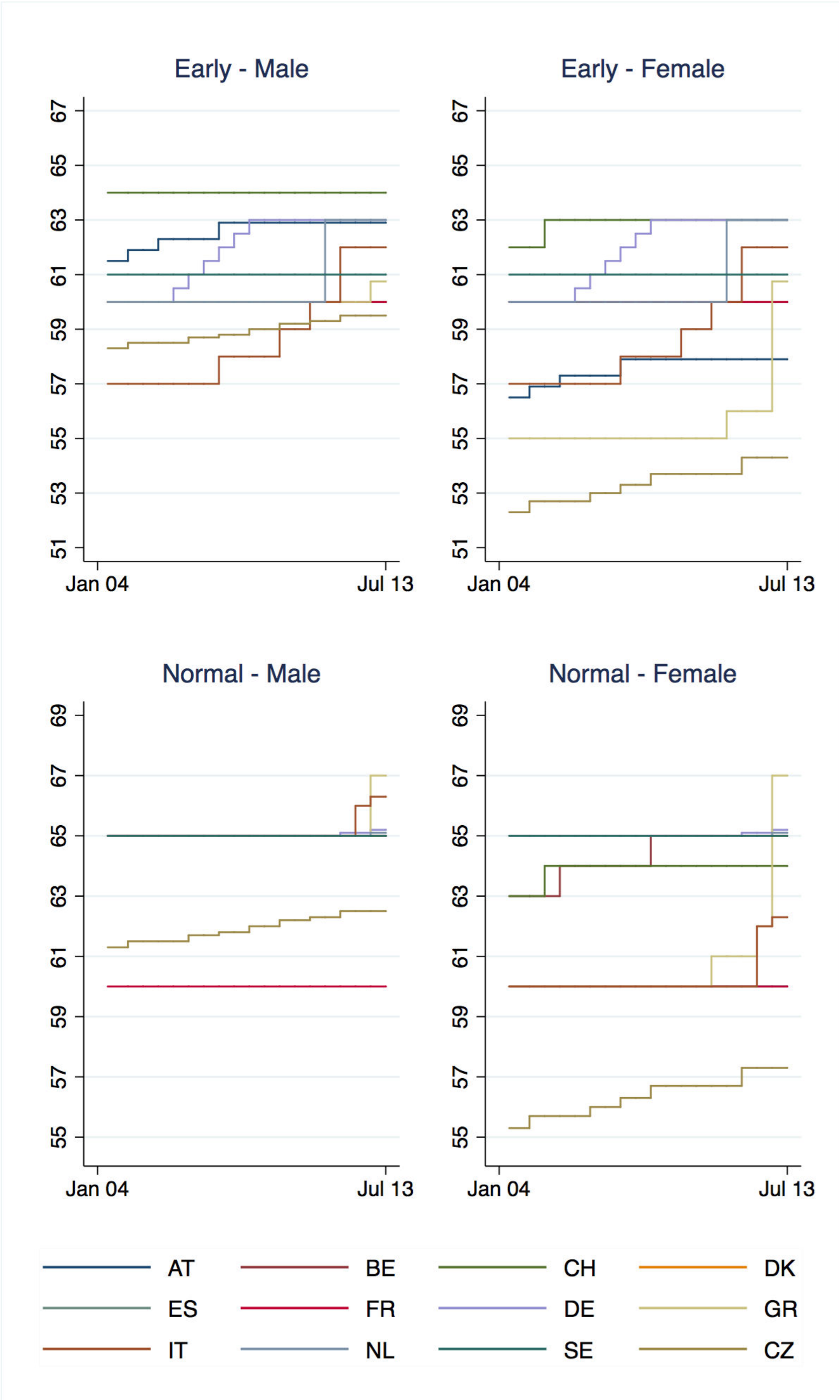


Figure 1: Variation in the retirement eligibility ages

Table 2: Employment rates at the retirement eligibility ages (%)

| Eligibility age | Full-time worker | Part-time worker | Full-time retiree |
|---|---|---|---|
| Under early ret. age | 67.83 | 23.74 | 8.43 |
| Between early and normal ret. age | 33.70 | 11.85 | 54.45 |
| Over the normal ret. age | 6.51 | 3.08 | 90.41 |
| Under early ret. age (P) | 56.33 | 18.52 | 25.15 |
| Between early and normal ret. age (P) | 27.88 | 12.26 | 59.86 |
| Over the normal ret. age (P) | 10.91 | 5.37 | 83.72 |

Notes: 1. P: Married or unmarried partner. 2. Other employment statuses, 'disabled', 'not in the labor force', and 'unemployed' are excluded from the analysis.

### 3.4 Descriptive statistics

Table 3 presents descriptive statistics for the full sample selected using the exclusion criteria in Section 3. It also presents the statistics for the first and last wave of the survey so that changes in the statistics can be compared over time.

Over the whole survey period, the average age of the sample is 62.5 years, where 14.3 percent are between the early and normal retirement ages and 43.6 percent are above the normal retirement age. 28.9 percent have a higher education. 80.1 percent of the sample are married or have a partner.

22.5 percent report that their health is fair or poor. The clinical threshold for depression based on the EURO-D Score is 4 or more symptoms. Since the average depression score is 1.89 out of 12, it appears that a significant fraction of the sample is close to or over the clinical threshold. As objective indicators of general physical health, the average number of difficulties in daily activities or in mobility or muscle use seems low. The average number of chronic diseases is 0.90 out of 9. 42.9 percent of the sample are overweight and 16.6 percent are obese. While the average score of the word recall test is just below half of its maximum, the average score of the numeracy test seems much higher.

33.6 percent of the sample report working 35 hours or more per week, while 13.2 percent report working less than 35 hours at the time of the survey.

There are plausible changes in the statistics between the first and last waves. The most notable change is that health status deteriorates across most health indicators.

Table 3: Descriptive statistics

| | Percent | | |
|---|---|---|---|
| | All waves | 2004 wave | 2013 wave |
| Demographics | | | |
| Age (50–75) (avg.) | 62.47 | 61.82 | 62.75 |
| Under early ret. age | 42.09 | 43.03 | 42.57 |
| Between early and normal ret. age | 14.27 | 16.31 | 13.78 |
| Over normal retirement age | 43.64 | 40.66 | 43.65 |
| High education | 28.89 | 24.56 | 31.73 |
| Spouse or unmarried partner | 80.09 | 79.01 | 79.96 |
| Female | 44.32 | 40.78 | 47.30 |
| Health status | | | |
| Self-perceived fair or poor health | 22.54 | 20.42 | 22.09 |
| Number of ADL limitations (0–5) (avg.) | 0.08 | 0.07 | 0.08 |
| Number of IADL limitations (0–5) (avg.) | 1.80 | 1.58 | 2.02 |
| Number of mobility limitations (0–5) (avg.) | 0.25 | 0.26 | 0.23 |
| Number of difficulties in muscle use (0–4) (avg.) | 0.43 | 0.45 | 0.42 |
| Number of chronic diseases (0–9) (avg.) | 0.90 | 0.89 | 0.91 |
| Hospital stay in the previous two years | 11.75 | 10.55 | 12.09 |
| Overweight | 42.94 | 44.87 | 41.55 |
| Obese | 16.63 | 15.84 | 16.82 |
| Word recall test score (0–20) (avg.) | 9.80 | 8.82 | 10.28 |
| Numeracy (0–5) (avg.) | 4.06 | 3.57 | 4.36 |
| Depression scale EURO-D (0–12) (avg.) | 1.89 | 1.86 | 1.90 |
| Fluency (0–100) (avg.) | 21.45 | 20.30 | 22.40 |
| Employment status | | | |
| Full-time worker | 33.64 | 34.61 | 33.76 |
| Part-time worker | 13.23 | 12.46 | 13.96 |
| Full-time retiree | 53.13 | 52.93 | 52.28 |
| N obs. | 70949 | 13390 | 25478 |
| N ind. | 41514 | 13390 | 25478 |

Notes: 1. Totals may not add due to rounding error. 2. Number of observations is based on the information available on employment status.

# 4 Exploratory graphical analysis

In our empirical approach, identification of the effects of working part-time and full-time on health relies on the discontinuities in the probabilities of working part-time and full-time upon reaching own or the partner's retirement eligibility ages. While identifying the effects of working part-time and full-time, we also condition on a continuous function of age of the individual. Here we provide exploratory graphical analysis of the jumps in the conditional mean of the treatment (the number of hours worked) and outcome (health) variables at the points of discontinuity in the assignment (retirement eligibility ages) variable. We also rely on the graphical analysis when motivating the functional forms of age in the first stage regressions of part-time and full-time work probabilities, and in the second stage regressions of health outcomes.

Based on univariate nonparametric regression, Figure 2 plots the individual weekly hours worked against the age of the individual, allowing for jumps at the retirement eligibility ages. We also draw 95 percent confidence bounds around each curve. Note, however, that the plot is based on univariate regression and does not control for the effect of the partner's age. Figure 3 plots the individual weekly hours worked against the age of the partner. For all countries, there are obvious discontinuities at the cutoff ages, and the jumps are in the expected direction. Except for Italy, the bounds never cross the curves, suggesting that the jumps are statistically significant. The jumps are more pronounced at the cut-off ages of the individual than at those of their partner, however. The jumps suggest that part-time and full-time work probabilities change significantly at the early and normal retirement ages, which supports our identification strategy. In the next section, we present formal tests of whether the dummy variables for the discontinuities are jointly powerful enough to serve as good instruments for both part-time and full-time work probabilities.

In Figures 4-9, six health indicators are plotted against the age of the individual, allowing for jumps at the retirement eligibility ages of the individual. Significant jumps are apparent at the retirement ages in the health index, word recall score, numeracy score, and depression score, for all countries.

The 2SLS regression model we employ requires us to specify a functional form of age that is same in the first and second stage regressions of weekly hours worked and health outcomes, respectively. However, Figures 2 and 3 suggest a cubic relationship between the weekly hours worked and the age of the individual, while Figures 4-9 suggest linear or quadratic relationships between the health outcomes and the age of the individual, although the curvature of the quadratic polynomials differ considerably across the health outcomes but not across the countries. Therefore, in our robustness analyses in Section 5.3, we present results from estimation of regression models employing different functional forms of age, and argue that a linear function of age is statistically significant in both the first and second stage regressions for the majority of the health outcomes.

# 5 Results

## 5.1 Instrument relevance and validity

Table 4 presents the coefficient estimates from the first stage estimation of the linear probability model with fixed effects given by Equation (2.5). The errors of the linear probability model are heteroskedastic by construction of the model, and the predictions of the model may lie outside the unit interval. We correct the standard errors of the estimates for heteroskedasticity. In 6898 cases the predictions of the model lie outside the unit interval for the regression explaining full-time work status. However, this does not affect the consistency of the fixed effects instrumental variables estimator we use. Dropping these cases also does not change our qualitative results.

The results show that the retirement eligibility ages of the respondent significantly decrease the probability of working part-time and full-time. The effects on working full-time are about three times larger than those on working part-time. This is plausible because the majority of the workers opt out of full-time work when they are eligible for retirement benefits, as suggested by Table 2. The retirement ages of the partner also appear to be predictive of the respondent's own retirement behavior, but only for full-time working, and to a considerably lesser extent. This shows that older workers become less inclined to work full-time once their partner is eligible for social security benefits which confirms our expectation.

Angrist and Pischke (2009, pp. 217-18) introduced an F statistic for testing weak identification when there is more than one endogenous regressor. We find evidence against weak identification for both endogenous regressors. Table 4 also shows that the retirement age indicators are jointly significant at the 0.01 level.

Table 5 presents the results of the overidentification test when we consider the retirement eligibility ages of both the respondent and the partner, which constitute a total of four instrumental variables for two potentially endogenous regressors. In all regressions, the test results support the use of these instruments: the null hypothesis that all moment restrictions are valid is not rejected. However, in the regression of self-perceived health, we fail to reject the test at the 0.10 level when we allow for quadratic and cubic age effects, but these functional forms of age do not appear to be the correct approximations of the age effect as we will argue in Section 5.3.

These results show that retirement ages are important predictors of both part-time and full-time work probabilities, when we control for a linear function of age.

Table 4: Results for first-stage FE model explaining part-time and full-time work status

| | Part-time | | Full-time | |
|---|---|---|---|---|
| | Coeff | p-val | Coeff | p-val |
| Bet. early and normal ret. age | −0.041 | 0.000 | −0.148 | 0.000 |
| At or over the normal ret. age | −0.104 | 0.000 | −0.303 | 0.000 |
| Bet. early and nor. ret. age (P) | 0.005 | 0.420 | −0.026 | 0.001 |
| At or over the normal ret. age (P) | −0.000 | 0.972 | −0.009 | 0.350 |
| Age | −0.004 | 0.000 | −0.020 | 0.000 |
| Constant | 0.440 | 0.000 | 1.769 | 0.000 |
| F-test for four instruments | | 0.000 | | 0.000 |
| AP test of weak identification | | 0.000 | | 0.000 |
| N obs. | 63964 | | 63964 | |
| N ind. | 38221 | | 38221 | |

Notes: 1. Linear probability model with fixed effects. 2. P: Married or unmarried partner. 3. Standard errors are robust to heteroskedasticity and clustering on panel groups.

## 5.2 Physical and mental health

Table 5 presents the baseline results from the estimation of linear probability models with instrumental variables and fixed effects given by Equation (2.6). The estimation makes use of the full set of four instruments introduced above.

A first finding is that in most regressions the linear age effect is significant at the 0.01 level. This confirms the linear relationships observed between age and self-perceived health and health index in Figures 4-5. However, for other health indicators Figures 6-9 suggest that other functional forms of age can better approximate the age effect. Furthermore, as discussed in Section 4, a cubic function of age might better capture the effect of age on weekly hours worked in the first stage of the 2SLS estimation. Therefore, in our robustness analyses in the next section, we discuss additional results based on quadratic and cubic age functions, and argue that the linear function of age approximates the age effect better than the nonlinear functions of age do.

Regarding labor market participation at the extensive margin, we find that working (part-time or full-time) has a significant effect on self-perceived health, based on the F-test of joint significance, in line with the findings of Coe and Zamarro (2011) and Neuman (2008), who showed that retired people have better self-perceived health in Europe and in the US, respectively. We also find significant effects for word recall score and depression score, but not for the other health indicators.

Regarding labor market participation at the intensive margin, surprisingly, we find that working part-time and full-time have opposite and significant effects in almost all regressions. The two effects are also often statistically different from each other at the conventional significance levels. The results on physical health show that working full-time has a negative effect on self-perceived health, while working part-time has a positive but insignificant effect. The results on the objective health index are in line with the results on self-perceived health, but provides additional evidence that working part-time has a positive and significant effect on general health.

The results on the body mass index show that working full-time reduces body weight while working part-time increases it. It could be that part-time workers are more likely to have a higher body mass index because they are probably physically less active. However, we could expect that as part-time workers are more likely to have a higher body weight, they are also more prone to diseases related to being overweight or obese, such as diabetes or heart attacks, and therefore also have a lower health index. Our results do not confirm this expectation. Furthermore, we do not exclude the possibility that full-time workers are not physically active during work time as they have a desk job, while part-time workers have more spare time to spend on physical activities.

With respect to cognition, we find that working part-time has a large and positive effect on word recall score, while working full-time has a relatively smaller and negative effect. The results on numeracy score show that working part-time has a negative effect but working full-time has a positive effect, while both effects are at odds with the results on word recall score. Finally, we find that working part-time has a large and positive effect on the depression score while working full-time has a smaller and negative effect.

Overall, these results suggest that labor market participation has significant and surprisingly large effects on both the physical and mental health outcomes at the intensive margin, while some of these effects are not apparent at the extensive margin as we would have found if we have analysed the effect of retirement following the common practice in the earlier studies of the subject literature.

Table 5: Results for IV-FE model explaining health outcomes

| | Self-perceived health | | Health index | | Body mass index | | Word recall score | | Numeracy | | Depression score | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Model: LA, CE, 6 IV, FE | | | | | | | | | | | | |
| Part-time | −0.714 | 0.379 | −0.861 | 0.058‡ | 5.295 | 0.066‡ | 10.122 | 0.024‡ | −2.373 | 0.063‡ | −5.612 | 0.024‡ |
| Full-time | 0.572 | 0.041 | 0.282 | 0.037 | −1.779 | 0.044 | −3.942 | 0.012 | 0.832 | 0.060 | 2.272 | 0.008 |
| Age | 0.047 | 0.000 | 0.011 | 0.000 | 0.030 | 0.006 | 0.015 | 0.417 | 0.093 | 0.000 | 0.042 | 0.000 |
| End. test | | 0.000 | | 0.000 | | 0.001 | | 0.000 | | 0.040 | | 0.000 |
| Ove. test | | 0.143 | | 0.615 | | 0.439 | | 0.861 | | 0.241 | | 0.393 |
| F-test for employment terms | | 0.000 | | 0.110 | | 0.130 | | 0.037 | | 0.167 | | 0.019 |
| N obs. | 43248 | | 25105 | | 25747 | | 42531 | | 43069 | | 42389 | |
| N ind. | 17518 | | 10435 | | 10697 | | 17215 | | 17441 | | 17170 | |

Notes: 1. LA: Linear age. CE: Contemporaneous effect. 2. Linear model with instrumental variables and fixed effects. 3. Self-perceived health: 1 (Excellent), ..., 5 (poor). Health index takes similar values. Body mass index takes values from 13.5 to 77.2. Higher values indicate increasing body weight. Word recall score takes values from 0 to 20. Higher values indicate better memory. Numeracy takes values from 1 (bad) to 5 (good). Depression score takes values from 0 to 12. Higher values indicate more severe depression. 4. Standard errors are robust to heteroskedasticity and clustering on panel groups. 5. The double dagger symbol (‡) indicates the cases where equality of the coefficients of part-time and full-time is rejected at the 0.10 or smaller significance levels. 6. Endogeneity test tests the null hypothesis that the variables ‘part-time’ and ‘full-time’ are exogenous. The test is based on the C statistic. Overidentification test tests the null hypothesis that all instruments are uncorrelated with the unobserved error. The test is based on the Hansen J statistic. Both tests are robust to heteroskedasticity and clustering on panel groups. The F-test tests the null hypothesis that the coefficients of the age terms, or those of part-time and full-time working, are zero.

### 5.3 Robustness checks

#### Age specification

Our 2SLS regression model has allowed for a linear function of age to capture the changes in the health status due to advancing age observed in Figures 5-9. Table 5 showed that the linear age term is significant at the 0.01 level in most health outcome regressions. In fact, Figures 4 and 5 suggested that a linear function of age would capture well the changes in self-perceived health and health index due to advancing age. However, we find no significant effect for the linear age term in the word recall score regression. Furthermore, 5-9 suggest nonlinear relationships between the other health indicators and age. 2 also suggested that a cubic function of age might better capture the nonlinear changes in weekly hours worked due to advancing age.

Tables 6 and 7 present the results from the first stage regressions of part-time and full-time working when we employ quadratic and cubic functions of age, instead of a linear function. We also reproduce the results from the estimation of the model with a linear function of age for ease of comparison of the results. The results show that when we consider a cubic function of age, the effect of the cubic age term is virtually zero in both regressions of part-time and full-time working. Likewise, when we consider a quadratic function of age, the effect of the quadratic age term is virtually zero. Furthermore, note that the explanatory power of the instruments decrease when we allow for nonlinear continuous functions of age, especially in the regression of full-time work. We conclude that the linear function of age explains the age-related changes in part-time and full-time work preferences sufficiently well, while at the same time allows the instruments to maintain their explanatory power in explaining the discontinuous changes in work preferences around the retirement eligibility ages.

Table 8 presents the results from the second stage stage regressions of health outcomes when we employ quadratic and cubic functions of age. We also reproduce the results from the estimation of the health outcome regressions using a linear function of age. A first main finding is that when we consider a cubic function of age, the effect of the cubic age term is virtually zero and often not significant in the health outcome regressions. Furthermore, the coefficient of the quadratic age term is close to zero in all regressions employing a quadratic age function. These results suggest that health outcomes are fairly linearly related to age, when we account for the effects of working part-time and full-time. Viewed in combination with the results from the sensitivity analysis of the first stage regressions, we conclude that correct specification of the first and second stage equations of the 2SLS model requires age to enter these equations linearly.

A second main finding is that the coefficients of the main variables of interest are generally smaller in magnitude, and less precisely estimated when we allow for quadratic and cubic functions of age. Note that the C statistic for testing the endogeneity of the part-time and full-time work decisions also loses power in most of the regressions when we employ flexible functional forms of age. The obvious reason is that the instruments lose their explanatory power when we employ flexible functional forms of age, even though we observe only small changes in the magnitudes and significance of the effects of the instruments across the models employing linear and quadratic functional forms in Tables 6 and 7.

A third finding is that the effect of working full-time in the health index regression, and the effects of working part-time and full-time in the numeracy score regression remain significant at the 0.10 level in both models employing quadratic and cubic functions of age. This shows that our qualitative results for these two health outcomes in Section 5 are fairly robust to the functional forms of age.

A final result is that the J statistic for testing the exogeneity of the instruments loses power when we employ flexible functional forms of age in the self-perceived health regression.

### Instrument set

To analyze the causal effect of part-time and full-time working on health outcomes, we have used the retirement eligibility ages of the respondent as instruments for work behavior, and supplemented this instrument set with the retirement ages of the married or unmarried partner. Supplementing the instrument set with the retirement ages of the partner can affect the efficiency of our instrumental variables estimator in two directions. It can improve the efficiency of the estimator, yielding more precise and significant effects, since the predictive power of the instrument set increases. On the other hand, it can reduce the efficiency of the estimator since the number of observations used in the estimation decreases as we restrict the sample to those respondents with a partner only so that the retirement ages of the partner can serve as instruments. In order to investigate the sensitivity of the estimates for restricting the instrument set to the retirement ages of the respondent, the top panel of Table 9 presents the results using the retirement eligibility ages of the respondent only.

Compared to the results using the full instrument set in Table 5, we find that in all regressions the coefficients become less significant, and the magnitudes of the coefficients often become larger. It seems obvious that this is due that the predictive power of the instrument set has decreased. We conclude that the retirement ages of the partner improve the efficiency of the instrumental variables estimator, yielding more significant effects.

### Econometric model

Our econometric model makes use of instrumental variables to circumvent the endogeneity of hours worked, but also exploits the panel nature of the data to allow for fixed effects that control for unobserved individual heterogeneity. To show the extent to which the endogeneity of hours worked and individual heterogeneity affect the estimated coefficients, the bottom panel of Table 9 presents the results using three alternative regression models. In the first model, we do not exploit the panel dimension of the data and do not control for the endogeneity of hours worked; instead we follow a pooled OLS estimation. In the second model, we do not allow for endogeneity of hours worked, but exploit the panel dimension of the data and follow a panel FE estimation which uses the within group estimator (the within group transformation followed by OLS). In the third model, we do not exploit the panel dimension of the data, but allow for endogeneity of hours worked and follow a pooled IV estimation that uses the two-stage least squares estimator. The baseline panel IV-FE model in Table 5 uses the two-stage least squares estimator after the within group transformation.

A first result is that the coefficients become insignificant in almost all regressions when we do not control for the endogeneity of hours worked, but control for fixed effects. Both the signs and the magnitudes of the coefficients change when we do not control for the endogeneity of hours worked nor for fixed effects. This suggests that health conditions affect labor supply decisions at the intensive margin in old age.

The second result is that the signs and the magnitudes of the effects change when we control for fixed effects, regardless of if we take an instrumental variables approach. This suggests that individuals have health-related unobserved characteristics that are also correlated with their labor market behavior.

Overall, the results suggest that controlling for the endogeneity of hours worked, and for individual heterogeneity are essential in the analysis of the effect of labor market activity on health outcomes at older ages.

Table 6: Robustness checks for first-stage FE model explaining part-time work status

| | Part-time | | | | | |
|---|---|---|---|---|---|---|
| | Linear age | | Quadratic age | | Cubic age | |
| | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Bet. early and normal ret. age | −0.041 | 0.000 | −0.042 | 0.000 | −0.030 | 0.000 |
| At or over the normal ret. age | −0.104 | 0.000 | −0.104 | 0.000 | −0.080 | 0.000 |
| Bet. early and nor. ret. age (P) | 0.005 | 0.420 | −0.005 | 0.442 | 0.007 | 0.291 |
| At or over the normal ret. age (P) | −0.000 | 0.972 | −0.000 | 0.978 | 0.002 | 0.756 |
| Age | −0.004 | 0.000 | −0.002 | 0.727 | 0.450 | 0.000 |
| Age squared | | | 0.000 | 0.671 | −0.007 | 0.000 |
| Age cubed | | | | | 0.000 | 0.000 |
| Constant | 0.440 | 0.000 | 0.372 | 0.044 | −8.864 | 0.000 |
| F-test for two age terms | | | | 0.000 | | |
| F-test for three age terms | | | | | | 0.000 |
| F-test for four instruments | | 0.000 | | 0.000 | | 0.000 |
| AP test of weak identification | | 0.000 | | 0.000 | | 0.000 |
| N obs. | | | | | | |
| N ind. | | | | | | |

Notes: 1. Linear probability model with fixed effects. 2. P: Married or unmarried partner. 3. Standard errors are robust to heteroskedasticity and clustering on panel groups.

Table 7: Robustness checks for first-stage FE model explaining full-time work status

| | Full-time | | | | | |
|---|---|---|---|---|---|---|
| | Linear age | | Quadratic age | | Cubic age | |
| | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Bet. early and normal ret. age | −0.148 | 0.000 | −0.137 | 0.000 | −0.088 | 0.000 |
| At or over the normal ret. age | −0.303 | 0.000 | −0.301 | 0.000 | −0.197 | 0.000 |
| Bet. early and nor. ret. age (P) | −0.026 | 0.001 | −0.021 | 0.011 | −0.012 | 0.131 |
| At or over the normal ret. age (P) | −0.009 | 0.350 | −0.010 | 0.275 | −0.000 | 0.963 |
| Age | −0.020 | 0.000 | −0.072 | 0.000 | 1.886 | 0.000 |
| Age squared | | | 0.000 | 0.000 | −0.031 | 0.000 |
| Age cubed | | | | | 0.000 | 0.000 |
| Constant | 1.769 | 0.000 | 3.391 | 0.000 | −36.594 | 0.000 |
| F-test for two age terms | | | | 0.000 | | |
| F-test for three age terms | | | | | | 0.000 |
| F-test for four instruments | | 0.000 | | 0.000 | | 0.000 |
| AP test of weak identification | | 0.000 | | 0.000 | | 0.000 |
| N obs. | | | | | | |
| N ind. | | | | | | |

Notes: 1. Linear probability model with fixed effects. 2. P: Married or unmarried partner. 3. Standard errors are robust to heteroskedasticity and clustering on panel groups.

Table 8: Robustness checks on the functional form of age for IV-FE model explaining health outcomes

| | Self-perceived health | | Health index | | Body mass index | | Word recall score | | Numeracy | | Depression score | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Model: LA, CE, 4 IV, FE | | | | | | | | | | | | |
| Part-time | −0.714 | 0.379 | −0.861 | 0.058‡ | 5.295 | 0.066‡ | 10.122 | 0.024‡ | −2.373 | 0.063‡ | −5.612 | 0.024‡ |
| Full-time | 0.572 | 0.041 | 0.282 | 0.037 | −1.779 | 0.044 | −3.942 | 0.012 | 0.832 | 0.060 | 2.272 | 0.008 |
| Age | 0.047 | 0.000 | 0.011 | 0.000 | 0.030 | 0.006 | 0.015 | 0.417 | 0.093 | 0.000 | 0.042 | 0.000 |
| End. test | | 0.000 | | 0.000 | | 0.001 | | 0.000 | | 0.040 | | 0.000 |
| Ove. test | | 0.143 | | 0.615 | | 0.439 | | 0.861 | | 0.241 | | 0.393 |
| F-test for employment terms | | 0.000 | | 0.110 | | 0.130 | | 0.037 | | 0.167 | | 0.019 |
| Model: QA, CE, 4 IV, FE | | | | | | | | | | | | |
| Part-time | −0.231 | 0.825 | −0.462 | 0.175 | 3.058 | 0.198 | 3.991 | 0.338 | −4.017 | 0.084‡ | −3.444 | 0.193 |
| Full-time | 0.405 | 0.265 | 0.175 | 0.082 | −1.150 | 0.115 | −1.801 | 0.221 | 1.400 | 0.085 | 1.521 | 0.099 |
| Age | −0.001 | 0.955 | −0.032 | 0.000 | 0.296 | 0.000 | 0.533 | 0.000 | 0.247 | 0.001 | −0.110 | 0.191 |
| Age squared | 0.000 | 0.124 | 0.000 | 0.000 | −0.002 | 0.000 | −0.004 | 0.000 | −0.001 | 0.023 | 0.001 | 0.057 |
| End. test | | 0.002 | | 0.042 | | 0.065 | | 0.083 | | 0.008 | | 0.108 |
| Ove. test | | 0.076 | | 0.632 | | 0.883 | | 0.550 | | 0.278 | | 0.324 |
| F-test for employment terms | | 0.000 | | 0.150 | | 0.233 | | 0.233 | | 0.221 | | 0.087 |
| F-test for age terms | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 |
| Model: CA, CE, 4 IV, FE | | | | | | | | | | | | |
| Part-time | −0.336 | 0.749 | −0.493 | 0.187 | 3.080 | 0.209 | 3.609 | 0.384 | −3.852 | 0.097‡ | −3.141 | 0.218 |
| Full-time | 0.499 | 0.253 | 0.242 | 0.082 | −1.173 | 0.219 | −2.355 | 0.177 | 1.735 | 0.075 | 1.649 | 0.119 |
| Age | −0.290 | 0.566 | −0.353 | 0.112 | 0.375 | 0.803 | 4.356 | 0.030 | −1.924 | 0.085 | −1.450 | 0.247 |
| Age squared | 0.005 | 0.544 | 0.005 | 0.125 | −0.003 | 0.889 | −0.066 | 0.047 | 0.034 | 0.066 | 0.022 | 0.271 |
| Age cubed | −0.000 | 0.579 | −0.000 | 0.152 | 0.000 | 0.958 | 0.003 | 0.067 | −0.000 | 0.061 | −0.000 | 0.305 |
| End. test | | 0.015 | | 0.054 | | 0.250 | | 0.059 | | 0.000 | | 0.168 |
| Ove. test | | 0.090 | | 0.619 | | 0.884 | | 0.471 | | 0.259 | | 0.269 |
| F-test for employment terms | | 0.016 | | 0.195 | | 0.442 | | 0.151 | | 0.202 | | 0.219 |
| F-test for age terms | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 | | 0.000 |

Notes: 1. CA: Cubic age. LA: Linear age. QA: Quadratic age. CE: Contemporaneous effect. 2. Self-perceived health: 1 (Excellent), ..., 5 (poor). Health index takes similar values. Body mass index takes values from 13.5 to 77.2. Higher values indicate increasing body weight. Word recall score takes values from 0 to 20. Higher values indicate better memory. Numeracy takes values from 1 (bad) to 5 (good). Depression score takes values from 0 to 8. Higher values indicate more severe depression. 3. Standard errors are robust to heteroskedasticity and clustering on panel groups. However, the latter correction is not done in the Pooled OLS and Pooled IV regressions so that the FE and IV-FE regressions fully reflect the effect of exploiting the panel dimension of the data. 4. The double dagger symbol (‡) indicates the cases where equality of the coefficients of part-time and full-time is rejected at the 0.05 level. 5. Endogeneity test tests the null hypothesis that the variables 'part-time' and 'full-time' are exogenous. The test is based on the C statistic. Overidentification test tests the null hypothesis that all instruments are uncorrelated with the unobserved error. The test is based on the Hansen J statistic. Both tests are robust to heteroskedasticity and clustering on panel groups.

Table 9: Robustness checks on the econometric model for IV-FE model explaining health outcomes

| | Self-perceived health | | Health index | | Body mass index | | Word recall score | | Numeracy | | Depression score | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val | Coeff | p-val |
| Model: LA, CE, 2 IV, FE | | | | | | | | | | | | |
| Part-time | 0.037 | 0.971 | −1.091 | 0.101‡ | 8.584 | 0.113 | 10.900 | 0.105‡ | −3.941 | 0.099‡ | −10.304 | 0.082 |
| Full-time | 0.286 | 0.424 | 0.359 | 0.080 | −3.013 | 0.079 | −4.291 | 0.076 | 1.462 | 0.085 | 3.878 | 0.066 |
| Age | 0.043 | 0.000 | 0.012 | 0.000 | 0.021 | 0.124 | 0.028 | 0.146 | 0.097 | 0.000 | 0.043 | 0.009 |
| End. test | | 0.000 | | 0.000 | | 0.000 | | 0.001 | | 0.003 | | 0.000 |
| Ove. test | | – | | – | | – | | – | | – | | – |
| F-test for employment terms | | 0.000 | | 0.210 | | 0.182 | | 0.157 | | 0.217 | | 0.166 |
| Model: LA, CE, Pooled OLS | | | | | | | | | | | | |
| Part-time | −0.276 | 0.000‡ | −0.076 | 0.000‡ | −1.385 | 0.000‡ | 0.804 | 0.000‡ | 0.113 | 0.000‡ | 0.058 | 0.030‡ |
| Full-time | −0.367 | 0.000 | −0.104 | 0.000 | −0.596 | 0.000 | 0.411 | 0.000 | 0.182 | 0.000 | −0.357 | 0.000 |
| Model: LA, CE, FE | | | | | | | | | | | | |
| Part-time | 0.049 | 0.014 | 0.003 | 0.580 | −0.059 | 0.269 | −0.093 | 0.201 | −0.004 | 0.866 | 0.144 | 0.001 |
| Full-time | 0.026 | 0.121 | 0.008 | 0.160 | −0.021 | 0.646 | −0.041 | 0.500 | 0.002 | 0.926 | 0.165 | 0.000 |
| Model: LA, CE, Pooled 4 IV | | | | | | | | | | | | |
| Part-time | 1.105 | 0.000‡ | 0.341 | 0.001‡ | −0.993 | 0.423 | 4.330 | 0.000‡ | −1.731 | 0.000‡ | 3.028 | 0.000‡ |
| Full-time | −0.584 | 0.000 | −0.201 | 0.000 | −1.481 | 0.000 | −1.258 | 0.000 | 0.880 | 0.000 | −1.395 | 0.000 |
| End. test | | 0.000 | | 0.000 | | 0.001 | | 0.000 | | 0.000 | | 0.000 |
| Ove. test | | 0.000 | | 0.000 | | 0.051 | | 0.238 | | 0.217 | | 0.000 |

Notes: 1. LA: Linear age. CE: Contemporaneous effect. 2. Self-perceived health: 1 (Excellent), ..., 5 (poor). Health index takes similar values. Body mass index takes values from 13.5 to 77.2. Higher values indicate increasing body weight. Word recall score takes values from 0 to 20. Higher values indicate better memory. Numeracy takes values from 1 (bad) to 5 (good). Depression score takes values from 0 to 8. Higher values indicate more severe depression. 3. Standard errors are robust to heteroskedasticity and clustering on panel groups. However, the latter correction is not done in the Pooled OLS and Pooled IV regressions so that the FE and IV-FE regressions fully reflect the effect of exploiting the panel dimension of the data. 4. The double dagger symbol (‡) indicates the cases where equality of the coefficients of part-time and full-time is rejected at the 0.05 level. 5. Endogeneity test tests the null hypothesis that the variables 'part-time' and 'full-time' are exogenous. The test is based on the C statistic. Overidentification test tests the null hypothesis that all instruments are uncorrelated with the unobserved error. The test is based on the Hansen J statistic. Both tests are robust to heteroskedasticity and clustering on panel groups.

# 6 Conclusion

In this paper we have examined the causal effects of working part-time and full-time on the physical and mental health conditions of elderly people in Europe. We have accounted for unobserved heterogeneity across individuals that are likely to bias the causal effects examined. Identification is achieved through exploiting the rich variation in the eligibility ages both within and across 12 countries as well as over time.

Our findings suggest that working full-time negatively influences the objective general health, but working part-time has a positive and larger effect. On the other hand, we find that part-time workers are more likely to have a higher body mass index, while the opposite is true for full-time workers. With respect to cognition, we find mixed effects. Working part-time has a large and positive effect on word recall score, while working full-time has a relatively smaller and negative effect. On the other hand, the opposite holds for results on numeracy score. Depressive symptoms are increased by full-time work, while they are decreased by a larger magnitude by part-time work. Both effects are significant at the conventional significance level of 0.05.

Overall, all the effects on physical and mental health are large in absolute terms which is an important finding considering the recent trend of increasing physical and mental health problems in the population and their potential effects on health expenditures and labour market outcomes. In these respects our findings seem relevant for policy making to improve the work and health conditions of older people.

Kantarcı (2016) has found that working part-time and full-time have negative effects on self-perceived general health and memory, while they decrease body weight among the elderly people in the United States. He also finds that the magnitude of the effect of working part-time is substantially larger than that of working full-time, while both effects act in the same direction. Like Kantarcı, we also find that the magnitude of the effect of working part-time is substantially larger than that of full-time, but, surprisingly, part-time and full-time working have opposite health effects in Europe. Furthermore, self-perceived health, body mass index, and cognition appear as the same health indicators where we find significant effects as in the United States. These comparisons suggest that United States and Europe differ in terms of how working part-time and full-time affect health in old age, but not always by the types of the health outcomes affected. Future research might aim at explaining the differences in health outcomes due to working different number of hours in old age in the United States and Europe.


# Acknowledgments

This paper uses data from SHARE Wave 5 release 1.0.0, as of 31.03.2015 (DOI: 10.6103/SHARE.w5.100) or SHARE Wave 4 release 1.1.1, as of 28.03.2013 (DOI: 10.6103/SHARE.w4.111) or SHARE Waves 1 and 2 release 2.6.0, as of 29.11.2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1.0.0, as of 24.11.2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N 211909, SHARE-LEAP, N 227822 and SHARE M4, N 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

## References


Eric Bonsang, Stphane Adam, and Sergio Perelman. Does retirement affect cognitive functioning? *Journal of Health Economics*, 31(3):490–501, 2012. doi: 10.1016/j.jhealeco.2012.03.005.

Kerwin Kofi Charles. Is retirement depressing? Labor force inactivity and psychological well-being in later life. *Research in Labor Economics*, 23:269–299, 2004.

Norma B. Coe and Gema Zamarro. Retirement effects on health in Europe. *Journal of Health Economics*, 30(1):77–86, 2011.

Fabrizio Mazzonna and Franco Peracchi. Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710, 2012. doi: 10.1016/j.euroecorev.2012.03.004.

Kevin Neuman. Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, 29(2):177–201, 2008.

Susann Rohwedder and Robert J. Willis. Mental retirement. *Journal of Economic Perspectives*, 24 (1):119–138, 2010.

Tunga Kantarcı. Effects of working part-time and full-time work on physical and mental health in old age. *Netspar Discussion Paper*, No. 10/2013-037, 2016.

Irina Murtazashvilia and Jeffrey M. Wooldridge. Fixed effects instrumental variables estimation in correlated random coefficient panel data models. *Journal of Econometrics*, 142(1):539–552, 2008. doi: 10.1016/j.jeconom.2007.09.001.

Brenda Gannon and Jennifer Roberts. Part-time work and health among older workers in Ireland and Britain. *Applied Economics*, 43(30):4749–4757, 2011.

John Bound, Micheal Schoenbaum, Todd R. Stinebrickner, and Timothy Waidmann. The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, 6(2):179–202, 1999.

Floortje Mols, Melissa S. Y. Thong, Pauline Vissers, Tamar Nijsten, and Lonneke V. van de Poll-Franse. Socio-economic implications of cancer survivorship: Results from the PROFILES registry. *European Journal of Cancer*, 48(13):2037–2042, 2012. doi: 10.1016/j.ejca.2011.11.030.

David S. Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010. doi: 10.1257/jel.48.2.281.

John Bound. Self-reported versus objective measures of health in retirement models. *The Journal of Human Resources*, 26(1):106–138, 1991.

David M. Blau. Labor force dynamics of older married couples. *Journal of Labor Economics*, 16(3):595–629, 1998. doi: http://www.jstor.org/stable/10.1086/209900.

Alan L. Gustman and Thomas L. Steinmeier. Retirement in dual-career families: A structural model. *Journal of Labor Economics*, 18(3):503–545, 2000. doi: 10.1086/209968.

Alan L. Gustman and Thomas L. Steinmeier. Social security, pensions, and retirement behavior within the family. *Journal of Applied Econometrics*, 19(6):723–737, 2004. doi: 10.1002/jae.753.

Joshua D. Angrist and Jrn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

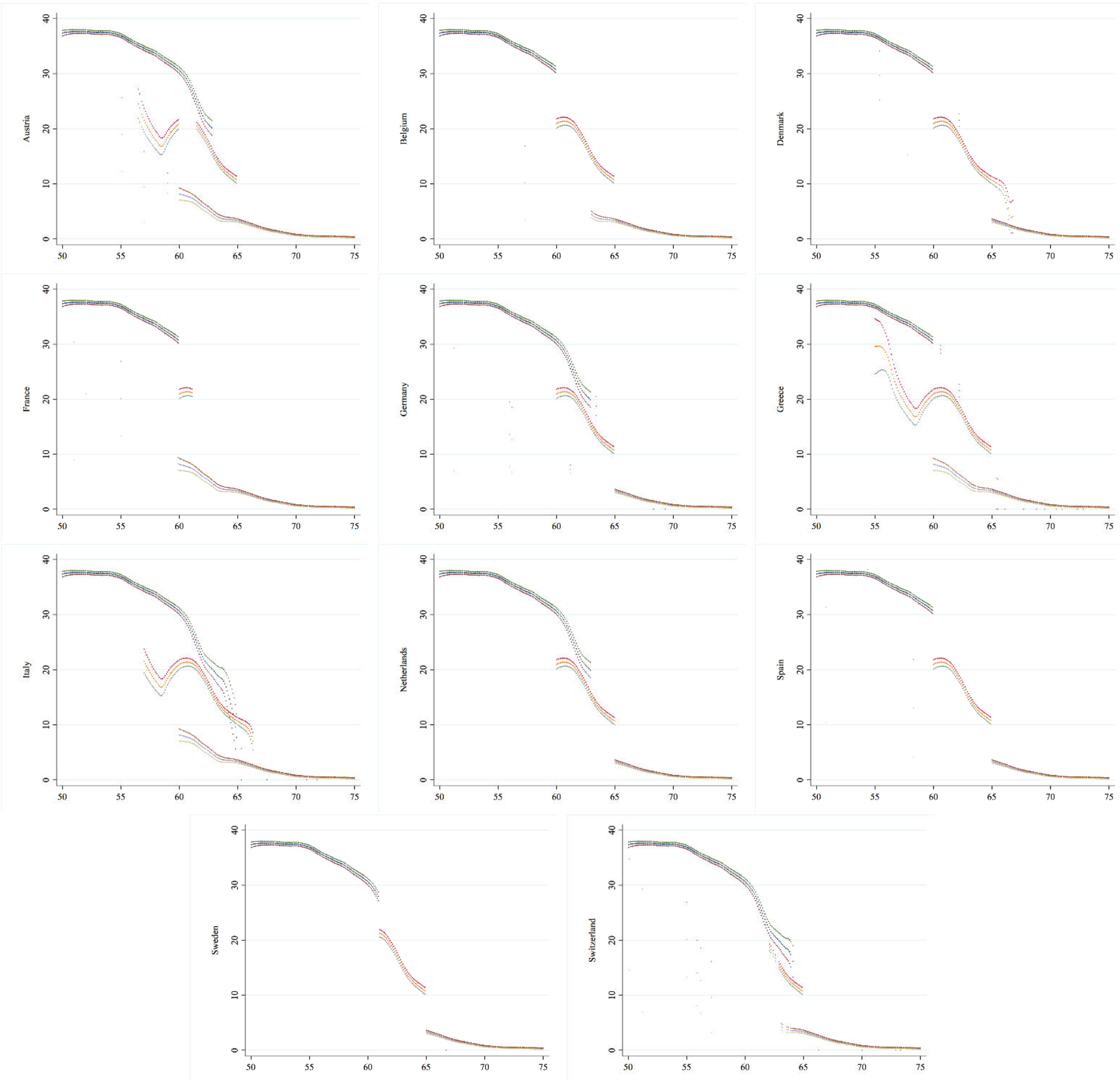


Figure 2: Hours worked per week by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

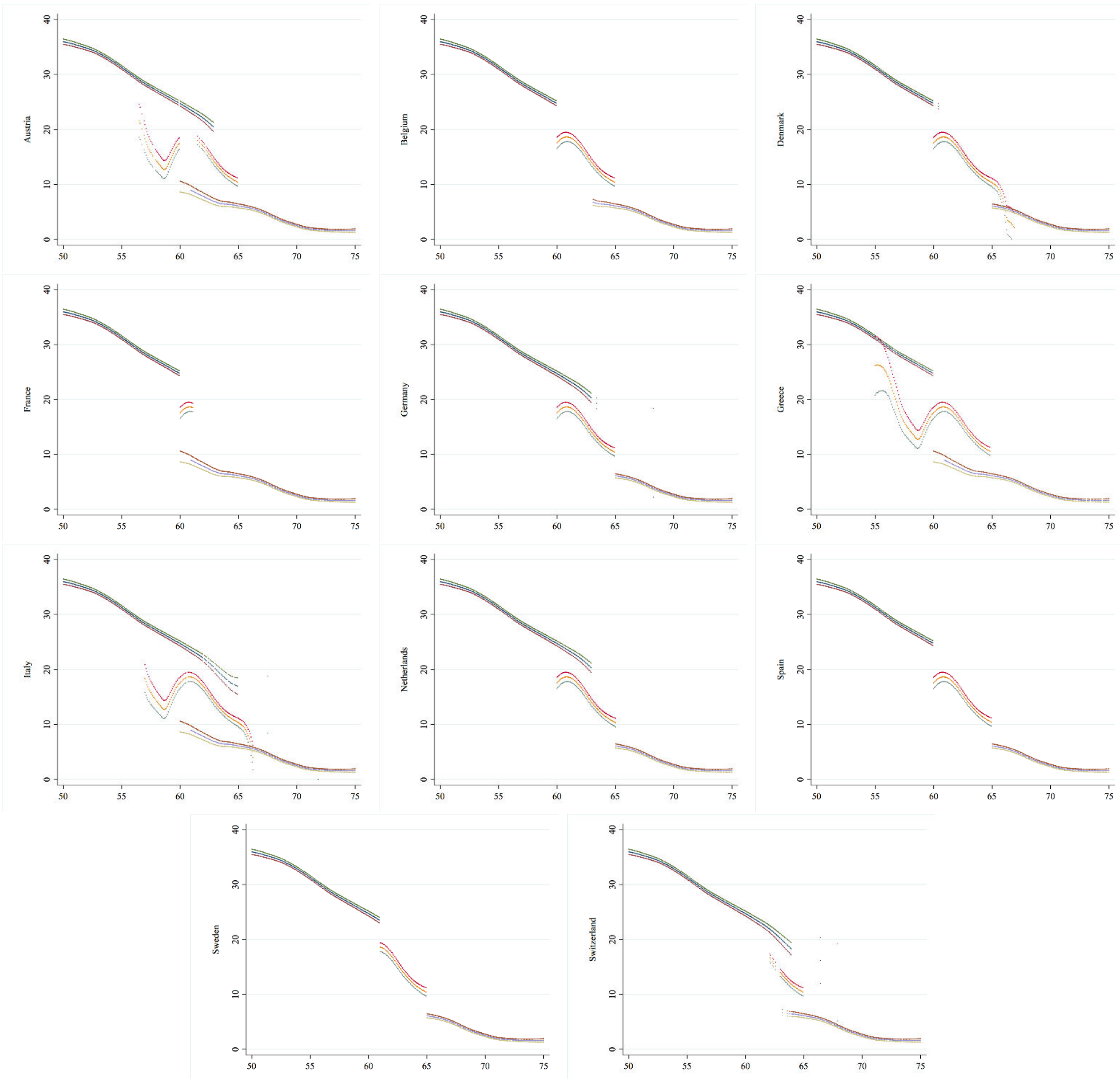


Figure 3: Hours worked per week by age of the partner. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

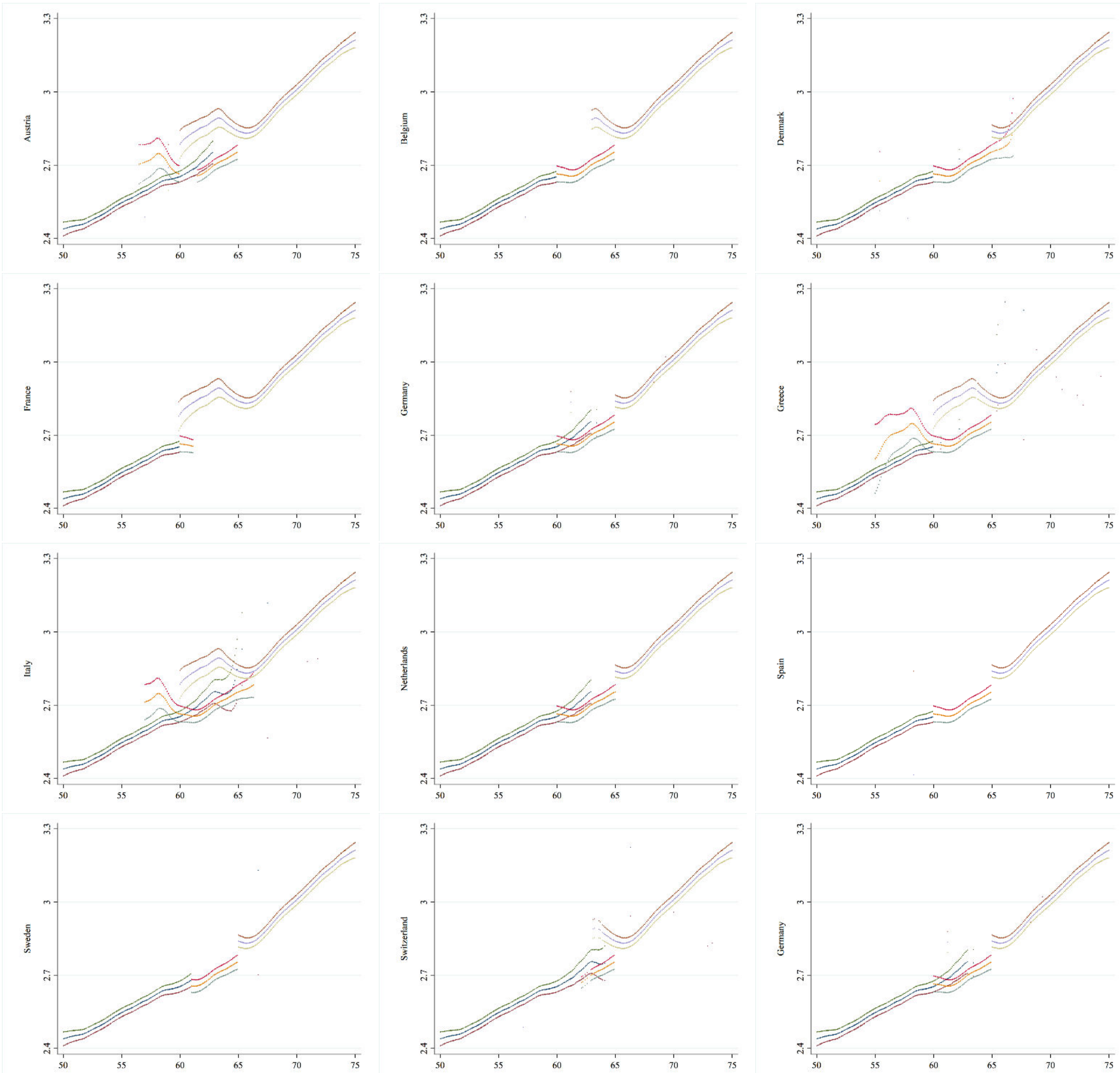


Figure 4: Self-perceived health by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

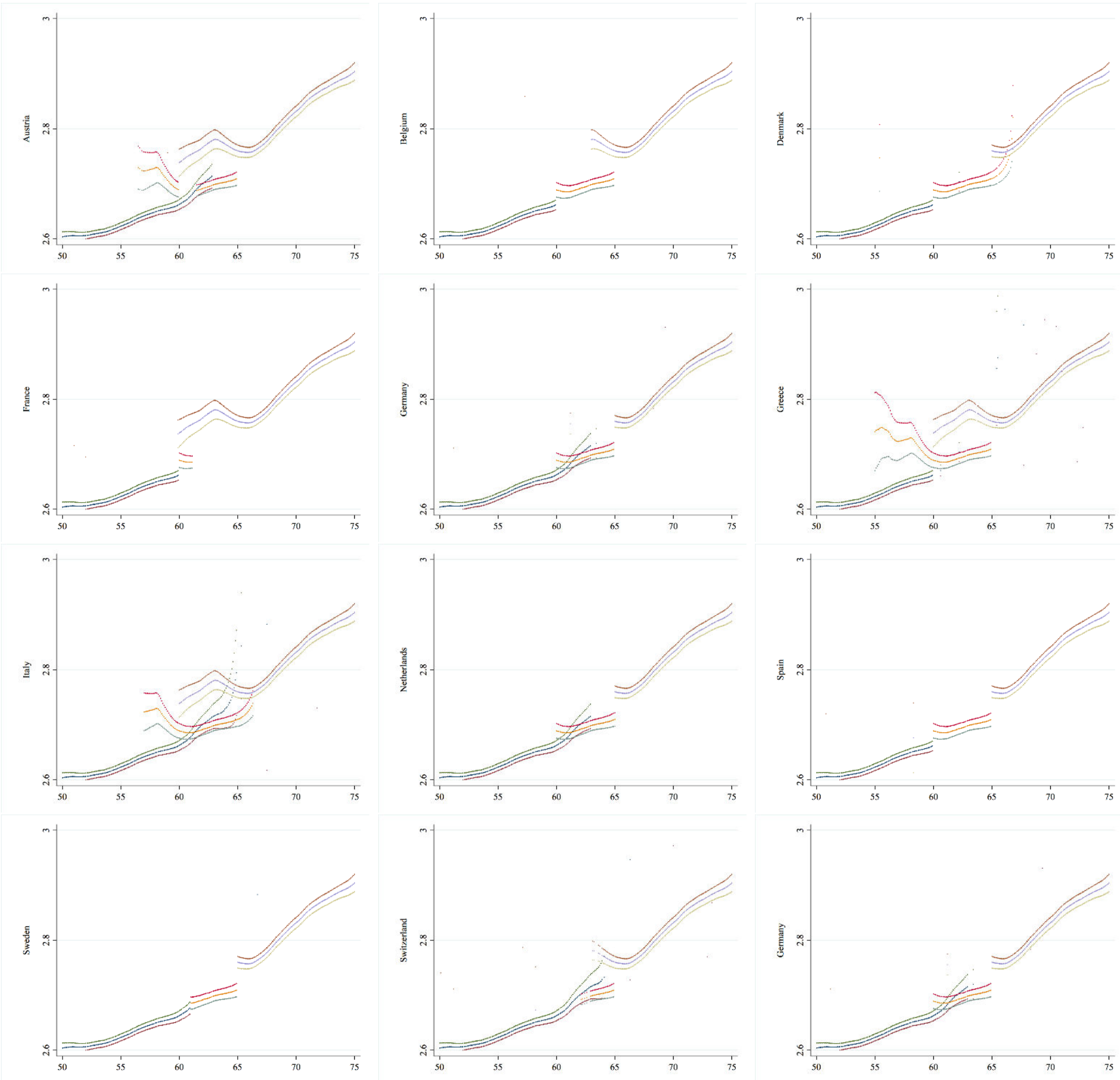


Figure 5: Health index by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

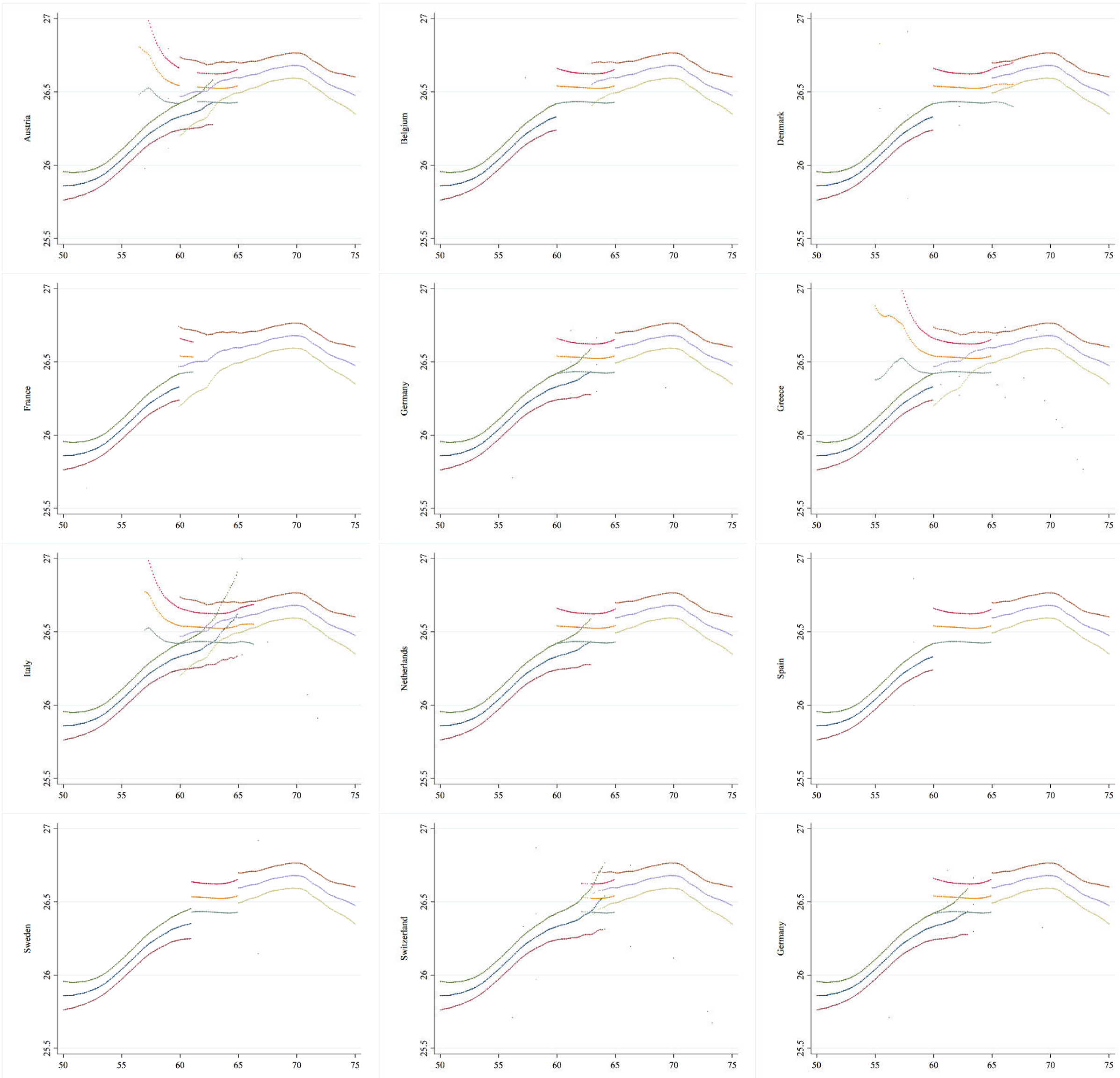


Figure 6: Body mass index by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

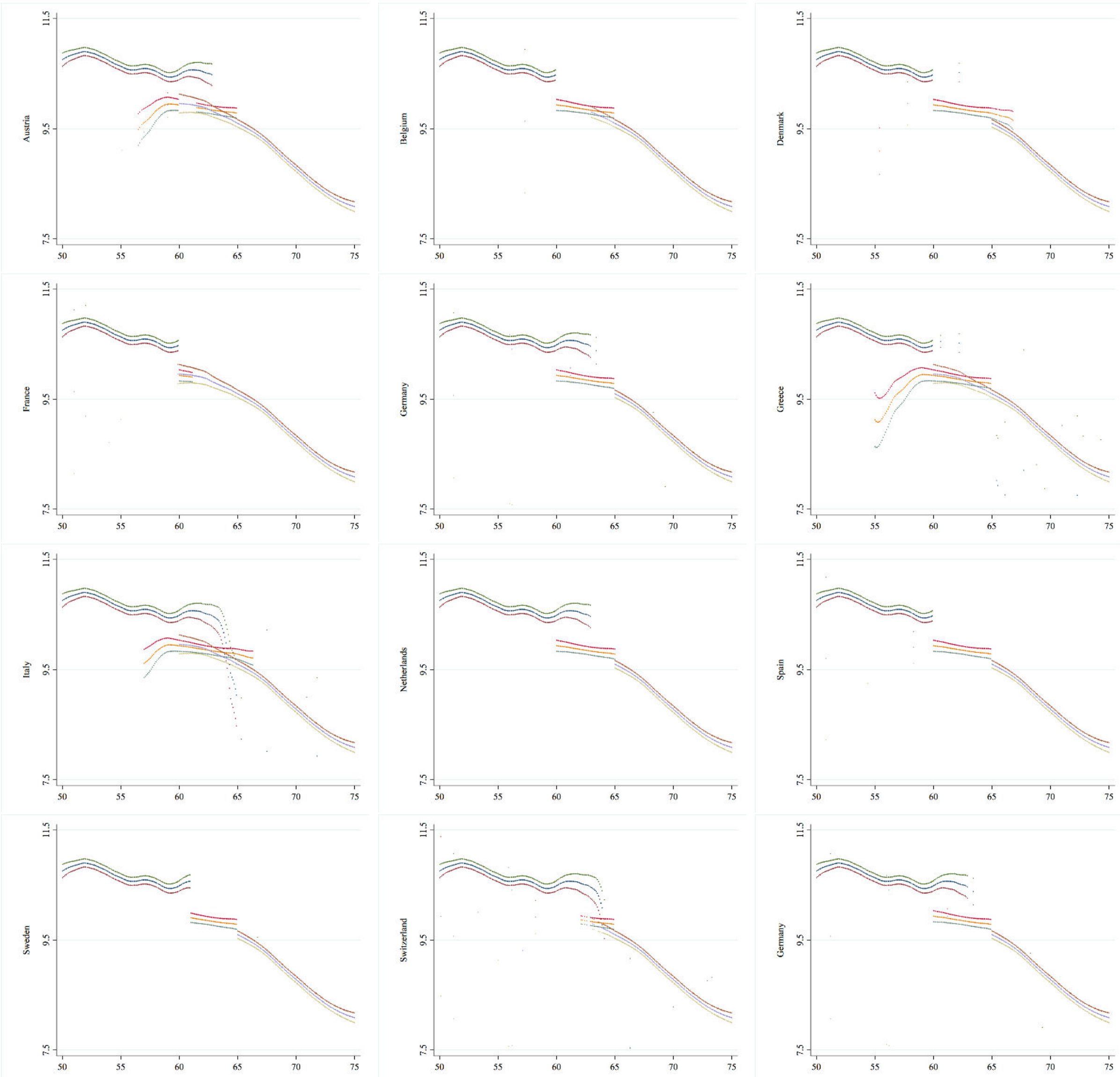


Figure 7: Word recall score by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

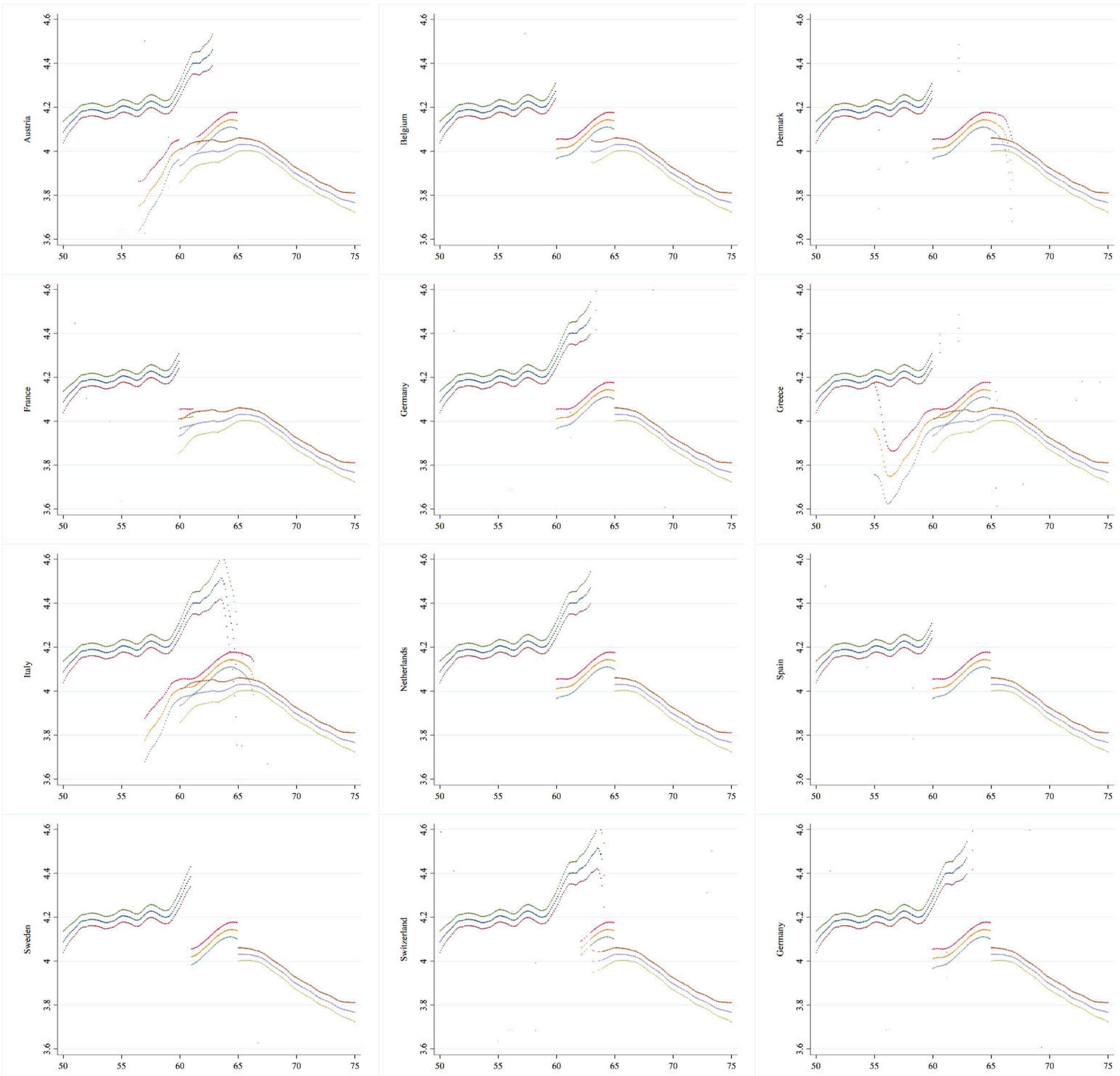

Figure 8: Numeracy score by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.

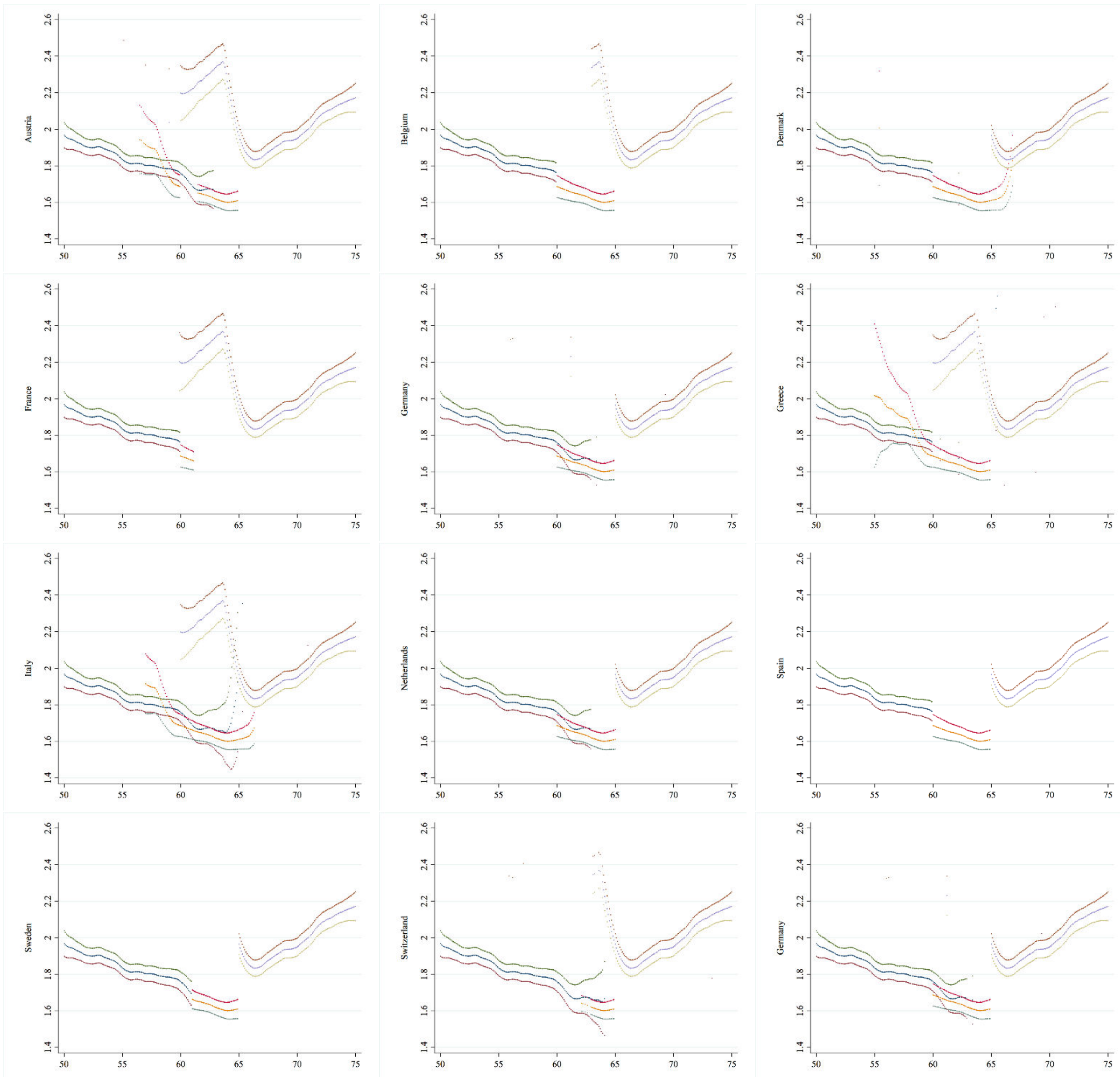

Figure 9: Depression score by age of the respondent. Kernel smoothed local polynomials and 95 percent confidence intervals around them.