UNIVERSITÀ POLITECNICA DELLE MARCHE

DOTTORATO DI RICERCA IN ECONOMICS

XXXV CICLO

Timing of labor market entries and exits: career dynamics and health outcomes

Ph.D. Candidate:
Mattia Filomena
S1092184

Supervisor:
Prof. Matteo Picchio

Ph.D. Coordinator:
Prof. Marco Cucculelli

Contents

List of Tables

List of Figures

OVERVIEW

Labor market entries and exits in Italy are of utmost importance from both a socio-economic point of view and from a policy perspective. On the one hand, the persisting high unemployment in Italy is particularly penalizing for new labor market entrants, who face significant difficulties in Italy, where the average duration of the school-to-work transition (2.88 years for those aged 18-34) is the highest in Europe, discouraging young people from investing in tertiary education (Pastore et al., 2020, 2021). Together with the immediate loss of income and the lack of accumulation of human capital, early career nonemployment may also have long-term negative consequences in terms of labor earnings and labor market participation (Gregg and Tominey, 2005; Mroz and Savage, 2006). On the other hand, Italy is seeing its population ageing, like many developed countries. For this reason, increasing pressures are thus being generated on the financial sustainability of the pension system and are faced with policies that have gradually delayed the retirement age, increased the required contributory period, and changed the pension calculation schemes (Carone et al., 2016). These institutional changes lead researchers to investigate the health consequences for workers who will have to stay on the labor market longer than they had anticipated when they were younger. Although in the last 20 years the scientific literature in the health economics field has seen the publication of various works that have studied the impact of retirement on income, consumption, leisure activities, domestic activities, and physical and mental health (see e.g. Fé and Hollingsworth, 2016; Nishimura et al., 2018), the gradients of heterogeneity among workers and how important the timing of retirement is still unclear. Thus, both these issues become the object of investigation of this research project.

This thesis is divided into two parts, which are both composed by two chapters: i) labor market entries, focusing on unemployment and subsequent labor market outcomes; ii) labor market exits, with a particular interest on the health effects of retirement and its timing. Both parts include extensive review of the related literature, which provide an overview of empirical studies for both unemployment scarring effects and health impact of retirement, respectively. Furthermore, I performed a battery of meta-regression techniques aimed at estimating the average precision effects after correcting for publication bias and the heterogeneities related to several study characteristics. The two empirical analyses in this thesis made use of the AD-SILC data. These are Italian data obtained from the combination of the database produced by the Survey on Income and Living

Conditions in Italy (IT-SILC), made available by ISTAT, with the administrative data on employment contracts, provided by INPS. In both cases, the econometric models allow the identification of the causal effects of youth nonemployment by taking into account a series of individual and time-varying unobserved factors related to personal characteristics and the socio-economic context; and the causal impact of retirement exploiting the exogenous shock of a pension reform which aims at increase the normal retirement age.

The structure of the thesis is as follows. The first chapter reviews the empirical literature on unemployment scarring effects, that is the negative effect of past experiences of unemployment on subsequent labor market outcomes, such as lower probabilities to be re-employed, or higher earning losses after re-employment (Arulampalam et al., 2001). More in detail, it presents an overview of empirical studies that applied causal inference techniques and focused on different causes of previous unemployment (job displacement, youth unemployment after school completion, plant closure..). This exercise reveals both wage penalties following unemployment spells and state dependence in unemployment persistence as a common conclusion in the literature, although little differences across empirical findings relate to the magnitude of the scarring effects. To shed light on the heterogeneity dimensions under different study features, I employed model averaging strategies in meta-regression analysis (Stanley, 2005, 2008; Magnus et al., 2010) and estimated the expected partial correlation coefficients for different combinations of these study-related characteristics. Main results show that unemployment scarring effects are particularly penalizing for laid-off workers, and the negative impact is greater for men and in the short-term.

Although the issue of unemployment scarring effects is crucial, the Italian labor market has not received much attention so far on this topic. Thus, the second chapter aims at investigating the presence of scarring effects in Italy, focusing on the impact of nonemployment episodes experienced during the first 3 years after high school diploma on subsequent yearly labor earnings and participation in employment in short- and in long-term. From the methodological point of view, I employed a factor analytic model (Carneiro et al., 2003; Heckman and Navarro, 2007) which allow to take into account time-varying unobserved heterogeneity jointly affecting selection into nonemployment after diploma and subsequent labor market outcomes later in life. Once unobservables characteristics are accounted for, I obtain evidence that school-leavers in Italy who experienced nonemployment during the first 3 years after attained high school diploma suffer from relevant scarring effects. The negative effects are very persistent in terms of earnings: they are still sizable and statistically significant 25 years after school completion. Labor market participation, measured as the fraction of days spent at work in a year, is negatively affected by early nonemployment for a shorter span, as it disappears for both men and women by the 10th year after the school completion. When I control for time-varying unobserved heterogeneity, the negative effect of early nonemployment on labor earnings becomes smaller in magnitude, whereas the penalties in terms of labor participation are present only up to 5 years after school completion. This suggests that the inclusion in the model of the time-varying latent factor allow to capture those latent traits which affect both selection into early nonemployment and future labor market performances.

The second main focus on the Italian labor market concerns the health consequences of labor market exits through retirement options. For this reason, the third chapter approaches this topic by collecting published articles in peer-reviewed journals which focused on the impact of retirement on several measures of health: physical and mental health, self-assessed general conditions, healthcare utilization, and mortality. I exploited meta-regression analysis techniques to check for the presence of publication bias and a genuine effect of retirement on health, under the assumption of a common true effect (Stanley, 2005, 2008; Stanley and Doucouliagos, 2012, 2014). Furthermore, through the use of model averaging techniques (Magnus et al., 2010), the chapter explores possible sources of heterogeneity of the true effect, taking into account all the main factors that might lead to different findings, such as the institutional context, the causal effect identification strategy, the type of previous occupation, and further several study-related characteristics. Main findings reveal a positive but extremely small average effect of retirement on health, whereas different reported estimates are linked to the differences in health outcomes used by researchers and in the type of retirement scheme.

The fourth and last chapter exploits the 1992 pension reform in Italy to evaluate the causal effect of retirement and its timing on mortality on a sample of private employees born between 1929 and 1944. The identification of the causal health effects of retirement involves methodological issues that are not easy to deal with (Kuhn, 2018). This chapter add to the debate an empirical innovation by adopting a factor analytic model with dynamic selection into treatment to evaluate the causal impact of retirement and its timing on mortality, in which workers differ in unobserved characteristics jointly affecting selection into retirement and subsequent health outcomes. These unobserved traits, such as labor force attachment, liquidity constraints, different health problems or behaviors, may affect the retirement decision and subsequent health outcomes and make difficult the identification of the causal effect of retirement. Through a factor-analytic dynamic model (Carneiro et al., 2003; Heckman and Navarro, 2007), I achieved the nonparametric identification of the treatment effect while taking into account selection on the time-varying unobservables by the factor structure with a latent trait and time-varying factor loadings. In this framework, the 1992 pension reform works as a further exclusion restriction in the treatment equation. In line with previous empirical literature, results suggest that, on average, retirement does not affect mortality, with the exception of a positive effect of postponed retirement for men on the probability to be alive at 78 years old.

Chapter 1

Unemployment scarring effects: an overview and meta-analysis of empirical studies

1.1 Introduction

Since the 1980s, many labor economists have focused their research activity on studying the impact of past unemployment on subsequent labor market outcomes, such as earnings and employability, and whether these effects are temporary or permanent. The detrimental effects of unemployment are confirmed in more recent studies: a past history of unemployment tends to increase the likelihood of experiencing future unemployment and generate earning losses after re-employment, inflicting a permanent "scar" (Arulampalam et al., 2001). The literature on scarring effects is very large, and researchers have approached it from different angles: for instance, substantial literature looks at the impact of job displacement on future labor market outcomes, while other studies focus on how unemployment experience affects school-to-work transitions.[1] Literature reviews on the unemployment scarring effects on subsequent labor market outcomes date back to the 1990s or even earlier: Hamermesh (1989) collected 12 studies on US worker layoffs; Fallick (1996) reported on the effects of worker displacement distinguishing in employment,

[1] Moreover, further outcomes discussed by the literature on scarring are family formation, crime and negative psychological implications in terms of well-being, life satisfaction, and mental health (see e.g. Helbling and Sacchi, 2014; Strandh et al., 2014; Mousteri et al., 2018; Clark and Lepinteur, 2019).

earnings, and human capital outcomes; while Kletzer (1998) extensively discussed the state of knowledge on the issues of job displacement. Recent reviews are also provided by Baumann (2016), who discussed the consequences of job displacement on displaced workers' occupational situation, sociability, and well-being, and by Borland (2020), who mainly focused on Australian studies.[2]

The present article collects a large number of papers studying these phenomena, including both single-country and multi-country analyses. We contribute to the literature filling the absence of a systematic review of studies that apply causal inference to identify the causal effects of previous unemployment episodes and through a meta-analysis. In doing so, we retrieved point estimates from each study and performed a meta-regression analysis to highlight the magnitude of the scarring effects after investigating the issue of publication bias. Moreover, we took into account the main factors that might be sources of different effect sizes among studies, such as identification strategies, geographical areas, different causes of previous unemployment experiences, and other study-related characteristics.

Although the analysis of the scarring effects of unemployment is not at the frontier of the research in labor economics, it is of utmost importance to provide evidence on the magnitude and duration over time of the unemployment scarring effects for both socioeconomic and policy reasons. Firstly, the economic crises of the Great Recession and the Covid pandemic should spark a renewed interest in understanding and avoiding the negative consequences of experiencing unemployment. Secondly, Adascalitei and Morano (2016) counted at least 642 changes in labor market institutions which aim at reducing the existing level of regulation and therefore they may facilitate job dismissals. Thirdly, recent studies documented longer school-to-work transition periods, in particular in Southern Europe (Pastore et al., 2021). This may lead to the lack of accumulation of human capital and skills, to less chance of generating a network, and therefore to negative effects on subsequent labor market outcomes. For these reasons, we reviewed the literature on unemployment scarring effects to provide policy makers further suggestions on how to avoid persistent scars. Knowledge about the scarring effect of unemployment on

[2] A further strand of the recent literature focuses on the effect of adverse labor market conditions at graduation, for example focusing on the effect of local unemployment rate or graduating during a recession (see e.g. Raaum and Røed, 2006; Kahn, 2010; Oreopoulos et al., 2012; Kawaguchi and Murao, 2014; Altonji et al., 2016). The consequences of economic downturns on wages, labor supply and social outcomes for young labor market entrants have been recently surveyed by Cockx (2016), Von Wachter (2020), and Rodríguez et al. (2020).

future outcomes is important from a societal perspective, as it informs whether the social cost of unemployment extends beyond the period in which it is experienced. Therefore, this review aims to offer scholars and policy makers a consistent collection of empirical evidence relating to the scarring effects of previous unemployment on later working career, focusing on job displacement, plant closure, early unemployment episodes after graduation and more general causes of individually experienced unemployment. In this way, policy makers could have a solid body of proof on the magnitude of unemployment scarring effects on subsequent labor market status, wage penalties, and job stability, and use these results as a support to the economic policy responses that aim at preventing long-term unemployment and avoiding such consequences on living and working conditions.

The remainder of the article is organized as follows: Section 1.2 defines the theoretical background relative to the unemployment scarring effects. Section 1.3 presents the search strategy and summarizes both methodological issues and causal effect identification strategies mainly adopted in the empirical literature. Section 1.4 describes the magnitude of the scarring effect in the literature through the use of meta-analysis techniques and focuses on the heterogeneity of the results of the empirical evidence according to several study-related characteristics. Section 1.5 draws some conclusions.

1.2 Theoretical background: mechanisms of scarring

Following Gregg (2001) we can summarize at least three reasons that explain the possible association between previous unemployment and future labor market persistence and scarring. Firstly, some people may be more inclined than others to worse job careers due to persistent differences in unobservable characteristics (e.g. ability and motivation; differences in the search intensity or in the methods of search; different liquidity constraints and, as consequences, different reservation wages). Secondly, a young worker may become unemployed due to persistent labor market conditions. Thirdly, the experience of past unemployment can generate further unemployment in the future, i.e. the “true state dependence”, which is what the scarring effects literature is interested in. Moreover, the duration of unemployment can affect labor market outcomes directly and indirectly. The direct effect is through negative duration dependence in the transition from unemployment to employment or through its lagged effect on the starting wage and on the subsequent employment stability. The indirect effect is through the employment experience that is foregone, influencing thereby both the duration of subsequent employment (or unemployment) spells and wages in subsequent employment periods (Cockx and Picchio, 2013).

But what are the causes of these scarring effects? Theoretical explanations for the presence of the labor market scars are laid down in three main theories: the human capital theory, the signaling theory, and the job search theory. According to the human capital theory, as long as workers accumulate firm-specific skills, their productivity increases and so earn more (Becker, 1975; Mincer, 1974; Pissarides, 1992). In the human capital theory, employment and wage scars are related to i) the depreciation, following an unemployment spell, of general skills and knowledge that workers possess; ii) the lack of accumulation of human capital that occurs if an individual faces early unemployment spells. In particular, when the contract between workers and the firm is terminated, workers are likely to lose their specific human capital, be less productive in their subsequent jobs, and to obtain lower subsequent wages than if they did not experience unemployment.

A second explanation derives from the signaling theory or imperfect information theory. It suggests that, since productivity is not easily observable at the time of hiring, the employer uses past history of unemployment of a worker, such as the number of unemployment spells, their duration or frequency, as a signal of low productivity. Workers are therefore penalized, at least initially, by lower employment probabilities and subsequent wages (Spence, 1973; Vishwanath, 1989; Lockwood, 1991). However, this penalty at the time of re-employment should vanish over time if the worker shows greater productivity than expected from employer. In presence of asymmetric information, employers observe also the type of separation from previous job: e.g. plant closures give a less negative signal about productivity compared to layoffs, so the "stigma" effect[3] and the consequent wage loss should be lower. For instance, Gibbons and Katz (1991) found that workers who were laid off experienced a short-term wage loss that was 5.5 percentage points greater than that of workers who were displaced by plant closures. Additionally, laid-off workers had post-displacement unemployment spells that were about 25% longer.

An attempt to disentangle the effects of stigma, human capital decay and heterogeneity across the earlier literature on unemployment scarring was provided by Omori (1997), who found that one month more in the duration of past episodes of nonemployment lengthens the expected duration of future nonemployment by 0.39 months on average.

[3]The stigma effect means that individuals who have been unemployed face lower chances of being hired because employers may use their past history of unemployment as a negative signal.

Moreover, this effect is larger the lower the local unemployment rate was at the time of past unemployment episodes, confirming the stigmatization effect on US workers. Further examples of papers that highlight the signaling as the main mechanisms at work are Biewen and Steffes (2010) and Tanzi (2022).

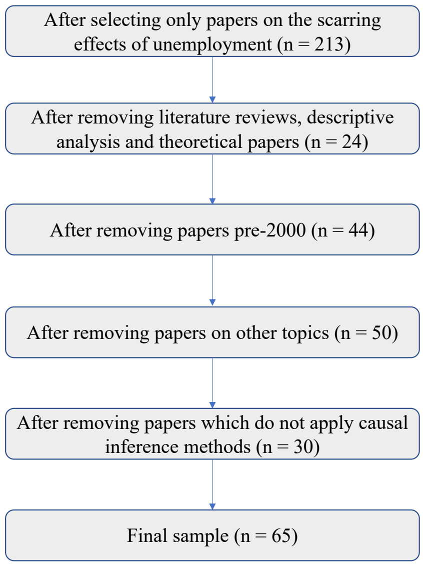
However, another important role is played by the job-match, according to the job search theory. When a bad match is terminated, future earnings will be higher if the subsequent unemployment spell allows the worker to get a better match with future employer. The job-search model predicts a positive effect of job mobility on subsequent earnings because workers are assumed to continue searching for more efficient job matches (Burdett, 1978; Jovanovic, 1979a; Mortensen, 1987, 1988). Moreover, workers could leave jobs if they do not experience improvements in productivity with seniority. If it is true, a stable matching over time will be considered as a signal of high productivity and then a highly profitable job (Jovanovic, 1979b). Finally, Lazear (1986) suggested that job movers are high-skilled workers and firms, competing for this type of employees, offer them higher wages.

1.3 Data and empirical framework

1.3.1 Selection criteria and study features

Following a set of standards in summarizing the literature suggested by the Meta-Analysis of Economics Research Network (MAER-Net) guidelines (Havránek et al., 2020), we carried out our literature search through a comprehensive search in Web of Science and Google Scholar databases and focused only on articles in English, for the sake of accessibility (Vooren et al., 2019). Our search strategy was performed up to December 2021 using combinations of the following keywords: “unemployment scarring effects”, “wage losses”, “duration dependence”, “employment probability”, “labor earnings”, “unemployment spell”. In order to review empirical studies that control for endogeneity and estimate the causal effect of unemployment by using more robust econometric strategies, we apply some inclusion criteria. First, since much of the earlier literature suffers from selection bias, we focused on more recent articles that offer more credible ways to deal with this issue. We applied filters to take only articles published during the period 2000-2021 in refereed journals, working papers, technical reports and contributions to books. Second, we excluded studies that did not use a rigorous methodological approach.[4] Third, we included only articles dealing with the causal effect of individual unemployment spells on subsequent wage and employment opportunities. For this purpose, we excluded articles focusing on other topics, such as intergenerational scars, macroeconomic conditions at graduation, psychological implications, or well-being analysis.[5] Figure 1.1 graphically reports the rules we followed to include/exclude articles in our final sample. The final selection included 65 papers, which are listed by outcomes in Tables A1.1 (subsequent employment) and A1.2 (labor earnings) in the final Appendix, briefly reporting their main characteristics.

Figure 1.1: Flow chart describing selection criteria



Notes: Starting with a sample of 213 papers, from the second to the fifth block n represents the number of excluded studies at each step.

The empirical literature on the scarring effects of unemployment covers lots of coun-

[4]Thus, papers using traditional multivariate descriptive analysis, duration models, or OLS regressions with a reduced number of controls which do not properly assess endogeneity issues and are unlikely to reveal causal interpretation (endogeneity issues are discussed in Subsection 1.3.2).

[5]For intergenerational scars we mean that studies focused on the effect of parents' unemployment experiences on the children' future employment status (see e.g. Karhula et al., 2017). For macroeconomic conditions at graduation we mean that we exclude that literature focused on the local unemployment rate at graduation or other local labor market conditions, rather than on individual unemployment experience and state dependence (see e.g. Oreopoulos et al., 2012; Raaum and Røed, 2006).

tries and uses several databases. Most of the US studies used databases such as the Displaced Worker Survey (DWS), which is related to the Current Population Survey (CPS), followed by the Panel Study of Income Dynamics (PSID), and the National Longitudinal Survey (NLS). As for Europe, studies about the British labor market used the British Household Panel Survey (BHPS), the National Child Development Survey (NCDS), or the Joint Unemployment and Vacancies Operating System (JUVOS). Papers concerning Germany used the German Socio-Economic Panel (GSOEP), while the 12 studies about Scandinavian countries mainly exploited administrative register datasets. Several databases are used in studies concerning Belgium: Panel Study on Belgian Households (PSBH), Crossroads Bank for Social Security (CBSS), VDAB, and SONAR Survey Database. This empirical literature not only concerns studies conducted on single countries, but also comparative analysis between two or more countries, in particular within the European Union and using the EU-SILC database or the European Community Household Panel (ECHP). Five papers compared several countries, and Gangl (2006) included also the US in the analysis using the Survey of Income and Program Participation (SIPP) for USA and the ECHP for 12 European countries.

1.3.2 Methodological approaches

From a methodological point of view, the analysis conducted about unemployment scarring effects usually have a large number of control variables. These include individual and demographic characteristics (e.g. age, gender, nationality, marital status), human capital indicators (education, experience, tenure), unemployment indicators (e.g. duration or number of unemployment spells), job characteristics (type of contract, number of working hours, sector, firm size, union membership) and macro measures to check for business cycle variations and differences in the state of the local labor market (GDP growth, annual rate of unemployment).

However, there are some econometric issues to take into account. First, in estimating wage losses one requires individuals to be found in employment with non-missing wage information, otherwise this might cause sample selection bias. Thus, many studies adopt the procedure proposed by Heckman (1979), that is including the Heckman correction term as a regressor in the wage equation. Second, the individual fixed characteristics may drive the unemployment scarring effects. Therefore, it is important to separate true state dependence from their spurious effect induced by the correlation with unobserved individual propensities to remain unemployed, to avoid biased estimates due to reverse causality or measurement errors. These unobservable traits may jointly determine both selection into treatment (unemployment) and future labor market outcomes: labor force attachment, motivation, ability, self-confidence, job search behavior, liquidity constraints, and family or cultural background are indeed very likely to affect labor market performances, but they are not observed by the analyst in most cases. As a consequence, the relationship between previous unemployment and subsequent labor market outcomes may be not causal but reflect this kind of unobserved heterogeneity, which may be both time-constant or time-varying. Thus, in cases of biased results, policies aimed at reducing the incidence or duration of unemployment spell might be misdirected.

Among the collected studies, some articles adopted a field experiment approach (for instance, by randomly assigning fictitious resumes to real job postings). Randomization guarantees that individuals belonging to the treated and counterfactual groups are equal with respect to all observed and unobserved characteristics except for treatment reception. However, randomization of treatment is often unfeasible in labor market studies because most of individuals, either employed or non-employed, cannot be forced to receive the treatment of the RCT. The decision to participate or not may be correlated to the benefits of the treatment, meaning that self-selection into treatment occurs, and selection bias still arises when the treatment variable is correlated with the error in the outcome equation. This correlation could be induced by incorrectly omitted observable variables ("selection on observables") or by unobserved factors ("selection on unobservables").

The problem in the former is solved using regression and matching methods. Two studies adopted the "control function estimator" which is motivated by the possibility that a set of observables determining the treatment variable may be correlated with the outcome, under the assumption that conditioning linearly on observed covariates is adequate to remove selection bias. These studies are Heylen (2011), who looked at the effects of an unemployment episode at the beginning of the career in Belgium, and Gartell (2009) who focused on the college-to-work transition in Sweden. Differently, 7 studies adopted the "Propensity Score Matching" (PSM) method (see e.g. Nilsen and Reiso, 2014; Helbling and Sacchi, 2014; Abebe and Hyggen, 2019), which handles the selection problem by non-parametric techniques and some underlying assumptions, such as no systematic differences between the two groups in unobserved characteristics that influence the outcome after matching. These methods are based on selection on the observables but tell nothing about selection on unobservables.

In the selection on unobservables methods the issue is addressed using instrumental variables (IV), diff-in-diffs estimators (DiD), the timing of events approach, or other tools such as dynamic panel fixed or random effects methods. The literature investigating the causal effect of unemployment on labor market outcomes has made use of a variety of these methodologies to overcome selection bias and endogeneity problems. The most commonly employed strategy among the collected papers is the within-group estimation in fixed-effects panel regression (17 studies), in particular to estimate the scarring effects on wages. Thus, even recent papers relied on the diff-in-diffs approach proposed by Jacobson et al. (1993), at least concerning job displacement and wage equation, by comparing the changes in outcomes over time between treated and untreated units (e.g. between workers who experienced a job loss and subsequent unemployment and a control group of continuously employed workers). In contrast, the dynamic random-effects probit models (DREP) were mainly used to evaluate the unemployment state dependence. In these models, unlike in linear ones, the unobserved heterogeneity is treated as randomly distributed in the population and any bias in the estimated parameters is allowed by parametric approximations (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005). Further articles made use of IV estimators, in particular within the strand of the literature which looks at the scarring effect of youth unemployment. The main instruments used are the local unemployment rates at age 16 (Gregg, 2001; Gregg and Tominey, 2005), or before graduation (Ghirelli, 2015; Schmillen and Umkehrer, 2017; Tanzi, 2022), while Möller and Umkehrer (2015) instrumented early-career unemployment with the event of a plant closure of the training firm, taking place in the year of graduation. The use of these instruments is based on the idea that the variation in the labor market conditions at such a young age or at school leaving is exogenous since individuals do not choose the area in which they live or the time to graduate. Therefore, this variation in an individual's early unemployment is unrelated to unobserved characteristics that could influence both early and adult labor market performances.

1.4 Meta-analysis

In what follows we summarized the empirical evidence about the magnitudes of the scarring effects through both a graphical approach and a meta-regression analysis. For each study included in our survey we retrieved the t-statistic (effect size) of the relationship between past unemployment and future labor market outcomes.[6] Multiple point estimates were delivered if, for instance, the analysis is disaggregated by gender, by incidence, duration or number of previous unemployment spells, time horizons analyzed, or if multiple labor market outcomes were tested. However, we only retrieved the estimates related to the effect of occurrence, duration or incidence of previous unemployment, or the number of unemployment spells. Thus, we excluded estimates based on interactions between unemployment and other features, such as discouragement (Ayllón, 2013; Ayllón et al., 2021) or stigma (Ayllón, 2013; Ayllón et al., 2021; Biewen and Steffes, 2010). We obtained a final meta-sample of 616 observations.[7] Although the empirical literature provides homogeneous evidence in support of the unemployment scarring effects, that is the effect of previous unemployment experiences on subsequent labor market success is negative, in what follows we highlighted some differences in terms of the magnitude of these penalties. The average t-statistic after distinguishing between point estimates focused on the outcomes of earnings and subsequent employment[8] is -6.48 and -9.30, respectively.

However, if we used the t-statistics as a measure of the relation between previous unemployment and labor market outcomes we would lose information about the size of the link between them. Thus, we computed the partial correlation coefficient r_i, which has been commonly used in meta-analyses in economics, business and social sciences since Doucouliagos (1995). It is a measure that allows us to keep a quantification of the strength of the statistical association between two variables and which is independent of the metrics of the dependent and independent variables (Ugur, 2014). The r_i is computed as follows:

$$r_i = \frac{t_i}{\sqrt{t_i^2 + dk_i}}, \tag{1.1}$$

where dk_i are the degrees of freedom in the model from which each effect size is derived.[9]

[6] When we could not directly retrieve the t-statistics because not reported among the study results, we computed them as the ratio between the estimated unemployment effects (β_i) and their standard errors. If studies only displayed the estimated effects and their 95% confidence intervals, the standard error can be calculated by $SE_i = (ub - lb)/(2 \times 1.96)$, where ub and lb are the upper bound and the lower bound, respectively

[7] We removed from the meta-regression analysis 8 articles because they did not contain sufficient information to compute the t-statistic of the estimated scarring effect. They are simply discussed in Tables A1.1 and A1.2 and are reported in italics.

[8] For employment outcomes we mean the likelihood of experiencing future unemployment, the probability to have a job later (employability), the fraction of days spent at work or the hours worked during the following years (labor market participation), the call-backs from employers in case of field experiment. Earning outcomes include hourly wages, labor earnings, income, etc.

[9] Since many studies did not provide precise information on the number of covariates, we approximated

Its standard error is given by

$$SE(r_i) = \sqrt{\frac{1 - r_i^2}{dk_i}}. \tag{1.2}$$

The partial correlation coefficient is a unitless measure, which takes a value between -1 and 1 and enables direct comparisons among the different ways to approach and measure outcomes in the empirical literature. This measure drops as the degrees of freedom or the sample size increase and, therefore, nearly similar t-statistics will produce very different partial correlations if the sample sizes are too diverse. Table 1.1 shows preliminary descriptive statistics by distinguishing between the two labor outcome variables and across different identification strategies. The overall mean of the r_i values is -0.029 when the outcome is labor earnings, and -0.055 in case of employment status as dependent variable.

Table 1.1: Summary statistics: average effect sizes

	Labor earnings			Employment		
	Average r	t-statistic	Observations	Average r	t-statistic	Observations
a) Overall sample	-0.029	-6.480	352	-0.055	-9.302	264
b) By identification strategy						
Field Experiment	–	–	–	-0.021	-0.986	66
Selection on observables	-0.070	-8.300	32	-0.064	-5.419	19
Selection on unobservables	-0.025	-6.298	320	-0.067	-12.781	179

Notes: Selection on observables include the control function estimator and the propensity score matching; Selection on unobservables includes instrumental variables; diff-in-diffs and within group estimation in panel fixed effects; dynamic random effects probit models; and other methods (Timing of Events, Discrete Factor Maximum Likelihood, exclusion restrictions).

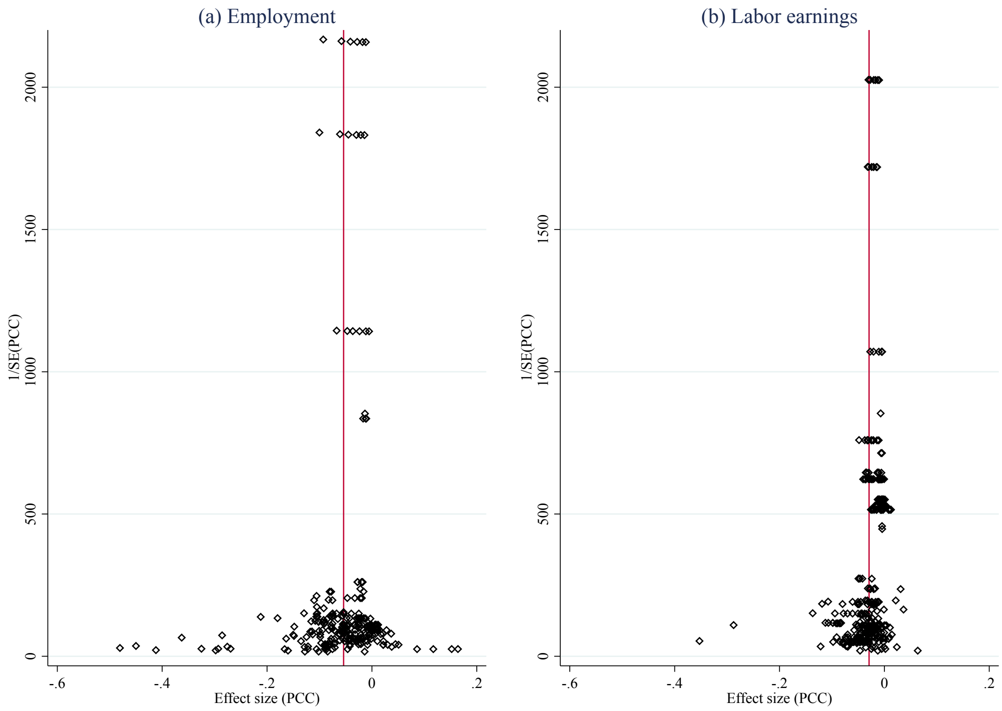
However, the simple overall mean effect should be interpreted with caution due to the possibility of publication bias which may affect the reported estimates in this strand of research.[10] To check whether this may be an issue, we first show the funnel plot in Figure 1.2. It displays the relationship between the effect size r_i and its precision, measured by the inverse of its standard error.

From the two graphs in Figure 1.2, we only note a mild asymmetry on the left tail when the labor market outcome concerns employment. We can conclude from this preliminary visual inspection that there is no publication bias because the effect size varies

dk_i with the number of observations minus 2. Indeed, given that in microeconometric applications the sample sizes are very often much larger than the number of the parameters, the calculation of the partial correlation coefficient is quite robust to errors in deriving dk_i (Picchio, 2022).

[10]The publication bias is the bias arising from the tendency of editors to publish more easily findings consistent with a conventional view or with statistically significant results, whereas studies that find small or no significant effects tend to remain unpublished (Card and Krueger, 1995).

Figure 1.2: Funnel plot of effect size (r_i) versus its precision ($1/SE(r)$)



Notes: The number of observations is 616 (264 for employment and 352 for labor earnings). The vertical lines are the average of the partial correlation coefficients for a) employment (r = -0.055) and for b) labor earnings (r = -0.029).

randomly around its mean, which corresponds to an authentic empirical effect. Thus, we formally test for the presence of publication bias by estimating the Funnel Asymmetry Test-Precision Effect Test (FAT-PET) model (Stanley, 2005, 2008). It is a linear model where the effect size is regressed on a constant term and its standard error:

$$r_i = \delta_0 + \delta_1 SE(r_i) + \varepsilon_i, \tag{1.3}$$

First, the Funnel Asymmetry Test (FAT) tests the hypothesis of no publication bias, i.e. $H_0 : \delta_1 = 0$. Second, the Precision Effect Test (PET) tests $H_0 : \delta_0 = 0$, where the rejection of the null hypothesis can be interpreted as the presence of an authentic empirical effect of past unemployment, corrected for publication selection. Table 1.2 reports the results of meta-regression analysis separated by the two labor market outcomes using the

Fixed Effects Weighted Least Squares (FE-WLS) model with $1/(SE_i)^2$ as weights to take into account heteroskedasticity affecting the distribution of r_i. We do not find evidence of publication bias, so we repeat our meta-regression analysis by performing the PET test only. Our results show that the precision effect of previous unemployment on labor earnings and subsequent employment is about -0.018 and -0.041, respectively, under the assumption of a common true effect.

Table 1.2: Meta-regression analysis (MRA)

	Labor earnings (N = 352)		Employment (N = 264)	
	WLS-FE (FAT-PET)	WLS-FE (PET)	WLS-FE (FAT-PET)	WLS-FE (PET)
a) MRA				
Precision effect	-0.016*	-0.018**	-0.040***	-0.041***
	(0.079)	(0.021)	(0.000)	(0.001)
Publication bias	-0.909		-0.651	
	(0.455)		(0.565)	
R^2	0.011	0.000	0.002	0.000
b) MRA by estimation strategy				
Precision effect (Field Experiment)	–	–	0.006	-0.011
			(0.653)	(0.146)
Precision effect (Selection on observables)	-0.051	-0.066**	-0.036**	-0.047
	(0.280)	(0.050)	(0.034)	(0.179)
Precision effect (Selection on unobservables)	-0.017*	-0.017**	-0.039***	-0.041***
	(0.051)	(0.020)	(0.001)	(0.003)
Publication bias (Field Experiment)	–		-1.410	
			(0.159)	
Publication bias (Selection on observables)	-2.113*		-1.693	
	(0.054)		(0.166)	
Publication bias (Selection on unobservables)	-0.082		-1.691	
	(0.950)		(0.269)	
R^2	0.680	0.679	0.669	0.666

Notes: We report wild clustered bootstrap p-values obtained from the wild clustered bootstrap-t procedure proposed by Cameron et al. (2008a), with clusters at study level (5,000 bootstraps using the Webb's (2014) six-point distribution as weights). *** Significant at 1%, ** significant at 5%, * significant at 10%.

These regression-based methods can be adjusted to deal with potential different tendencies to p-hack and selective reporting results across different features. For instance, Brodeur et al. (2020) found that p-hacking and publication bias tendencies in economics varies greatly by the estimation method of the causal effect. Brodeur et al. (2016) showed that p-hacking is less likely in studies using RCT as study design. Since in our case preliminary evidence suggests the possibility of significant differences in the magnitude of the effects between different methodologies, we present in panel (b) of Table 1.2 the estimates of FAT-PET and PET models dividing the sample by identification strategy. There is very weak evidence of negative publication bias at 10% for studies using selection on observables, but only when the outcome concerns labor earnings. In summary, the precision effect of unemployment on earnings is larger using selection on observables methods ($\delta_0 = -0.066$), whereas the precision effect on future employment is similar in magnitude between selection on observables and unobservables (between -0.036 and -0.041). In contrast, the precision effect is not statistically different from zero when the estimation strategy relies on field experiments.

As second step, in order to address the effect of heterogeneity we retrieved the research dimensions that may be relevant and include them into Equation 1.3 to perform a multiple meta-regression analysis. Firstly, we coded the following auxiliary regressors: i) methodology (field experiment, selection on observables and selection on unobservables); ii) type of data (survey vs. administrative data); iii) cause of unemployment episode (youth unemployment, job displacement and plant closure, experiences of nonemployment during recessions and unemployment episodes for which the reason is not clearly specified); iv) country (Anglo-Saxon, Scandinavian, European countries and multi-country analysis); v) sex (males, females, and both); vi) measure of previous unemployment (unemployment status or occurrence vs. duration of previous unemployment experience or number of unemployment spells); vii) time-horizon of the outcome (short-term vs. medium-long term, i.e. 5 years later or more); viii) study-quality measures, such as year of publication and a dummy equal to 1 for articles published in peer-reviewed journals. Thus, we estimated by FE-WLS the following equation:

$$r_i = \delta_0 + \delta_1 SE(r_i)^2 + \boldsymbol{\gamma}_1 \mathbf{x}_i + \varepsilon_i, \tag{1.4}$$

where $\mathbf{x}_i$ is a vector of auxiliary variables containing all the study characteristics and γ is the $1 \times k$ vector of parameters.[11]

To address model uncertainty, we used Bayesian Model Averaging (BMA, Magnus et al. (2010)) for model selection. BMA is a standard way in meta-analysis of selecting the best model by considering all possible models, by estimating them with different subset of potential explanatory variables, and by computing the weighted averages of the estimated coefficients. It provides the Posterior Inclusion Probability (PIP) for each regressor, where a PIP above 0.5 is usually used as a rule of thumb to include the auxiliary variable into the final model (Eicher et al., 2011). For each covariate, BMA returns the posterior coefficient distribution, which yields the posterior mean (PM) of the regression coefficient and the posterior standard deviation (PSD). Since we had 14 auxiliary covari-

[11] We employed the Precision Effect Estimate with Standard Error (PEESE) specification because its quadratic form of the standard errors has been proven to be less biased and often more efficient to check for heterogeneity than the FAT-PET specification when there is a nonzero genuine effect (Stanley and Doucouliagos, 2014).

ates (i.e. those variables that are suspected to be relevant in explaining heterogeneity), BMA required the estimation of 2^{14} models.

Furthermore, we followed Magnus et al. (2010) and used a further model-average procedure: the Weighted Average Least Squares. WALS is a Bayesian combination of frequentist estimators which has the intermediate position between Bayesian and frequentist model-average estimators. The advantages of WALS over BMA are that i) it does not impose an *ad hoc* assumption on the prior on the model space (in general BMA uses a uniform prior assigning equal probability to each model), but it is theoretically based (Magnus and De Luca, 2016); ii) it relies on preliminary orthogonal transformations of the auxiliary variables and their parameters, reducing the computational burden from, in our case, 2^{14} to 14 models. An auxiliary variable is considered to be correlated with the outcome if the t-ratio of its coefficient is greater than 1 in absolute value (De Luca and Magnus, 2011). Finally, following the meta-analysis literature (see e.g. Havranek et al., 2015; Xue et al., 2021; Picchio, 2022), we provided OLS estimates as a frequentist check using those variables that are relevant according to BMA results, i.e. by restricting the set of regressors to those with PIP > 0.5.

Table 1.3 reports the estimated results. First, scarring effects impair labor market outcomes in particular for men. Indeed, men and women could be differently affected by past unemployment episodes. For instance, women might be more likely to react by permanently withdrawing from the labor market. Moreover, if unemployment is more common among women, an early unemployment event experienced by a woman may generate a weaker signal and less adverse effects on future labor market performances. Second, the magnitude of the labor penalties is larger in the short-term (up to 4 years) rather than in the medium- or long-run. Third, results about identification strategies are confirmed because of the larger negative effect when using selection on observables methods. Moreover, the reason of unemployment matters: displaced workers are more penalized especially in terms of future earnings, but also concerning future employment according to WALS results. The use of survey data is significant only in the labor earnings regression. Focusing on employment as a labor market outcome, a noteworthy point is related to the treatment variable used in the studies: the occurrence of unemployment is more important and penalizing than its duration, in line with Böheim and Taylor (2002). Further covariates, like study-quality measures and geographical area, do not explain effect heterogeneity across results.

Finally, in order to shed lights on the magnitudes of the scarring effect under different

Table 1.3: Heterogeneity in the estimated effect

	Labor earnings			Employment		
	BMA	WALS	OLS	BMA	WALS	OLS
Precision effect	**-0.006**	-0.004	-0.003	**0.029**	**0.060**	0.026***
	(0.005)	(0.009)	(0.002)	**(0.011)**	**(0.032)**	(0.002)
Publication bias	**-49.055**	**-40.365**	-46.732	**-65.300**	**-73.222**	-61.878**
	(25.607)	**(29.850)**	(27.345)	**(37.669)**	**(42.667)**	(27.083)
Sex (Ref. category = Males + Females)						
Males	**-0.002**	**-0.002**	-0.003**	-0.001	**-0.003**	
	(0.002)	**(0.001)**	(0.002)	(0.003)	**(0.003)**	
Females	0.001	**0.003**		**0.006**	**0.007**	0.010***
	(0.002)	**(0.002)**		**(0.006)**	**(0.004)**	(0.001)
Estimation strategy (Ref. category = Selection on unobservables; Field Experiment)						
Selection on observables	**-0.039**	**-0.028**	-0.051***	-0.002	**-0.058**	
	(0.022)	**(0.016)**	(0.009)	(0.010)	**(0.037)**	
Selection on unobservables				-0.000	-0.021	
				(0.005)	(0.023)	
Cause of unemployment episode (Ref. category = Not specified)						
Youth unemployment	0.002	-0.000		-0.002	-0.011	
	(0.005)	(0.006)		(0.006)	(0.014)	
Job displacement	**-0.008**	**-0.011**	-0.010***	-0.001	**-0.014**	
	(0.005)	**(0.006)**	(0.001)	(0.006)	**(0.013)**	
Data (Ref. category = Administrative data)						
Survey data	**-0.008**	**-0.016**	-0.010***	-0.000	-0.008	
	(0.007)	**(0.010)**	(0.003)	(0.004)	(0.014)	
Country (Ref. category = Anglosaxon countries)						
Scandinavian countries	-0.011	**-0.019**		0.000	-0.005	
	(0.020)	**(0.019)**		(0.004)	(0.021)	
European countries	0.001	-0.002		-0.001	0.015	
	(0.002)	(0.007)		(0.003)	(0.016)	
Multi-country analysis	0.000	0.003		0.002	0.023	
	(0.002)	(0.014)		(0.022)	(0.068)	
Treatment measure (Ref. category = Unemployment duration)						
Unemployment status	0.000	0.004		**-0.041**	**-0.028**	-0.040***
	(0.002)	(0.007)		**(0.009)**	**(0.016)**	(0.010)
Time horizon (Ref. category = Medium- and long-term)						
Short term	**-0.008**	**-0.007**	-0.008**	**-0.039**	**-0.037**	-0.042***
	(0.001)	**(0.001)**	(0.003)	**(0.003)**	**(0.003)**	(0.002)
Study-quality measures						
Year of publication	-0.000	-0.000		-0.000	**-0.002**	
	(0.000)	(0.000)		(0.001)	**(0.001)**	
Published in peer-reviewed journal	-0.000	0.002		-0.000	-0.017	
	(0.002)	(0.007)		(0.004)	(0.025)	

Notes: Covariates for which the PIP is above 0.5 in BMA and for which the t-ratio of its coefficient is greater than 1 in absolute value for WALS are reported in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

combinations of study features, we computed the expected partial correlation coefficients from the frequentist check after BMA for the most frequent combinations of these sources of heterogeneity (95% of the sample), while assuming the absence of publication bias (δ_1 set to zero). Table 1.4 displays the results. The expected r varies from -0.004 to -0.056 in case of studies focused on employment outcomes, and from -0.006 to -0.071 when the labor market outcome concerns labor earnings. The most important penalties arise when the unemployment spells occur due to job displacement, as well as in case of studies using survey data or estimating short-term effects of previous unemployment incidence (rather than its duration). As for the identification strategies, the strongest negative effect comes from study results with an identification strategy based on observables. Finally, the lower labor market penalties for women compated to men may suggest that unemployment experiences for the former may generate a weaker negative signal and fewer adverse consequences.

Table 1.4: Expected partial correlation coefficients of the scarring effects for combinations of covariates

	Coeff.	Std. Error	Observations
a) Labor earnings			
+ survey data + short term	-0.021***	0.003	72
+ survey data + short term + males	-0.024***	0.003	58
+ males	-0.006***	0.001	42
+ short term + job displacement + survey data	-0.031***	0.003	25
labor earnings	-0.003	0.002	24
+ survey data + males	-0.016***	0.003	23
+ short term	-0.011***	0.002	17
+ short term + job displacement + survey data + males	-0.034***	0.002	16
+ survey data	-0.012***	0.004	12
+ short term + job displacement	-0.021***	0.001	11
+ selection on observables	-0.054***	0.009	10
+ short term + males	-0.014***	0.003	9
+ short term + job displacement + selection on observables	-0.073***	0.009	8
+ short term + males + job displacement	-0.024***	0.002	4
+ job displacement + selection on observables	-0.065***	0.009	4
+ selection on observables + short term	-0.062***	0.009	4
+ selection on observables + short term + survey data	-0.071***	0.009	2
+ selection on observables + survey data	-0.064***	0.009	2
b) Employment			
+ unemployment status + short term	-0.056***	0.001	122
+ short term	-0.014***	0.002	75
+ unemployment status + short term + females	-0.046***	0.001	24
+ unemployment status	-0.016***	0.000	19
+ unemployment status + females	-0.006***	0.001	10
+ short term + females	-0.004*	0.002	7

Notes: *** Significant at 1%, ** significant at 5%, * significant at 10%. Covariates not mentioned in each line are fixed at the reference category: youth unemployment, experiences of nonemployment during recessions, and unemployment episodes for which the reason is not clearly specified; administrative data for type of data; selection on unobservables for identification strategy; males & females for sex; medium- and long-term for time horizon; unemployment duration/number of unemployment spells for treatment measure.

1.5 Conclusions

Although the empirical literature has produced a lot on the study of unemployment scarring effects since the 1980s, to our knowledge there are no rigorous and recent surveys on the matter. The present article aimed to fill this gap by presenting an overview of empirical evidence that applies causal inference and is related to the scarring effects of previous unemployment episodes on subsequent wages and employment opportunities. Moreover, the second contribution of this paper consisted in the use of meta-regression techniques that allowed us to check for the presence of publication bias and explore sources of effect heterogeneity across several study-related features.

Empirical evidence is clear and homogeneous in detecting significant, and often persistent, wage losses following unemployment spells and strong state dependence in unemployment persistence. Moreover, the phenomenon of unemployment scarring effects collected empirical confirmations despite different datasets used, countries considered, time span covered and the methodology used in order to identify the causal effect. Little differences across empirical findings concern the magnitude of these detrimental effects, based on the reason and length (or number) of nonemployment spells: while in the literature the unemployment periods experienced by school-leavers or by laid-off workers are particularly penalizing, the negative effect on subsequent labor market performances seems to be less stigmatizing in the case of plant closures or when the local unemployment rate is high, as suggested by the signaling theory. Moreover, further heterogeneity dimensions briefly discussed in the Appendix might concern age, tenure and education level: empirical evidence suggests that penalties after job displacement are larger for older workers because of their longer tenure and more-accumulated firm specific human capital that new employers do not value, or because they have less recent education and training about new skills demanded by firms (Eliason and Storrie, 2006).

To empirically test the magnitude of the scarring effects under different study features, we performed a meta-regression analysis by focusing on some of these heterogeneity dimensions and providing results divided by labor market outcome. Our findings confirm the presence of scarring effects on both future employment and labor earnings, under the assumption of a common true effect. By exploiting several study-related characteristics, we used model averaging strategies to explore possible sources of effect heterogeneity and estimated the expected partial correlation coefficients for different combinations of these features. We conclude that scarring effects on labor earnings are larger when unemployment is due to job displacement, and when the identification strategy is based on selection on observables. Furthermore, unemployment incidence, rather than its duration, seems to have the major negative impact on future employment. Finally, for both the labor market outcomes, the negative effect of previous unemployment is greater in the short-term and more penalizing for men than for women.

We can draw some policy implications from the collected empirical evidence and from our meta-regression results. Focusing on the reason behind unemployment spells, on the one hand, the creation of conditions that favor work experience as quickly as possible after school completion appears to be an urgent issue. On the other hand, policy makers should not continue to follow the path of labor market reforms that facilitate layoffs if they want to avoid the stigma effect found in the empirical literature, which is particularly scarring for dismissed workers. Policy makers should also favor training programs to avoid losses of human capital for younger unemployed but even for older workers (see e.g. Picchio and van Ours, 2013). Finally, one way to mitigate the wage scars highlighted in this study and facilitate the search for a better job match could be suggested by that strand of the literature that analyzes the duration of unemployment insurance (see e.g. Gangl, 2004; Tatsiramos, 2009; Nekoei and Weber, 2017). However, this is not the focus of our study and could be a topic of investigation for future research.

Chapter 2

Off to a bad start: youth nonemployment and labor market outcomes later in life

2.1 Introduction

Since 1980s, many labor economists have focused their research activity on studying the impact of early unemployment on subsequent labor market career and whether these effects are temporary or permanent. Empirical literature provides several findings about the so-called "unemployment scarring effects" as regard both wage losses and the probability of remaining unemployed in the future. Indeed, in addition to the immediate loss in terms of not perceived income and lack of human capital accumulation, past history of unemployment can also have longer-term or permanent effects by increasing the likelihood of experiencing future unemployment and generating lower subsequent wages (Arulampalam et al., 2001; Gregg and Tominey, 2005).

Empirical evidence is clear in detecting significant, and often persistent, wage penalties and lower employment probabilities after unemployment experiences, despite different dataset used, countries considered, time span covered and econometric strategies applied in order to identify the causal effect. Little differences concern the magnitude of the scarring effects: for instance, unemployment episodes experienced by school-leavers or by laid-off workers are particularly penalizing (see e.g. Jacobson et al., 1993; Burda and Mertens, 2001; Mroz and Savage, 2006), while the negative effect is less stigmatizing in cases of plant closures (Gibbons and Katz, 1991) or during economic downturns (Omori, 1997).

This article aims to provide evidence about the impact of youth unemployment experiences after school completion on subsequent labor market performances in Italy, estimating the effect up to 25 years after school completion. Thus, our paper adds to the debate on scarring effects by answering the following questions: i) What is the causal impact of remaining unemployed for Italian school-leavers on subsequent labor earnings and their participation in labor market? ii) How does it changes over time and how long do these penalties take to fade away? The contribution of our analysis is twofold. First, we shed further light into the scarring effects of early nonemployment by estimating short, medium and long-term impacts, measured up to 25 years after school completion. Second, we focused on the Italian case, which is particularly interesting. Indeed, we do not have many empirical analysis on this topic related to the Italian labor market, although the issue of unemployment scarring effects (and, in particular, the case of early unemployment) is crucial both from a socio-economic point of view and from a policy perspective, if we note that the average duration of the school-to-work transition for young people aged 18-34 was 2.88 years in Italy, which corresponds to the highest average duration in Europe (Pastore et al., 2020, 2021).

To our knowledge, only two articles studied the stigma effects of nonemployment for Italian youth. Lupi et al. (2002) investigated only the effect of individual unemployment experiences on future wages and showed that they tend to be scarring only in the North, where the aggregate unemployment rate is lower than in the South. Similar results are provided by Tanzi (2022), who highlighted that the negative effects of early non-employment on the propensity to experience further non-employment periods in subsequent years are smaller during recession or in regions characterized by high unemployment rates.[1]

To investigate these issues, we made use of the AD-SILC database, which is obtained by matching the IT-SILC database and administrative data from the National Social Insurance Agency (INPS). For each interviewee of the IT-SILC, the dataset contains and allows us to reconstruct all the working history as an employee up to the end of 2013.

[1]Moreover, labor market performances have deteriorated across cohorts, with 11% lower entry wages for the younger cohorts according to Naticchioni et al. (2016). Finally, Raitano and Fana (2019) estimated that not only new entrants start to work more frequently through atypical contracts, but they are also characterized by lower wages at the entry and along the first six years of career. However, differently from our study, they used the AD-SILC database to evaluate whether the increase in labor market flexibility has been associated to changes in post-reform entrants' economic conditions.

We followed an approach similar to Cockx and Picchio (2012, 2013),[2] contrasting the job profiles from the time of leaving upper secondary school of individuals with different nonemployment experiences at the end of the studies. Moreover, we followed Picchio et al. (2021) as concern the identification of the treatment effect, by including a series of individual time-varying factors related to unobserved characteristics. These latent variables are crucial to model the unobserved heterogeneity due to persistent differences in unobservables characteristics such as ability, motivation or difference in search intensity. In this sense, we set up a factor analytic model (Carneiro et al., 2003; Heckman and Navarro, 2007) in which individuals differ in unobserved characteristics jointly affecting selection into treatment and subsequent labor market outcomes. We made use of a non-parametric identification strategy where the unobserved determinants of the treatment and the outcomes are time-varying. The longitudinal structure of our dataset allowed us for the reconstruction of a complete working history for each individual and provides multiple observations over time of the endogenous variables. Moreover, we exploited two selection-free measures of the latent factor: a measure of employment experiences before high school diploma and the number of siblings when the individual was 14 years old in order to capture social and family background.

The article is organized as follows. Section 2.2 summarizes the empirical literature. Section 2.3 describes data and sample. Section 2.4 illustrates the econometric strategy. Section 2.5 discusses our findings. Section 2.6 concludes.

2.2 Literature review

Theoretical predictions on unemployment scarring effects can be derived from two main strands of the economic theory: the human capital theory and the signalling theory. According to the former, scarring effects are related to the depreciation of workers' general skills and knowledge following unemployment spells, or to the lack of accumulation of human capital occurring in case of early unemployment experiences (Mincer, 1974; Becker, 1975; Pissarides, 1992). Following the signalling theory, employers use past history of unemployment of a worker as a signal of low productivity, and the magnitude of the stigma effect on worker's subsequent labor market outcomes may depend on the cause

[2] Cockx and Picchio (2012) focused on the dependence of job stability on past labor market states, while Cockx and Picchio (2013) analysed the employment stability and integrate wages as an endogenous variable in the analysis.

of previous unemployment spells (Spence, 1973; Vishwanath, 1989; Lockwood, 1991).

Empirical evidence about scarring effects are present not only concerning youth unemployment episodes, but considering plant closures, job displacements, and unemployment experiences in general. Large and permanent wage scars caused by displacements or mass-layoffs are found in US labor market (see e.g. Ruhm, 1991; Jacobson et al., 1993; Stevens, 1997). About Europe, permanent wage penalties are estimated in UK (Arulampalam, 2001; Gregory and Jukes, 2001) as well as in Germany or Scandinavian countries for displaced workers (see e.g. Burda and Mertens, 2001; Eliason and Storrie, 2006), or after plant closure (Couch, 2001). Strong evidence of significant structural dependence induced by previous unemployment experience is highlighted by several authors too (Arulampalam et al., 2000; Gregg, 2001; Böheim and Taylor, 2002; Stewart, 2007; Biewen and Steffes, 2010; Deelen et al., 2018).

Within this strand of the literature, we are interested in the unemployment after school completion, focusing on the penalties concerning employability and subsequent wage dynamics for unemployed school-leavers. According to Corcoran (1982) and Ellwood (1982), early nonemployment causes lower future earnings also 10 years after school completion, whereas Mroz and Savage (2006) did not find long-lived persistence in unemployment spells but evidence of blemishes from unemployment and lower wage. No lagged duration dependence is found in Doiron and Gørgens (2008), who estimated that an additional spell of unemployment increases the probability of being unemployed in the future. The longer the unemployment spell upon graduation the more substantial are subsequent individual earning losses and higher the unemployment probability after 5 years for Swedish youths (Gartell, 2009; Nordström Skans, 2011). The same is found in Belgium, where increasing time spent in nonemployment in the first 2.5 years since graduation decreases both annual earnings and hours worked by 10 per cent and 7 per cent 6 years later (Ghirelli, 2015); or increases subsequent unemployment probability as well as its duration, an effect that remain substantial even a decade after leaving school (Heylen, 2011). According to Cockx and Picchio (2013), job finding probability decreases from 60 per cent to 16 per cent for men and from 47 per cent to 13 per cent for women in the following 2 years if the entry is delayed by one year. Similar findings about young people who finished their studies in Finland, where the incidence of unemployment on future labor market performances is a scarring effect of 20 percentage points in terms of unemployment probability (Hämäläinen, 2003). Burgess et al. (2003) found evidence of heterogeneity in responses of school-leavers, estimating adverse effects on later unemployment of early career unemployment for the unskilled and the reverse for the more skilled. Using an instrumental variables approach, Gregg and Tominey (2005) estimated large and significant wage penalties caused by youth unemployment in the magnitude of 13-21 per cent at age 42 in UK. Tanzi (2022) found that the size of this scarring effect in Italy depends on regional labor market characteristics: in particular, the scarring effect is smaller the higher is the regional unemployment rate or during economic downturns. Early unemployment would increase the probability of future unemployment by 3.42 percentage points and each additional nonemployment spell increases this probability by 0.078 percentage points in Germany (Manzoni and Mooi-Reci, 2011), and these scarring effects are likely to be significant and long-lasting in Schmillen and Umkehrer (2017). According to Möller and Umkehrer (2015), wage penalties are found to be different across the earning distribution since an increase in early-career unemployment causes persistent earning losses of about 56 per cent for workers at the bottom, whereas workers with higher income only 7 per cent.

Slightly different research question are considered by other researchers: Hällsten (2017) analysed the link between educational failure and future adverse outcomes, estimating that university dropouts spend 2.4 percentage points more of their first 8 years in a state of low earnings compared to never entrants. Helbling and Sacchi (2014) investigated scarring effects of early unemployment among young adults with vocational studies, whereas Kahn (2010), Oreopoulos et al. (2012) and Kawaguchi and Murao (2014) estimated large and persistent negative effects on wages of graduating in a worse economy, which could persist for 5 or even 15 years after college graduation.

2.3 Data and sample

2.3.1 Sample selection criteria

Our empirical analysis was based on the AD-SILC database, which is obtained by matching two data sources: i) the IT-SILC database covering the period 2004-2012 gathered by the Italian National Institute of Statistics (ISTAT); ii) the administrative data on labor market contracts from the National Social Insurance Agency (INPS). The latter manages social security so contains gross earnings and the number of working days for each working episode in each year for all the salaried employees, and allowed us to rebuild the working history of each individual as an employee up to the end of 2013. Furthermore, we matched the AD-SILC with the regional time series of unemployment and employment rates, real GDP and GDP growth rate from ISTAT, used as time-varying controls in our empirical analysis.

We extracted data on Italian individuals interviewed in 2005 and 2011: these two waves are the only ones with the *ad hoc* module on intergenerational transmission of poverty and disadvantages, which provides information on the family situation when the respondents were 14 years old. We exploited this predetermined information to model unobserved heterogeneity, such as the attachment to the labor market or the cultural, family and social background. Each individual is interviewed for 4 consecutive years.[3]

The starting sample of 98,529 units contained personal and child-related information on all individuals of 2005 and 2011 surveys. We further selected individuals who were not in education in 2003 if interviewed in 2005 and in 2009 if interviewed in 2011, to have at least 3 years of labor market information between school leaving and the IT-SILC interview. Moreover, we restricted the sample to individuals who exited formal education after 1976, because ISTAT database provides the regional time series which we used as time-varying controls only from 1977. The following match with data on province of birth reduces the sample to 34,180 individuals. Since we had no information on business cycle at province level about other countries, individuals born abroad were not included in the analysis. Table 2.1 reports in more detail the selection criteria which reduced the sample to individuals for whom, thanks to the INPS administrative data, we rebuilt all their past labor market histories up to the moment in which they were interviewed for the IT-SILC. Individuals not included in the INPS database were dropped because self-employed or inactive, so our analysis is based only on salaried employees.

However, we focused only on those ones who obtained the high school diploma as the highest level of education, since while for them the observation period always starts from the following September 1, we did not have information about the month in which each graduate achieved the tertiary degree. We excluded individuals younger than 26 at the time of the interview because the *ad hoc* module was submitted only to individuals older than 25. At the same time, we excluded individuals with missing data about the number of siblings at 14 years old because this predetermined information is used as a measurement equation in our identification strategy. After applying these selection criteria, our final sample consisted of 10,295 observations, of which 5,396 males and 4,899 females.

[3]For the 139 individuals interviewed both in 2005 and 2011, we only keep the 2011 data, since more recent and therefore richer in the construction of the working history.

Table 2.1: Sample size across selection criteria

	Individuals left in the sample	Individuals removed
Individuals in IT-SILC, waves 2005 and 2001	98,529	–
After removing individuals with errors on gender	98,513	16
After removing individuals observed twice from the wave 2005	98,374	139
After taking only individuals who exited formal education after 1976 and individuals who are not in education in 2003 if interviewed in 2005 and in 2009 if interviewed in 2011	34,180	64,194
After removing individuals with missing county of birth	34,167	13
After removing individuals born abroad	31,134	3,033
After removing individuals not included in the INPS database	29,576	1,558
After removing individuals due to incorrect information related to working periods	29,481	95
After removing graduates and individuals without high school diploma	12,834	16,647
After removing individuals younger than 16 or older than 21 at the time of their highest diploma	11,787	1,047
After removing individuals with yearly earnings greater than 800,000€ or daily wages greater than 5,000€	11,781	6
After removing individuals younger than 26 at the time of the interview	10,559	1,222
After removing individuals not observed at least 5 years after high school diploma	10,447	112
After removing individuals with missing data about the number of siblings at 14	10,375	72
After removing individuals with daily wages greater than 250€ (outliers)	10,295	80
Final sample	10,295	88,234

2.3.2 Descriptive statistics

Our sample is composed only by individuals who obtained high school diploma more than 3 years before the IT-SILC interview are kept. While we can observe their labor market outcomes at least up to 3 years after school leaving, the number of individuals we can follow for a longer labor market histories is decreasing with the size of the time window considered after graduation. Table 2.2 shows the number of observations from 5 to 25 years after school completion grouped by periods of 5 years. At the same time, we provided some descriptive statistics concerning our treatment variable, that is the fraction of days of nonemployment during the first 3 years after school completion, and other time-invariant characteristics by distinguishing among males and females. As we can see, the number of observation 25 years later amounts to 2,792 males employees and 2,423 females. Table A2.1 in the Appendix gives complete information on the age distribution at diploma. The main differences among males and females appear to be related to the average number of kids at different year after school exit and to the fraction of days in employment the year before high school diploma.

Table 2.3 shows summary statistics of our dependent variables from 5 to 25 years after high school diploma, distinguishing them between yearly labour earnings and participation in the labor market, i.e. yearly fraction of days spent in employment. The main differences among males and females are related to earnings outcomes: while men obtain a 21% higher average yearly labor earnings than women 5 years after school completion, this gap reaches 45% 25 years after high-school diploma. Indeed, Table 2.3 shows that the

Table 2.2: Sub-samples by different years after school completion

Year after school completion	Observations	Nonemployment during 3 years after school exit	Father's education	Mother's education	Father at work	Mother at work	Number of siblings at 14	Number of kids	Employment 1 year before school exit
				Males					
5	5,396	0.66	1.30	1.27	0.87	0.31	1.27	0.03	0.08
10	5,310	0.65	1.30	1.27	0.87	0.31	1.28	0.19	0.08
15	4,864	0.66	1.29	1.26	0.87	0.30	1.30	0.52	0.08
20	3,947	0.66	1.27	1.23	0.86	0.28	1.35	0.81	0.08
25	2,792	0.64	1.23	1.18	0.86	0.26	1.42	1.05	0.08
				Females					
Year after school completion	Observations	Nonemployment during 3 years after school exit	Father's education	Mother's education	Father at work	Mother at work	Number of siblings at 14	Number of kids	Employment 1 year before school exit
5	4,899	0.66	1.27	1.23	0.88	0.32	1.29	0.13	0.04
10	4,722	0.66	1.27	1.23	0.88	0.32	1.28	0.49	0.04
15	4,235	0.66	1.28	1.23	0.88	0.31	1.29	0.90	0.04
20	3,383	0.65	1.28	1.21	0.88	0.30	1.34	1.17	0.04
25	2,423	0.64	1.25	1.18	0.88	0.29	1.38	1.29	0.04

Notes: The table shows the mean values of the treatment variable and other time-invariant characteristics predetermined with respect to the treatment for the number of individuals observed from 5 to 25 years after school completion.

average value of yearly labor wages for males increases by 130% along the time window considered, whereas for females it does not even double at the 25th year. Differences concern also the fraction of days spent in employment: the average values are quite similar during the first 5 years, but this gap increases with time in favor of higher values for men.

Table 2.3: Outcome variables at different years after school completion

	Males					Females				
		Yearly labor earnings (€)[(a)]		Days in employment [(b)]			Yearly labor earnings (€)[(a)]		Days in employment [(b)]	
Year after school completion	Observations	Mean	Std. Dev.	Mean	Std. Dev.	Observations	Mean	Std. Dev.	Mean	Std. Dev.
5	5,396	12,352.83	10,627.46	0.62	0.45	4,899	10,190.33	9,658.65	0.57	0.46
10	5,310	18,374.76	12,607.04	0.79	0.38	4,722	13,077.41	11,113.04	0.67	0.44
15	4,864	22,759.81	14,176.09	0.85	0.34	4,235	14,770.19	11,989.04	0.73	0.41
20	3,947	25,909.90	16,449.95	0.87	0.31	3,383	17,242.18	13,109.76	0.79	0.37
25	2,792	28,344.23	18,118.38	0.89	0.28	2,423	19,601.58	13,708.33	0.83	0.33

[(a)] Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
[(b)] These outcome variables measure the fraction of days spent in employment.

Tables A2.3 and A2.4 report first marginal correlations between nonemployment during the first 3 years after high school diploma and the two main outcome variables. Our preliminary findings reveal that the impact of increasing the time spent in nonemployment after high school diploma has a large negative effect on yearly labor earnings and participation in labor market for both males and females until 25 years later. A 10 percentage point increase in the time spent in nonemployment in the first three years after school completion is associated to a decrease of €1,230 (€1,280) in yearly earnings and of 5.5

(5.2) p.p. in the yearly fraction of time spent in employment for women (men) 5 years after school completion. Later, these negative correlations fade away but they are still sizable and significant; 25 years after school completion a 10 percentage point increase in early nonemployment is related to a decrease of €488 (€384) in yearly earnings and of 1.0 (0.3) p.p. in the yearly fraction of time spent in employment for women (men). The estimated β_t from Ordinary Least Squares (OLS) are graphically displayed in Figure A2.4 along with 95% confidence intervals, but such estimation results cannot be given a causal interpretation: labor market performances after diploma are indeed endogenous because of both time-constant and time-varying unobserved traits, which jointly determine both the labor market outcomes in the future career and the experiences after school completion, such as labor market attachment or different job search strategies.

More detailed descriptive statistics are depicted in section A of the Appendix. In the next sections we outlined an econometric model to evaluate the causal impact of unemployment episodes after school completion on future earnings and participation at different moments in the subsequent career. The proposed econometric model is aimed at disentangling the true causal effect of period of nonemployment from the spurious one induced by systematic differences across individuals with different labor market histories.

2.4 Econometric model

2.4.1 Estimation framework

Let $i = 1, ..., n$ index an individual and $t = 1, ..., T_i$ index the time elapsed since school completion. Obviously, the observable time elapsed since high school diploma (T_i) differs across individuals since it depends on the time between the IT-SILC interview and the year of school exit. We denoted as Y_{it}^j the j-th labor market outcome, with $j = 1, 2$ since we analysed both labour earnings and the fraction of days spent at work. For each individual i the observed labor market outcome j at time t can be written as

$$Y_{it}^j = \beta_t^j TR_i + \mu_t^j(X_{it}^j) + \epsilon_{it}^j \tag{2.1}$$

where β_t^j is the effect of the treatment variable TR on outcome j at time t, μ_t^j is a function of observed covariates X_{it}^j and ϵ_{it}^j collects the individual time-varying unobservables. We are interested in the effect of early unemployment experiences on future wages and participation in the labor market, so our treatment is a continuous variable corresponding to the fraction of days spent in nonemployment during the first 3 years after school completion. The intensity of the treatment TR_i is specified as follows

$$TR_i = \nu(Z_i) + u_i \tag{2.2}$$

where $\nu(\cdot)$ is a function of a vector of covariates Z_i, which are realized either before the end of secondary school (for example mother's highest education) or in the three years after school exit (like number of kids, labor market status or GDP growth at regional level) and u_i is individual unobserved heterogeneity. As previously illustrated, we can follow individuals over time up to 25 years after school completion. Thus, we estimated the impact of our treatment variable on both yearly labor earnings and yearly fraction of days spent at work every five years after diploma, and so $t \in \{5, 10, 15, 20, 25\}$. In summary, we estimated the parameters for 10 outcomes separated by sex, along those entering selection and two measurement equations.

2.4.2 Identification strategy

The identification of the effect of unemployment spells after school completion on future labor market outcomes requires to take into account unobserved heterogeneity across individuals, which might affect the occurrence of early nonemployment events after school exit and subsequent labor market outcomes. This is related, for example, to differences in unobserved characteristics such as labor force attachment, ability, motivation, liquidity constraints and job search strategies. Moreover, unobserved heterogeneity is likely to change over time. For instance, the liquidity constraints of those individuals who experience more intensively longer nonemployment events may become tighter and more relevant over time, increasing the job search intensity, lowering the reservation wages and having therefore an impact on labor market outcomes that may be varying over time. Finally, some determinants of early nonemployment, like preferences for family formation or parenthood may also change over time. At some point after school exit, individuals may form a family and have kids, modifying the preference towards the work-family balance, which is a time-varying unobservable very likely to matter for future labor market outcomes. Hence, we need to specify the joint distribution of the unobserved components determining both the labor market outcomes and the selection into treatment.

In this sense, we set up a factor analytic model (Carneiro et al., 2003; Heckman and Navarro, 2007; Fruehwirth et al., 2016; Cockx et al., 2019; Picchio et al., 2021):[4] the unobserved terms of outcomes and selection into treatment equations are composed of a latent factor θ, which collects the unobserved differences among individuals that determine the selection into treatments and the unemployment effects on subsequent labor market outcomes, and error terms that are conditionally independent given the factor. Indeed, to account for such heterogeneity we can recover the joint distribution of the unobservables in the selection (u_i) and outcome equations (ϵ^j_{it}) by imposing a factor structure (Fruehwirth et al., 2016). Hence, we have

$$\epsilon^j_{it} = \alpha^j_t \theta_{it} + \varepsilon^j_{it} \tag{2.3}$$

$$u_i = \lambda\theta_i + v_i \tag{2.4}$$

where θ_{it} is a latent factor in $\theta_i = (\theta_{i1}, ..., \theta_{iT})$ with a multivariate distribution with $cov(\theta_{it}, \theta_{it'}) \neq 0$, for all $t \neq t'$. It is a vector of mutually independent factors, as well as the error terms. In summary, the unobserved terms in the outcome and treatment equations are made of a latent factor θ which collects unobserved differences among individuals, and a random component ε_{it} and v_i. Unobserved heterogeneity varies over time because of the factor distribution and a linear combination of the factor with time-varying coefficients α^j_t (the so-called *factor loadings*). Our framework differs from Fruehwirth et al. (2016), where latent variable is composed by general ability, cognitive ability and behavioral component. In our case, as in Picchio et al. (2021), unobservables are all included in a single latent factor θ, instead of differencing by several sources of unobserved heterogeneity. Following Carneiro et al. (2003), we relied on a set of selection-free measurements to control for the unobservables that jointly determine selection into treatment and its effect, and to reduce the degree of arbitrariness. We made use of predetermined information to specify our additional measures

$$M^l_i = \omega^l(S^l_i) + \xi^l\theta_{i5} + e^l_i \tag{2.5}$$

with $l = 1, 2$ and where M^l_i are predetermined information with respect to school com-

[4] Carneiro et al. (2003) studied the impact of different schooling levels on future returns; Fruehwirth et al. (2016) and Cockx et al. (2019) estimated how grade retention affects subsequent school performances; Picchio et al. (2021) investigated the effect of childbirth and its timing on female labour market outcomes in Italy.

pletion and selection into treatment. ω^l consists in a linear combination of observed covariates S_i^l that are illustrated in Table A2.5, ξ^l is a factor with time-varying coefficients and e_i^l is a zero-mean error term independent of both S_i^l and θ_{i5}.

We used two additional measurement equations which contain predetermined characteristics of each individual. These latent variables are crucial in order to model the unobserved heterogeneity due to persistent differences in unobservables characteristics such as ability or labor market attachment and/or unobservables persistent shocks that could simultaneously affect both selection into treatment and the outcome of the treatment. The first measure M_i^1 is a variable which corresponds to the fraction of days spent at work during the year before the school completion.

$$M_i^1 = s_i'\zeta^1 + \xi^1\theta_{i5} + e_i^1 \tag{2.6}$$

where e_i^1 has zero mean and variance $V(e_i^1) = \omega^2$. This measure is likely to be determined by a set of unobserved traits which include labor force attachment, motivation, ability, job search strategies, but also liquidity constraints, or family, social and cultural background. Such unobserved charachteristics should be relevant in explaining both labor market participation after school completion and labor outcomes in the future.

The second measure M_i^2 is the number of siblings the individual had when was 14 years old, which is a continuous variable specified as follows:

$$M_i^2 = s_i'\zeta^2 + \xi^2\theta_{i5} + e_i^2 \tag{2.7}$$

where e_i^2 has zero mean and variance $V(e_i^2) = \omega^2$. There is a strand of the literature which focused on the relation between the family size, investments in human capital and, in an indirect way, labor market outcomes. The idea is that increasing the number of siblings in a household might reduces the opportunity to study at college because of the resources dilution of parents' material resources on one hand, and increases need for other liquidity entries so determining an earlier participation in the labor market on the other hand. Blake (1981) suggested that the number of siblings, relative to other background variables, is found to have an important detrimental impact on a child's educational attainment and college plans, while families with fewer siblings provide more resources for the child and support the development of better educational outcomes. Indeed, the number of siblings has an indirect impact on future income through their influence on other control variables such as education, because of the limited resources parents must divide among their children greatly reduces the allocation of resources gained per child. Thus, siblingship could increase the risk that individuals will stop their education earlier than they should (Blake, 1989; Wijanarko and Wisana, 2019; Li and Hiwatari, 2020), although some studies did not find long-term traces of the negative effects of family size (see e.g. Åslund and Grönqvist, 2010).[5] As such, the second measure M_i^2 may encompass information on childhood household environment shaping the likelihood of success in the labor market in the adulthood.

Our factor structure is a special case of the one proposed by Carneiro et al. (2003). Furthermore, our model is a special case both of Fruehwirth et al. (2016) and Picchio et al. (2021). Their identification results related to the factor analysis can be invoked directly and specialized to fit our special case. Assuming that the regularity conditions (A-1 and A-2) in Carneiro et al. (2003) hold, the nonparametric identification of the deterministic parts of the model and of the joint distribution of the unobserved terms and their components, $(\epsilon_i^j, u_i, \upsilon_i)$, with $\epsilon_i^j = (\epsilon_{i1}^j, ..., \epsilon_{iT}^j)$, $\upsilon_i = (\upsilon_i^1, \upsilon_i^2)$, $\upsilon_i^l = \xi^l \theta_{i5} + e_i^l$, $j = 1, 2$ and $l = 1, 2$, is obtained as in Heckman and Smith (1998). As suggested by Carneiro et al. (2003), we satisfied their support condition (A-3) by including some continuous variables among the set of observed determinants of one outcome but excluded from the others. These variables are the regional employment rate, the regional unemployment rate, and the regional GDP growth rate: i) at the time when each individual was born in $\omega^l(S_i^l)$, for $l = 1, 2$; ii) at the time t in which the labor market outcome is evaluated in $\mu_t^j(X_{it}^j)$, for all j and t; iii) averaged across the three years after school completion in $\nu(Z_i)$. Both Bhargava (1991) and Mroz and Savage (2006) clarified why the variation of exogenous variables, like these regional rates, may be of help to identify the causal effects of endogenous variables in a dynamic discrete time panel data model. Indeed, these covariates implicitly provide additional identification conditions, resulting in significantly more degrees of freedom to control for endogenous determinants. Every lag of the exogenous time-varying regressor may indeed determine a separate effect on the current realization of the outcome. Table A2.5 clarifies in detail the exclusions across all the equations.

[5] Patrinos and Psacharopoulos (1997) showed that the age structure of siblings even matters, in conjunction with their activities: that is, having a greater number of younger siblings implies more age-grade distortion and a higher probability that the child works earlier, since schooling performance suffers when there are younger siblings in the household to care for. Alderman and King (1998) reviewed some studies in which family composition has differing effects among gender, suggesting the presence of unequal access to schooling and different parents' preferences.

2.4.3 Likelihood function

Let include all the parameters for our measurement, treatment and outcome equations in $\boldsymbol{\phi} = (\boldsymbol{\tau}^1, \boldsymbol{\tau}^2, \boldsymbol{\varphi}, \boldsymbol{\psi})$. The likelihood for individual i is the joint density of $(M_i^l, TR_i, \mathbf{Y}_i)$ conditional on observable and unobservable characteristics, so the individual contribution to the likelihood function can be written as

$$\mathcal{L}_i(\boldsymbol{\phi} \mid M_i^l, TR_i, \mathbf{Y}_i, S_i^l, Z_i, \mathbf{X}_i, \boldsymbol{\theta}_i) = g^l(M_i^l \mid S_i^l, \theta_{i5}; \boldsymbol{\tau}^l) h(TR_i \mid Z_i, \theta_{i5}; \boldsymbol{\varphi}) \prod_{j=1,2} \prod_{t=5,10,...,25} f(Y_{it}^j \mid TR_i, X_{it}, \theta_{it}; \boldsymbol{\psi})^{d_{it}}, \quad (2.8)$$

In order to account for the presence of individual time-varying unobserved heterogeneity, we recall that the vector of latent factor $\theta_i = (\theta_{i5}, ..., \theta_{i25})$ follows a multivariate discrete distribution with H support points. Thus, θ_i takes values θ^h, $h = 1, ..., H$ following a multi-logit parametrization

$$p^h = Pr(\theta_i = \theta^h) = \frac{exp(p^h)}{\sum_{u=1}^{H} exp(p^h)} \quad (2.9)$$

with normalization $\theta^1 = 0$ and $p^H = 0$. Moreover, we constrained the unobserved heterogeneity to be constant from 20 to 25 years after high school diploma, that is $\theta_{20}^h = \theta_{25}^h$, since the sample is halved approaching $T = 25$ and because we assumed that the unobserved traits tend to stabilize over time.

The i-th contribution to the likelihood becomes

$$\mathcal{L}_i(\boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\Theta} \mid M_i^l, TR_i, \mathbf{Y}_i, S_i^l, Z_i, \mathbf{X}_i) = \sum_{h=1}^{H} p^h \mathcal{L}_{ih}(\boldsymbol{\phi} \mid M_i^l, TR_i, \mathbf{Y}_i, S_i^l, Z_i, \mathbf{X}_i, \boldsymbol{\theta}_i = \boldsymbol{\theta}^h) \quad (2.10)$$

that is the likelihood in Equations (2.8), conditional on θ_i taking value θ^h and the matrix Θ contains the vectors of support points $(\theta^1, ..., \theta^H)$.

In order to estimate the abovementioned model, we made use of 3 different assumptions as concern the latent factor structure. In particular, we estimated our model i) without unobserved heterogeneity; ii) with time-constant unobserved heterogeneity with discrete distribution; iii) with time-varying latent factor with discrete distribution. The differences across these three specifications are reported in Table 2.4, which shows post-estimates characteristics such as the log-likelihood values and the Akaike and Bayesian information criteria (AIC and BIC, respectively).

Table 2.4: Summary statistics on the estimated models across different assumptions on the unobserved heterogeneity

	Without unobserved heterogeneity	Time-constant unobserved heterogeneity	Time-varying unobserved heterogeneity
a) Males			
Number of parameters	160	180	212
Log-likelihood	53673.34	48731.48	39755.52
AIC	107666.68	97822.96	79935.03
BIC	108721.63	99009.78	81332.83
Distribution of the latent factor	–	Discrete	Discrete
Number of support points of the latent factor	–	5	10
b) Females			
Number of parameters	160	180	212
Log-likelihood	44541.40	39831.57	32911.81
AIC	89402.80	80023.15	66247.62
BIC	90442.29	81192.57	67624.94
Distribution of the latent factor	–	Discrete	Discrete
Number of support points of the latent factor	–	5	10

Notes: AIC = Akaike information criterion; BIC = Bayesian information criterion.

Gaure et al. (2007) and Cockx and Picchio (2013) suggested that the best way in choosing the number of support points is by minimizing the Akaike information criterion (AIC). Following this suggestion, we stopped at $H = 5$ support points when the presence of time-constant unobserved heterogeneity is accounted for. When we took into account that the latent factor could assume a time-varying structure, we increased the number of support points H until we reach 10. We stopped because the estimated coefficients of the treatment become stable. With this specification we obtained a further substantial improvement in the optimization of the log-likelihood function and in terms of information criteria. Sections B and C in the supporting information report the full set of estimation results under the different latent factor structures, while the next section only considers the average treatment effects on the equations for the labor market outcomes for both males and females across the three alternative specifications of the unobserved heterogeneity.

2.5 Estimation results

2.5.1 Impact on labour market outcomes

The core question of the analysis is whether experiencing nonemployment after school completion inflicts a scar on future labor market outcomes as measured by labor earnings and yearly fraction of days spent in employment.[6] Table 2.5 and Figure 2.1 display the impact of the fraction of time spent in nonemployment in the first 3 years after school completion on yearly labor earnings evaluated at $t \in \{5, 10, 15, 20, 25\}$ after the diploma, along the three different latent factor structures.

Shifting from panel (a) to panel (c) of Table 2.5 or from graph (a) to graph (c) of Figure 2.1, it clearly emerges that if time-varying unobservables were not accounted for, the negative impact of early nonemployment on subsequent earnings would be largely overestimated. Even if the early nonemployment penalty is much smaller when we control for time-varying unobserved heterogeneity, it is statistically significant up to 25 years since school completion; the scarring effect of early nonemployment is long lasting for both men and women. The estimates reported in Table 2.5 and Figure 2.1 are the impact of the fraction of time spent in nonemployment in the first three years since school completion going from 0 to 1. Hence, if the time spent in nonemployment just after school completion increases by 10 percentage points (pp), male (female) yearly earnings decrease by €382 (€492) 5 years after school completion. This penalty for men (women) is reduced to €225 (€140) 25 years after the diploma. Figure 2.1 visually shows that men and women experience a similar nonemployment penalty in the short run ($t = 5$ and $t = 10$). However, men suffer larger penalties in subsequent years.

Table 2.6 and Figure 2.2 display the estimated impact of early nonemployment on the yearly fraction of days spent in salaried employment in the future. Also in this case, not controlling for time-varying unobserved heterogeneity generates a large overestimation of the scarring effect of early nonemployment, both in size and in duration. Once controlling for time-varying unobserved heterogeneity, we find that early nonemployment negatively affects the labor market participation only in the short-term; a 10 pp increase in the time spent in nonemployment after school completion reduced the fraction of days spent in employment 5 years after the diploma by 0.65 (0.99) pp for men (women). This penalty becomes very close to zero and not significantly different from zero by the 10th year after

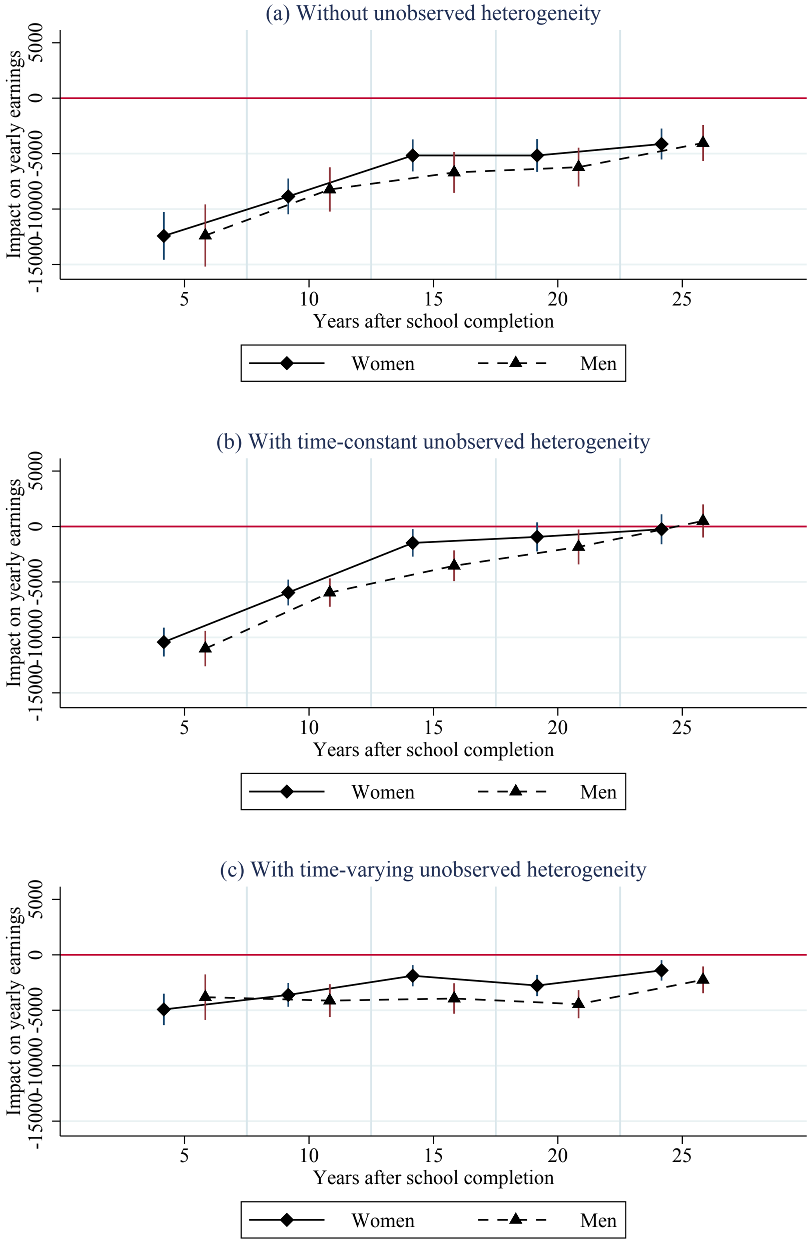
[6] In Table D2.11 in Appendix D we reported the estimated effects if we use daily earnings as outcome variable instead of yearly earnings.

Table 2.5: Average treatment effects of nonemployment during the first 3 years after school completion on yearly labor earnings (€)

Treatment variable across 3 different assumptions on the UH	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
(a) Without unobserved heterogeneity					
Men	-12382.93***	-8230.58***	-6699.58***	-6220.41***	-4041.89***
	(1431.78)	(1019.47)	(940.95)	(892.95)	(828.87)
Women	-12424.21***	-8856.36***	-5161.12***	-5170.76***	-4134.63***
	(1099.92)	(8220.13)	(736.22)	(755.33)	(711.54)
(b) With time-constant unobserved heterogeneity					
Men	-11006.14***	-5957.12***	-3537.09***	-1842.11**	502.03
	(814.19)	(654.60)	(707.99)	(802.19)	(762.80)
Women	-10422.62***	-5954.38***	-1474.79**	-935.41	245.864
	(663.50)	(593.01)	(633.67)	(669.56)	(690.10)
(c) With time-varying unobserved heterogeneity					
Men	-3815.09***	-4125.82***	-3935.38***	-4448.10***	-2254.04***
	(1048.98)	(757.46)	(701.46)	(646.82)	(616.17)
Women	-4919.67***	-3610.87***	-1880.88***	-2765.47***	-1399.38***
	(722.63)	(547.32)	(487.52)	(489.42)	(471.35)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: UH = Unobserved heterogeneity. Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index. The effect of one more year spent in nonemployment is equal to the estimated coefficients divided by three.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Figure 2.1: Impact of early nonemployment on yearly labor earnings (€)



Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index. The vertical segments are 95% confidence intervals.

school completion for both men and women. Finally, in the last year of observation ($t = 25$), individuals who experienced longer nonemployment events after school completion spend more time in the labor market, although the effect is small; an increase by 10 pp in the time spent in nonemployment after the diploma generates an increase by 0.43 (0.31) pp in the fraction of days spent at work 25 years later.

Table 2.6: Average treatment effects of nonemployment during the first 3 years after school completion on yearly fraction of days spent at work

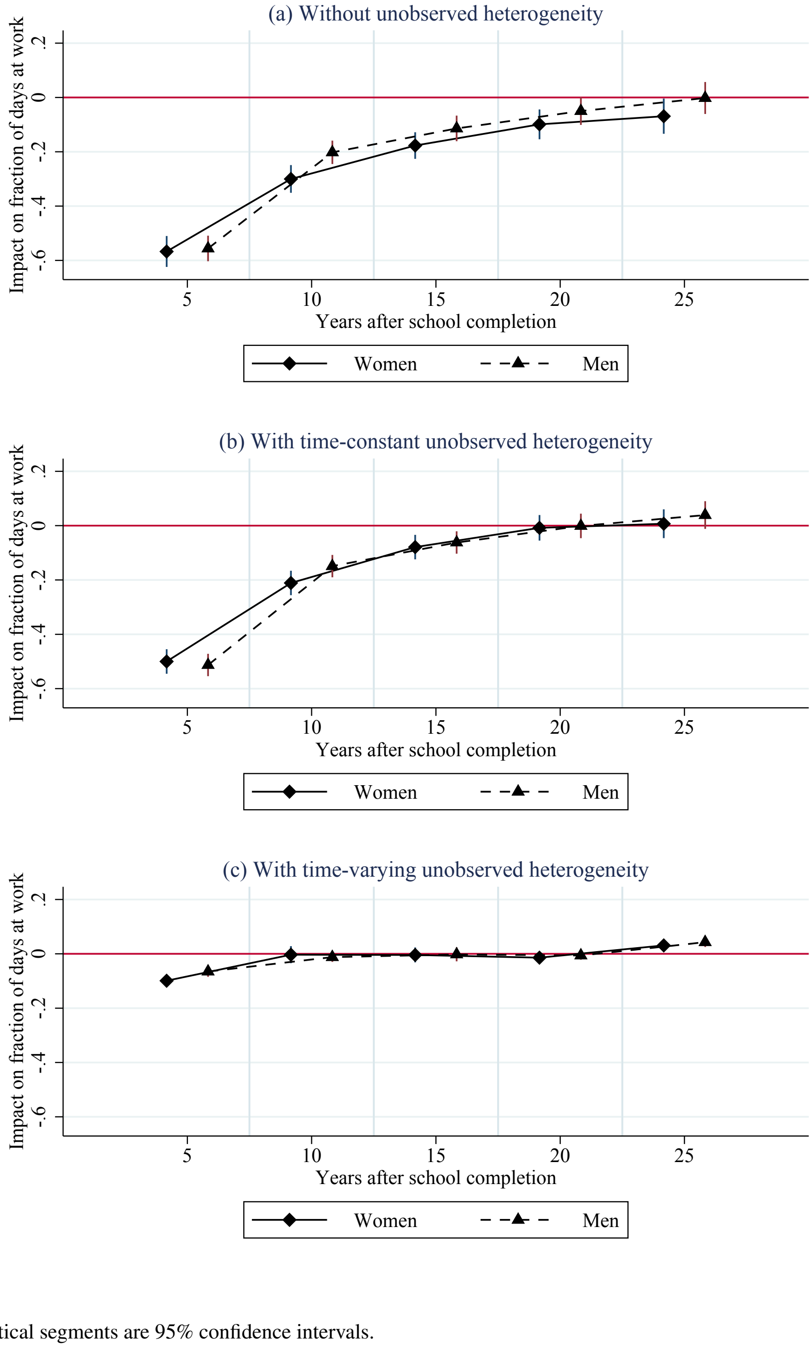
Treatment variable across 3 different assumptions on the UH	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
(a) Without unobserved heterogeneity					
Men	-0.556***	-0.202***	-0.114***	-0.050*	-0.002
	(0.024)	(0.022)	(0.024)	(0.026)	(0.030)
Women	-0.567***	-0.300***	-0.177***	-0.099***	0.069**
	(0.029)	(0.026)	(0.025)	(0.028)	(0.033)
(b) With time-constant unobserved heterogeneity					
Men	-0.513***	-0.149***	-0.062***	-0.001	0.039
	(0.021)	(0.021)	(0.021)	(0.023)	(0.026)
Women	-0.500***	-0.211***	-0.079***	-0.008	0.007
	(0.023)	(0.023)	(0.023)	(0.024)	(0.027)
(c) With time-varying unobserved heterogeneity					
Men	-0.065***	-0.012	-0.002	-0.006	0.043***
	(0.010)	(0.009)	(0.013)	(0.008)	(0.009)
Women	-0.099***	-0.003	-0.004	-0.014	0.031***
	(0.011)	(0.016)	(0.014)	(0.008)	(0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: UH = Unobserved heterogeneity. The effect of one more year spent in nonemployment is equal to the estimated coefficients divided by three.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table 2.7 quantifies the estimated impacts of early nonemployment on future labor earnings and participiation relatively to the average labor market outcomes by individuals who did not experienced nonemployment during the first 3 years after school completion.

When we control for time-varying unobserved heterogeneity, the negative effect of early nonemployment on labor earnings becomes smaller in magnitude, whereas the penalties in terms of labor participation are present only up to 5 years after school completion.

Figure 2.2: Impact of early nonemployment on yearly fraction of days spent at work



Notes: The vertical segments are 95% confidence intervals.

Table 2.7: Estimated impacts of early nonemployment during the first 3 years after diploma on future labor market outcomes, relative to the average in t for individuals who did not experienced early nonemployment

Years since school completion	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings					
Men	-26.19%	-19.95%	-17.69%	-18.36%	-9.12%
Women	-29.88%	-20.13%	-10.72%	-14.67%	-6.89%
b) Yearly fraction of days at work					
Men	-8.67%	-1.43%	-0.23%	-0.68%	4.94%
Women	-11.65%	-0.34%	-0.47%	-1.69%	3.48%

Notes: These figures are computed by evaluating the change in the labor market outcomes in a year implied by the estimated coefficients reported in Tables 2.5 and 2.6 relative to the average labor market outcomes in t of individuals who did not experienced nonemployment after diploma.

This suggests that when we include in the model the time-varying latent factor, we capture those latent traits which affect both selection into early nonemployment and future labor market performances. As an example, career-oriented individuals with higher abilities and motivations are more likely to have success in the labor market and therefore the negative impact of nonemployment on labor market outcomes is subject to upward bias if these characteristics were not accounted for. Moreover, differences in the estimated penalties between the model with time-constant and the model with time-varying unobserved heterogeneity indicate that the latent factor is subject to relevant variations over time. For example, the influence of the family background may diminish as a person ages (Gregg, 2001); further, liquidity constraints may change over time and individuals may reduce their reservation wages as they experience longer nonemployment spells, accepting therefore low quality jobs and translating into worse labor earning profiles throughout the reminder of their working career (Ghirelli, 2015).

In summary, the main findings on the impact of early nonemployment on future labor market outcomes are the following. First, both men and women suffer sizable earnings penalties, which are persistent up to 25 years after the secondary school diploma. Second, experiencing early nonemployment causes a lower participation in the labor market only in the short-term for both men and women. Our results on earnings are consistent with the ones in Gregg and Tominey (2005), where wage scars of about 9-11% persist up to 20 years later. Our findings on labor market participation are also in line with those in Nordström Skans (2011), who found the negative effect of early unemployment on the likelihood of unemployment 5 years after graduation. Finally, both are findings in terms of earning and labor market participation are similar to those in Mroz and Savage (2006), who estimated that the effect of unemployment on hourly earnings is long-lived, whereas only a short-lived persistence of about 4 years in terms of future unemployment was detected. As suggested by Ellwood (1982), early work experience may have a large and positive earnings effect and therefore the biggest costs of being nonemployed during the first years after school completion are wage penalties and lower earning power.

Our findings are not fully in line with the predictions of the signaling theory. Early nonemployment events may be used as a signal of low productivity and employers may penalize those individuals who experienced them (Spence, 1973; Vishwanath, 1989; Lockwood, 1991). However, individuals incurring in random early nonemployment events, once hired, will show greater productivity than expected and the initial penalties should disappear after a while. Only our findings on labor market participation are in line with the signaling theory. This is not the case in terms of earnings, because we find that the earnings penalties persist up to 25 years after school completion. A potential explanation of the persistent scars on earnings may come from the job search theory. Given that people experiencing early nonemployment send a worse signal, accumulate less human capital relatively to their employed peers, and are more likely to face liquidity constraints, they could lower their reservation wage and be more likely to accept worse jobs, characterized by a career track of lower profile, which traps them in lower wages and lower chances of subsequent promotions.

2.5.2 Sensitivity analysis

We ran some sensitivity checks to assess the robustness of our findings in several directions. We started by modifying the definition of nonemployment. In the benchmark model, experiences like volunteer work, internships and stages are considered as a form of employment and do not contribute to the computation of the fraction of days spent in nonemployment after the diploma. We modified this definition by considering as nonemployment also all the forms of unpaid work, for example volunteer work and unpaid internships, stages and training. Indeed, volunteer work, stages, internships and training are non-standard and so unstable positions in the labor market that one may wonder if they could be viewed as proper employment in terms of building a career, accumulating human capital, generating a network, etc. Table D2.1 in Appendix D displays the results, which are in line with the benchmark ones.

Second, we changed the definition of the treatment intensity by using, instead of the fraction of days spent in nonemployment in the first 3 years after school completion, the fraction of days spent in nonemployment during the first 2 or 4 years. The choice of measuring the intensity of early nonemployment by looking at the first 3 years after the diploma may indeed be viewed as arbitrary. Tables D2.2 and D2.3 display the effects of the fraction of days spent in nonemployment during the first 2 and 4 years after the diploma, respectively. They are in line with those obtained using the benchmark definition of treatment intensity. The only difference is that the penalties are somewhat: i) smaller if early nonemployment is computed in the first 2 years after the diploma; ii) larger if early nonemployment is defined in the first 4 years after school completion.

Third, we used different combinations of exclusion restrictions to test if they play a relevant role in determining the findings. For example, one may wonder whether geographical area or local labor market conditions at birth or just after school exit may, not only affect the predetermined outcomes (the measures) and early nonemployment, but also determine future labor market outcomes. In our baseline specification, as Table A2.5 clarifies, we indeed included these controls measured at birth in the measurement equations, measured just after school completion in the early nonemployment equation and measured at time t for the labor market equation at time t. These exclusion restrictions would not be supported by the data if, for instance, being born and growing up in more disadvantaged regions or in areas characterized by worse economic conditions increases future penalties in terms of labor market success, conditional on the current status of the economy and labor market. More in detail, we proceeded by checking the main findings with two different combinations of the exclusion restrictions: i) we included both the dummies for geographical area at birth and the regional employment, unemployment and GDP growth rates at birth in the labor market outcome equations and in the treatment equation; ii) we further added in the specification of the labor market equations also the regional rates in the first 3 years after school completion which, in the baseline model, are only included in the treatment equation. The findings from these alternative specifications are all in line with the benchmark results and are reported in Appendix D.

We ran a fourth check with the aim of understanding whether the findings are driven by cohort effects. We splitted the sample in individuals born in the 1960s and those born later (see Table A2.1 for summary statistics). For both groups the results are very similar to those obtained in the benchmark model and the main conclusions hold for both those born in the 1960s and those born later (see Tables D2.6 and D2.7 in Appendix D). However, the point estimates suggest that the latter suffered larger earning penalties. We also estimated the benchmark model using only those individuals we can follow up to 25 years after school completion. As shown in Table D2.8, even in this case the main results are confirmed.

A final check focused on the effect of the youth nonemployment across geographical areas. In particular, we splitted the sample between individuals born and graduated in Central or Northern Italy on the one hand, and individuals born and graduated in Southern Italy or Islands. Tables D2.9 and D2.10 in Appendix D show the results which are in line with the benchmark model, although the earning penalties in Central and Northern regions are larger than the ones in the South up to the first 10 years.

2.6 Conclusions

We studied the impact of nonemployment experienced during the first 3 years after school exit on labor market outcomes for Italian graduated. The effect is traced up to 25 years since school completion and evaluated in terms of yearly labor earnings and participation in the labor market by splitting the sample between males and females.

Using a factor analytic model, we were able to take into account time-varying unobserved heterogeneity jointly affecting selection into treatment and subsequent labor market outcomes. Once unobservables characteristics were accounted for, we found evidence that individuals in Italy who experienced nonemployment during the first 3 years after attained high school diploma suffer from relevant scarring effects. The negative effects are very persistent in terms of earnings: they are still sizable and statistically significant 25 years after school completion. Labor market participation, measured as the fraction of days spent at work in a year, is negatively affected by early nonemployment for a shorter span, as it disappears for both men and women by the 10th year after the school completion. Finally, the early nonemployment effect on labor market participation turns to be positive and significant 25 years after school completion, suggesting those who were exposed to early nonemployment in the long-run suffer smaller earnings and try to compensate with a larger participation in the labor market.

From a policy viewpoint, our findings suggest that favouring work experience after school completion is an urgent goal. This is a general and apparently obvious advice, which may be however complemented by a second peculiarity of our findings. The fact that earnings are persistently and negatively affected, while participation at the intensive margins is able to catch up after a bunch of years, suggests that those individuals who randomly experienced nonemployment after school completion were able to get reintegrated after a while, but in a downgraded track. Individuals suffering early nonemployment could have experienced the depreciation of their human capital (or they could have lost the opportunity to accumulate general human capital) and, under tighter liquidity constraints, could have been forced to lower their reservation wages and accept worse job conditions, limiting the transition to better career profiles. The policy maker could confine these negative consequences operating at different levels. First, the policy maker could favor training programs and apprenticeships for those who were exposed to early nonemployment, so as to facilitate their recoup of general human capital. For example, as shown by Picchio and Staffolani (2019), apprenticeships are effective ways for Italian workers to increase the probability of promotion to an open-ended contract. Second, the policy maker could intervene facilitating the match between employers and the youth who suffered early nonemployment, for example by *ad hoc* subsidies for hiring school-leavers with difficulties in making the school-to-work transition. Finally, to limit the lowering of the reservation wage and the acceptance of bad jobs in downgraded tracks, the welfare state could play a role: benefits and, to circumscribe moral hazard, monitoring job search behaviors, so as to guide the school leavers exposed to nonemployment towards more efficient and better quality job matches.

Chapter 3

Retirement and health outcomes in a meta-analytical framework

3.1 Introduction

In recent years, interest in the effects of retirement on workers' physical and mental health has grown considerably, becoming a topic of interest not only in the medical or psychological field, but also among labour and health economists. For the financial sustainability of the pension systems, in most of the OECD countries the standard retirement age has indeed increased and will continue to increase in the future (OECD, 2019). Understanding the health consequences of retirement is of utmost importance to provide policymakers with a clearer picture for the design of pension policies, labour market reforms, and healthcare investments that are welfare improving.

The identification of the causal health effects of retirement is the crux of this strand of research, and it involves methodological issues that are not easy to handle. Kuhn (2018) provides a clear non-technical summary of these methodological issues. First of all, estimation biases due to reverse causality may arise, because causality may not only run from retirement to health but is also likely to go from health to retirement decisions. Second, estimation biases may be due to measurement errors when researchers adopt subjective health measures as outcome variables. Indeed, the decision to retire early may influence the replies to the subjective answers of the interviewees, because they may assess their own health differently after retirement. This may happen for example because, when people retire, their reference group changes (Johnston and Lee, 2009). To deliver credible estimates of the causal impact of retirement on health, more recent studies have addressed endogeneity issues by means of different methodological strategies, especially using instrumental variables methods or regression discontinuity design (RDD).

Different identification strategies of the causal health effects of retirement may explain different estimates among studies. However, different findings are also explained by other reasons. For example, some recent reviews of the literature suggest that the heterogeneity in the estimated health effects of retirement depends also on the country or countries involved in the studies or the time span considered by the authors or covered by pension reforms. Furthermore, also the degree of freedom in choosing whether and when to retire matters: Bassanini and Caroli (2015), when reviewing the literature on the effect of working on health, found that both being forced to continue to work while one would like to retire and being forced to retire when one would prefer to continue working have similar adverse effects on health. They also found that voluntary retirement often has a positive effect on mental health. They concluded therefore that different findings among studies may be related to the voluntariness of the retirement decision.[1] Nishimura et al. (2018) investigated the source of differences among different studies by focusing on the methodological aspect and considering 8 recent papers in the economic literature. They concluded that the key factors in explaining different results are the choice of the estimation method and the countries surveyed. They also found that their results were not sensitive to the definition of retirement. van der Heide et al. (2013) summarized 22 longitudinal studies on the health effects of retirement, describing differences in terms of voluntary, involuntary, and regulatory retirement and between blue-collar and white-collar workers. While they found strong evidence for retirement having a positive effect on mental health, their review also revealed that contradictory findings emerge when the studies use perceived general health and physical health as outcome variables. Picchio and van Ours (2020) presented a selection of most recent studies focusing on differences in set-up, identification strategy, dependent variables, and heterogeneity of the retirement effects. Pilipiec et al. (2020) investigated the empirical evidence on the effects of increasing the retirement age on the health, well-being, and labour force participation of older workers. Focusing on 19 studies, they found that the evidence that an increase of the retirement age impacts on health and well-being is scant and inconclusive, because of the heterogeneity of the retirement effect among different groups of workers, and between

[1]To study the health effects of retirement, Bassanini and Caroli (2015) refer to 14 studies: 5 of them report negative effects of retirement on health.

workers far from retirement and older workers closer to the retirement age. Finally, Zulka et al. (2019) focused on the impact of retirement on cognitive functioning by using a sample of 20 studies. They suggested that different effects may be due to different types of prior occupation.

Although detailed, the aforementioned literature reviews focus on single aspects of a multifaceted phenomenon (Kuhn, 2018) and their concluding summaries may be deceptive (Stanley et al., 2013). According to Kuhn (2018), a meta-analysis, i.e. a research methodology used to bring together in a systematic way and with a quantitative perspective all the findings from previous studies on a given issue, has the potential to yield significant insights into the factors that trigger various health effects of retirement. To the best of our knowledge, only van Mourik (2020) has taken up this challenge and proposed a meta-analysis on the effects of retirement on several measures of health by collecting 576 results from 61 manuscripts. However, this meta-analysis did not comply with the guidelines of the Meta-Analysis of Economics Research Network (MAER-Net) (Stanley et al., 2013; Havránek et al., 2020). The analysis, in fact, was built on a trinomial outcome instead of effect sizes, revealing that 15% of the studies reported negative health effects of retirement, 35% positive health effects, and 50% statistically insignificant results. Furthermore, it includes not only articles published in scientific journals, but also working papers and Ph.D. dissertations. Also Sewdas et al. (2020) have provided a meta-analysis, but with a focus limited to the link between mortality and early and on-time retirement. Using a sample of 25 studies, they estimated a random-effects meta-regression to identify the effects of retirement and to assess the influence of gender, prior health, and demographics. They concluded that early retirement, compared to continued working, is not associated with a higher risk of mortality. However, on-time retirement, compared to continued working, is associated with a higher mortality risk, which may reflect the healthy worker effect, i.e. people in the group of those who work beyond the standard retirement age are on average healthier than those who retire on-time. Finally, both Pabón-Carrasco et al. (2020) and Li et al. (2021) only focus on depressive symptoms:[2] according to the former, the retirees with the highest prevalence of depression are those who retire in a mandatory fashion or due to illness; the latter show that the association of involuntary retirement with more depressive symptoms is stronger than voluntary or regulatory retirement, and it is more pronounced in Eastern developed countries.

[2] Pabón-Carrasco et al. (2020) collect a total of 11 articles, while Li et al. (2021) have a sample of 25 longitudinal studies.

A rigorous and extensive meta-analysis on the subject is lacking. The main contribution of our article is to fill this gap by means of a meta-analysis on the evidence of the health effects of retirement which i) follows the MAER-Net guidelines (Stanley et al., 2013; Havránek et al., 2020); ii) is based only on articles published in peer-reviewed journals, to reduce the probability that they contain mistakes (Xue et al., 2021), and in English, for the sake of accessibility (Vooren et al., 2019); iii) does not focus on a particular measure of health but instead considers the ones most frequently used in the literature, such as self-reported general health, physical and mental health, healthcare utilization, and mortality; and iv) focuses on studies published from 2000 onward in order to consider a more homogeneous labour market and pension policy background. Indeed, in most European countries the intensity of pension reforms has been particularly strong since the 2000s, with changes in eligibility criteria like the retirement age, the required contributory period, and the pension calculation scheme.[3] These changes have been implemented gradually and over long time periods. Thus, the increasing attention of policy-makers toward pension system reforms due to financial sustainability reasons and increasing workers' life expectancy after the mid-1990s has generated a large research interest among labour and health economists since the 2000s.

Our meta-analysis was carried out on 85 articles. It included the estimation of meta-regression models which enabled us to investigate the issue of publication bias and to look for patterns among different study characteristics after correcting the findings for it. We took into account all the main factors that might lead to different estimates of the effect sizes among studies, such as the institutional context, the research design, the causal effect identification strategy, and other study-related characteristics.

The rest of the paper is organized as follows. Section 3.2 focuses on the meta-analytical approach, describing the databases used, the research methods, and presenting preliminary and descriptive results of our meta-analysis. Section 3.3 assesses whether there is publication bias in this empirical literature. Section 3.4 provides heterogeneity analysis by using meta-regressions with the inclusion of covariates on the basis of Bayesian criteria for model selection. Section 3.5 concludes.

[3] Carone et al. (2016) report that the average number of pension measures per year in Europe was less than 10 during the late 1990s and rose to 44 between 2009 and 2014.

3.2 Meta-dataset

3.2.1 Search strategy and study selection criteria

The empirical literature does not report clear-cut results on the health effect of retirement. Several reasons may explain different findings: different methodologies of analysis, different identification strategies of the causal effect, different countries, different time spans considered by the studies or covered by pension reforms. Hence, a simple comparison among the different studies and of their results may be misleading (Stanley et al., 2013). A rigorous meta-analysis enabled us to systematically review the literature by combining the results of multiple and different studies so as to identify patterns among diverse study results while taking into account the uncertainty behind each point estimate of the relation of interest and remove bias induced by publication biases. Publication bias (also named 'file drawer problem') is the bias arising from the tendency of editors to prefer to publish findings consistent with the conventional view or with statistically significant results, while studies that find small or no significant effects tend to remain unpublished (Card and Krueger, 1995).

Our search for studies followed the MAER-Net guidelines (Havránek et al., 2020). These guidelines are an attempt to create a shared subjectivity in conducting meta-analyses in economics and thereby improve the transparency, replicability and quality of the reported results. We searched studies from November 2020 to March 2021 in Ideas/EconPapers, Google Scholar, Scopus and Web of Science by using the following keywords: 'retirement', 'health' and one among 'mental health', 'physical health', 'psychological well-being', 'healthcare' and 'mortality'. We only considered articles published in peer-reviewed journals of health economics, labour economics, social sciences, psychology, and medicine and with the SCImago Journal Rank (SJR) indicator.[4] We excluded theoretical works and studies concerning only cross-partner retirement effects of retiring (Atalay and Zhu, 2018; Bloemen et al., 2019), or general life satisfaction as dependent variable (Abolhassani and Alessie, 2013; Bender, 2012; Horner, 2014; Kesavayuth et al., 2016), or only health behaviour analysis (Evenson et al., 2002; Henkens et al., 2008; Zhao et al., 2017; Motegi et al., 2020).[5] Hence, we selected only micro-level studies on the health

[4]See www.scimagojr.com/SCImagoJournalRank.pdf for details on the calculation of the SJR. The following studies were not included in the final sample because their journals are not indexed in SCImago: Lee and Smith (2009), Fonseca et al. (2014), and Son et al. (2020).

[5]Drinking, smoking, and physical activity are examples of health behaviour outcomes.

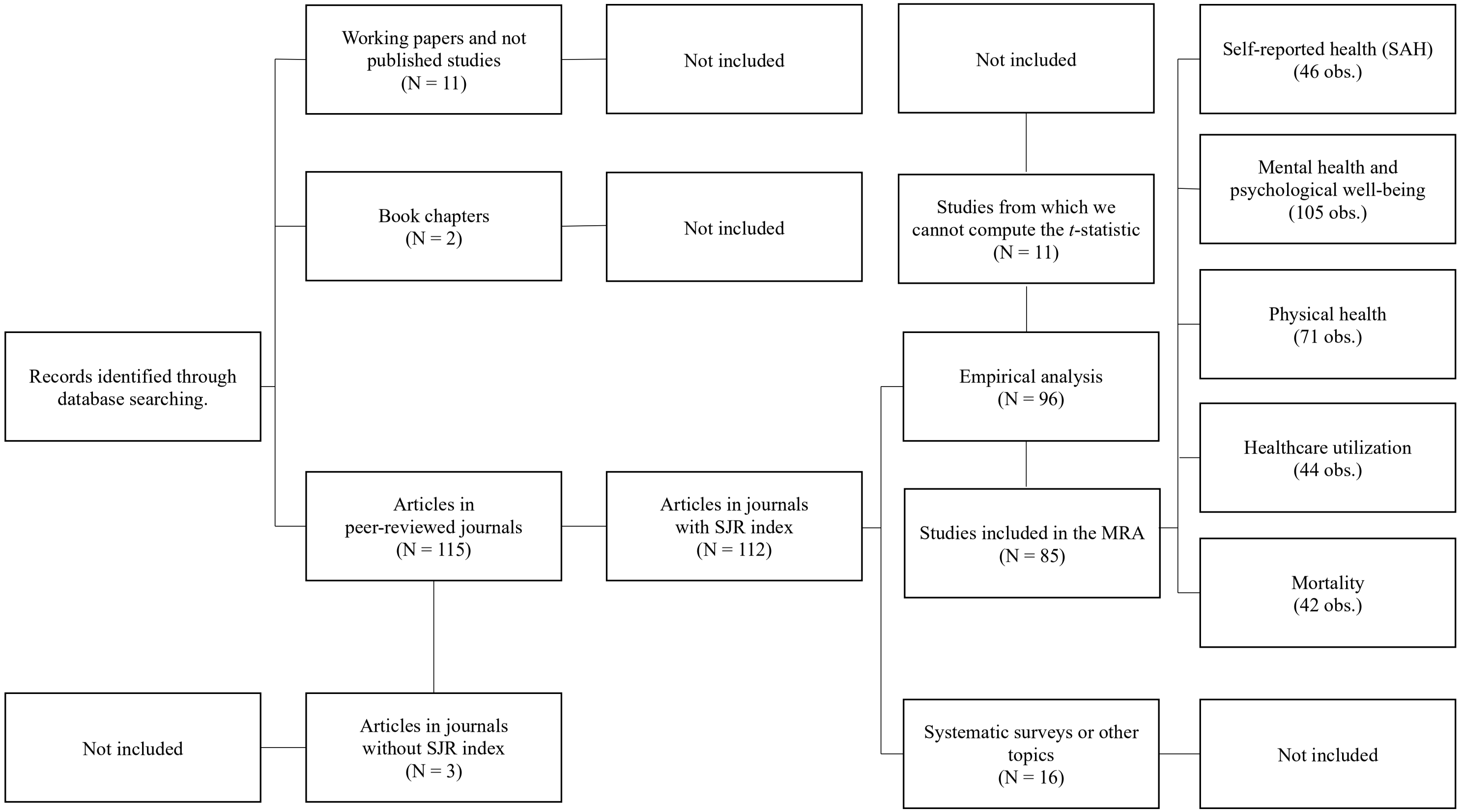
effects of retirement. We excluded 11 papers because they had not been published in peer-reviewed journals, i.e. discussion papers (see e.g. Waldron, 2001; Bound and Waidmann, 2007; Coe and Lindeboom, 2008; Lalive and Staubli, 2015; Zulkarnain and Rutledge, 2018) and two book chapters (Charles, 2004; Börsch-Supan and Schuth, 2014). At this point we had 96 articles. Finally, we had to remove 11 articles because they do not contain sufficient information with which to compute the t-statistic of the estimated retirement effect, on which we would build our meta-regressions.[6] Our final meta-analytic sample consisted of 85 articles, which are listed in Table A3.1 in the Appendix. Many studies dealt with the retirement effect on multiple health outcomes, and some others disaggregated the analysis by gender or by the type of previous occupation. In these cases, multiple data points were delivered and our final dataset consisted of 308 observations. Figure 3.1 is a PRISMA flow diagram (Moher et al., 2009): it graphically reports the rules we followed to include/exclude articles in our final sample.

From most of the articles, we directly extracted the estimated retirement effects ($\widehat{\beta}_i$) along which their standard errors ($SE_i(\widehat{\beta}_i)$) and computed the t-statistics as their ratio. In other cases, we could directly retrieve the t-statistics because they were reported among the study results. Finally, in some studies only the estimated effects and their 95% confidence intervals were displayed. In these cases, we approximated the standard errors in linear models (and then we computed the t-statistics) as follows: $SE_i = (ub - lb)/(2 \times 1.96)$, where ub and lb are the upper bound and the lower bound of the confidence interval, respectively. For studies with non-linear models, such as multinomial logit or Cox proportional hazard models, and reporting only the odds ratio (OR) and its 95% confidence interval, we calculated the standard error as $SE_i = [\ln(ub) - \ln(lb)]/(2 \times 1.96)$ and then the t-statistic as $t_i = [ln(\widehat{\beta}_{1i})\widehat{\beta}_{1i}]/SE_i$.

The health outcomes were quite different among, and sometimes within, studies. In some cases, when the sign of the coefficient of retirement was positive, this meant that there was a health improvement, like for general physical health indexes or self-assessed health. In some other cases, it was the negative sign that implied a health improvement, such as when mortality or depression were the health outcomes. We altered the sign of the t-statistics so that a "positive" ("negative") sign means a health improvement (deterioration), and all the rest of our analysis is based on this modification of the t-statistics.

[6] These 11 articles are: Allen and Alpass (2020), Barban et al. (2020), Carlsson et al. (2012), Dufouil et al. (2014), Finkel et al. (2009), Fisher et al. (2014), Kühntopf and Tivig (2012), Mazzonna and Peracchi (2012), Nishimura et al. (2018), Olesen et al. (2014), Rohwedder and Willis (2010).

Figure 3.1: PRISMA flow diagram



Graph a) of Figure 3.2 shows the distribution of t-statistics, which is quite dispersed, with a minimum of -15.66, a maximum of 14.70, and a standard deviation of 3.13. Most of the findings (60.4%, 186 outcomes) are not significantly different from 0, having a t-statistic smaller than 1.96 in absolute value; in 27.9% (11.7%) of the cases, 86 (36) results, the retirement effect on health is instead significantly positive (negative). Graph b) of Figure 3.2 plots the distribution of the square root of the observations used to estimate the retirement effects. The number of observations is also very heterogeneous, with a minimum of 49 and a maximum of 1,866,974. Since in what follows the t-statistics and the number of observations would then be used to build a comparable measure of the estimated effect across different studies, the presence of extreme values in these two key variables raised concerns about outliers, especially because the linear models typically used in meta-regressions may be particularly sensitive to them (Viechtbauer and Cheung, 2010). As suggested by Xue et al. (2021), who had a similar problem when conducting a meta-analysis on the education effect on health, we moderated the problem by winsorization of t-statistics and number of observations at the top and bottom of their distribution: we replaced values that were lower (larger) than the 5th (95th) percentile with the value of the 5th (95th) percentile.[7]

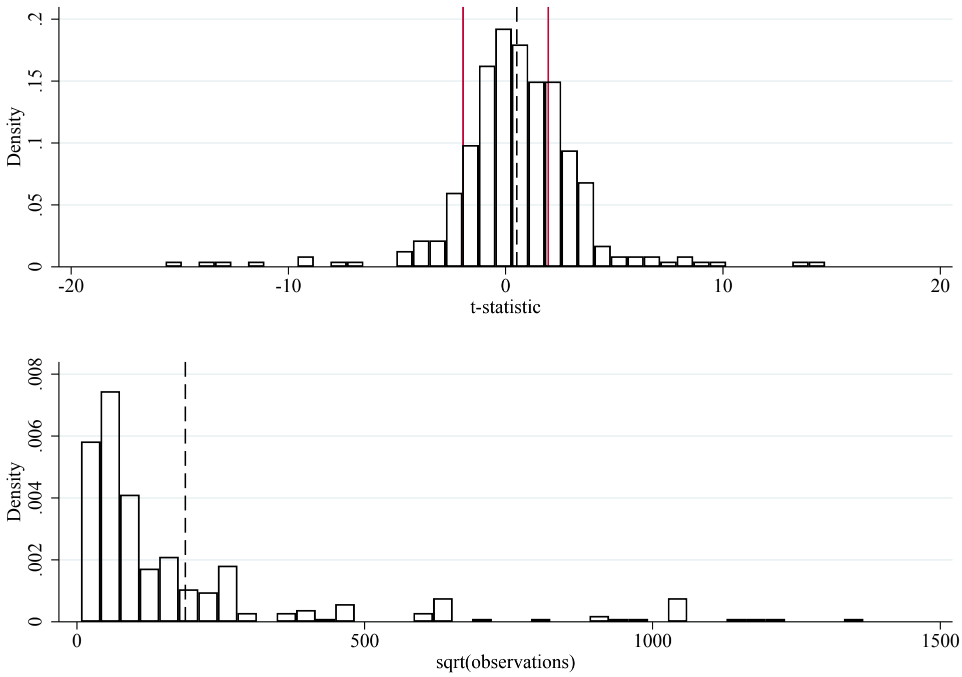
3.2.2 Descriptive statistics

We provide some basic descriptive statistics of our meta-analytic sample by research findings. Table 3.1 reports summary statistics by research outcomes[8] of those covariates that we used in the meta-regressions to capture the factors underlying the heterogeneous effects in the empirical literature: journal subject area, the number of citations on average per year (retrieved from Google Scholar on 05/04/2021), the journal SJR indicator at the time of publication, publication year, identification strategy, gender, institutional context, geographical area, type of previous occupation, birth cohort, and the way in which the t-statistic was calculated. We considered three subject areas according to the Scimago classification: i) Economics, Econometrics and Finance or Business, Accounting and Management (28.6% of our observations); ii) Medicine or Psychology (43.2% of the observations); iii) a residual category containing journals belonging to multiple subject

[7]We replicated the empirical analysis without winsorization as a sensitivity analysis. Our findings were unchanged. We report estimation results without winsorization in the Appendix.

[8]In Table A3.2 in the Appendix, we report similar summary statistics by the sign of the relation between retirement and health.

Figure 3.2: Distribution of t-statistics and observations of study outcomes



Notes: The number of study results is 308. The dashed vertical lines are the sample average of t-statistics in the upper graph (0.508) and of the square root of observations in the lower graph (188.23). The solid vertical lines in the upper graph denote the critical values for the 5% significance level in two-tailed tests (± 1.96).

areas (28.2% of the observations).[9]

The average number of yearly citations was the smallest (9.3) when the null hypothesis of no effect could not be rejected. It was instead the highest (20.1) when significant negative effects emerged and almost twice as large as the average number of yearly citations of findings supporting significant positive effects (11.5). Differences in the scientific influence of the journals where the articles had been published were smaller. In both cases, articles finding negative outcomes displayed a larger standard deviation. It is noteworthy that statistically insignificant results were not under-represented in journals of high scientific influence compared to those with more clear-cut findings; rather, they corresponded to almost 60% of our sample. This might suggest that, at a first and very descriptive level, publication bias is not an issue in this research strand.

Since health is a multidimensional concept, we referred to the main measures analysed in the empirical literature. Among the particular health measures evaluated, positive effect had the largest absolute frequency when we focused on general or self-assessed health. In all the other cases, no statistically significant effect was the prevailing outcome. These various health measures were physical health (23.1%), mental health (34.1%),[10] health-care utilization, which included doctor visits and hospitalization (14.3%), and mortality (13.6%).

Identifying the causal effect of retirement on health is not straightforward because there are several sources of potential endogeneity of the retirement decision, such as reverse causality, unobserved heterogeneity,[11] and measurement error.[12] These could affect not only the magnitude but also the sign of the estimated effect. Hence, we used a set of indicators to control for the methodology employed to identify and estimate the im-

[9]This category also comprises 2 observations by Kalwij et al. (2013), the only article in our sample published in a social-sciences journal.

[10]Physical health included chronic conditions, mobility, body mass index (BMI), activities of daily living (ADL) and a measure of general physical status. Mental health consisted of cognitive functioning, depression or anxiety, and a more general measure which comprised general mental health index and psychological well-being (in this case, it also comprised happiness as a proxy for well-being).

[11]Omitted variables biases might be induced by differences in unobserved individual characteristics that influence both health and retirement decisions (e.g. subjective life expectancy). Unobserved heterogeneity could be time-constant but also time-varying. To control for unobserved time-constant individual heterogeneity, researchers typically use individual fixed-effects panel data models (Eibich, 2015).

[12]Self-reported health measures are at risk of two kinds of measurement error: i) self-assessed health may not be comparable across individuals ("classical measurement error"); ii) individuals who do not work may justify their labour market status by their ill health ("justification bias"). The latter refers to retirees' tendencies to exaggerate their poor health conditions in order to provide socially acceptable justification for their retirement and observed health would be understated for retirees (Behncke, 2012; Insler, 2014).

Table 3.1: Descriptive statistics of explanatory variables used in the meta-regressions

	Negative effect			Null effect			Positive effect		
	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.
Scimago subject areas									
Multi area (Reference category)	9	0.250	0.439	53	0.285	0.453	25	0.291	0.457
Economics/Business	12	0.333	0.478	50	0.269	0.445	26	0.302	0.462
Medicine/Psychology	15	0.417	0.500	83	0.446	0.498	35	0.407	0.494
Health outcomes									
Mortality (Reference category)	8	0.222	0.422	32	0.172	0.378	2	0.023	0.152
General and self-reported health	5	0.139	0.351	17	0.091	0.289	24	0.279	0.451
Physical health	10	0.278	0.454	47	0.253	0.436	14	0.163	0.371
Mental health	12	0.333	0.478	60	0.323	0.469	33	0.384	0.489
Healthcare utilization	1	0.028	0.167	30	0.161	0.369	13	0.151	0.360
Identification strategies									
Other methods (Reference category)	4	0.111	0.319	21	0.113	0.317	13	0.151	0.360
Regression discontinuity design (RDD)	6	0.167	0.378	33	0.177	0.383	20	0.233	0.425
Instrumental variables (IV)	16	0.444	0.504	89	0.478	0.501	47	0.547	0.501
Difference-in-differences (DiD)	–	–	–	18	0.097	0.296	5	0.058	0.235
Propensity score matching (PSM)	3	0.083	0.280	14	0.075	0.265	–	–	–
Fixed-effects/First-differences	7	0.194	0.401	11	0.059	0.237	1	0.012	0.108
Institutional contexts									
Statutory retirement (Reference category)	22	0.611	0.494	110	0.591	0.493	67	0.779	0.417
Mandatory or involuntary retirement	8	0.222	0.422	17	0.091	0.289	7	0.081	0.275
Early retirement	5	0.139	0.351	36	0.194	0.396	8	0.093	0.292
Postponed retirement	1	0.028	0.167	23	0.124	0.330	4	0.779	0.417
Geographical areas									
Multi-country analyses (Reference category)	4	0.111	0.319	25	0.134	0.342	12	0.140	0.349
Europe	11	0.306	0.467	92	0.495	0.501	40	0.465	0.502
Extra-European countries	21	0.583	0.500	69	0.371	0.484	34	0.395	0.492
Sex									
Males (Reference category)	12	0.333	0.478	59	0.317	0.467	31	0.360	0.483
Females	7	0.194	0.401	62	0.334	0.473	24	0.279	0.451
Males+Females	17	0.472	0.506	65	0.349	0.478	31	0.360	0.483
Calculation of t-statistic									
from 95% CI or from OR (Reference category)	4	0.111	0.319	20	0.108	0.311	16	0.186	0.391
t-statistic from $\widehat{\beta}_i/SE_i$	32	0.889	0.319	166	0.892	0.311	70	0.814	0.391
Birth cohorts									
Other (Reference category)	28	0.777	0.422	121	0.651	0.478	63	0.733	0.445
Only birth cohorts $\leq$ 1950	8	0.222	0.422	65	0.349	0.478	23	0.267	0.445
Type of previous occupation									
White collar (Reference category)	–	–	–	13	0.070	0.256	3	0.035	0.185
Blue collar	2	0.056	0.232	16	0.086	0.281	6	0.070	0.256
Not specified	34	0.944	0.232	157	0.844	0.364	77	0.895	0.308
Study-related characteristics									
Google Scholar citations per year	36	20.104	14.078	186	9.339	9.225	86	11.534	10.314
Scimago Journal Ranking	36	2.210	1.938	186	1.771	1.082	86	1.757	1.185
Year of publication	36	2012.861	5.117	186	2015.962	4.387	86	2015.174	4.671
Observations		36			186			86	

Notes: Females+Males = observations for which authors do not separate estimates for men and women. Other methods = OLS regressions and non-linear models (logit, multinomial logit, ordered probit and Cox proportional hazard models). The sign of the effect is based on the value of t-stat: "negative" means $t \leq -1.96$; "positive" is for $t \geq 1.96$; "null" when $-1.96 < t < 1.96$. When articles found that postponed retirement had a negative effect, we labeled the effect of retirement as "positive".

[a] At the time of publication, some journals did not yet have the SJR index, either because they had been published in too recent years or because the journal was not yet indexed in Scimago. In these cases, we assigned to the journal the available value of the SJR index which was chronologically closer.

pact of retirement on health. The instrumental variables (IV) method was the one used most frequently (49.4%), followed by regression discontinuity design (RDD) (19.2%). The difference-in-differences (DiD) estimator was mostly used to evaluate policy reforms and represented 7.5% of our observations. In 12.3% of the study results, no particular methods was used to tackle the endogeneity of the retirement decision (e.g. linear model, multinomial logit or Cox proportional hazard models).

Some indicator variables were used to capture the institutional context and, in particular, the retirement scheme. The survey of the empirical literature provided by Bassanini and Caroli (2015) highlights the role played by choice vs. constraint in shaping the health impact of work and retirement. They focus on that strand of the literature which studies the voluntariness of retirement and from which evidence of adverse health effects arises when individuals are forced to stop working. In our analysis, we considered both the voluntariness of the retirement decisions and its timing: we distinguished among early (15.9%), postponed (9.1%), mandatory or involuntary (10.4%), and statutory retirement, i.e., retiring at the standard retirement age (64.6%).

A further control variable is the gender associated with the estimated effect. The retirement effects may be different for men and women, for example because the career trajectory and the involvement in the labour market are typically different by gender. We also controlled for the geographical areas. In particular, we considered results for Europe (46.4%), for extra-European countries (40.3%), and from multi-country analyses (13.3%).

The health effects of retirement could be associated with the birth cohort because working conditions and the attention to occupational health changed during the 20th century, impacting on the physical and mental stress at work (Cullen, 1999; Harrison and Dawson, 2016). We coded the birth cohort using two dummy indicators: a dummy equal to one if the result came from individuals who had all been born before 1950 (31.2%); a dummy equal to one for results not specifying the birth cohort or covering both the period before and after 1950 (68.8%).

For similar reasons, the health effects of retirement may depend on the kind of occupation. Although very few studies provide separate estimates related to the type of previous occupation, we distinguished between blue-collar (7.8%) and white-collar workers (5.2%), and we grouped in a residual category all the other results which did not distinguish between the types of occupation (87%).

Finally, we also controlled for the method used to calculate the t-statistics. 87% of our observations were based on t-statistics derived from the ratio between $\widehat{\beta}_i$ and the corresponding standard error. The remaining 13% were derived from 95% confidence intervals or starting from odds ratios (OR).

3.2.3 Comparable effect sizes

The estimated retirement effects on health $\widehat{\beta}_i$ are not easily comparable across the models used by the studies surveyed and the estimation techniques generating them. In this regard, we observed a large heterogeneity in the health measures used as outcome variables. For example, those most frequently used were self-reported general health, physical health indexes, like the body mass index (BMI) or the activities of daily living (ADL), mental health measures, like depression or the 5-item mental health inventory (MHI-5), healthcare utilization, and mortality. The units of measurement used by the studies were therefore not comparable. Moreover, even when a similar health outcome was used across studies, different model specifications and/or different estimation methods could alter their comparability. For example, although most of the estimated models were linear, in some cases nonlinear models, like multinomial logit or Cox proportional hazard models, were estimated.

To make the effect estimates comparable, we computed the partial correlation coefficient r_i, which has been commonly used in meta-analyses in economics, business and social sciences since Doucouliagos (1995). The partial correlation coefficient is a measure of the association between two variables, keeping other covariates constant. It is independent of the metrics with which the dependent and the independent variables are measured (Ugur, 2014). Very recent examples are Churchill and Mishra (2018) and Xue et al. (2021), who used the partial correlation coefficient in reviewing returns to education on the labour market and on health, respectively. Xue et al. (2020), in their meta-analysis on the health effects of social capital, used the partial correlation coefficient as a way to combine estimated effects that were not comparable because of different measures of health used and different types of econometric models estimated, as in our framework.[13]

The partial correlation coefficient is computed as

$$r_i = \frac{t_i}{\sqrt{t_i^2 + dk_i}}, \tag{3.1}$$

[13]See Reed (2020) and the meta-analyses cited therein for other examples of meta-analyses using the partial correlation coefficient as the effect size.

where dk_i is the degrees of freedom in the model from which the i-th t-statistic is derived. Keef and Roberts (2004) show that the estimate of r_i contains a small positive bias, since it increases as the number of independent variables in the regression model increases, i.e. as the degrees of freedom decrease. However, asymptotically this bias disappears. Moreover, in our meta-dataset many studies did not provide precise information about the number of covariates. Consequently, we could not recover the degrees of freedom. When this was the case, we approximated dk_i with the number of observations (minus 2).[14] Because the smallest number of observations, after the aforementioned winsorization, was 523, this approximation generated a very mild upward bias which asymptotically disappeared. The standard error of the partial correlation coefficient is given by

$$SE(r_i) = \sqrt{\frac{1 - r_i^2}{dk_i}}. \tag{3.2}$$

It can be shown that $r_i / SE(r_i) = t_i$.

The partial correlation coefficient r is a unitless measure, which takes a value between -1 and 1. It enables direct comparisons among the different ways to approach and measure health outcomes in the empirical literature and in the diverse literatures (Doucouliagos and Laroche, 2009). The partial correlation coefficient drops as the degrees of freedom or the sample size increase. This implies that nearly similar t-statistics will produce very different partial correlations if the sample sizes are diverse: the larger the sample size, the more the effect size measured by the partial correlation is scaled down.

Table 3.2 displays summary statistics of partial correlations, t-statistics, and number of observations of the full sample and of the results by the type of health measure. As in Xue et al. (2020) and Xue et al. (2021), we included in our meta-analysis different types of health measures, ranging from physical health, mental health, self-reported general health and healthcare service utilization. One may wonder whether we mixed together outcomes which measured too diverse phenomena. On the one hand, one of the aims of Section 3.4 is to understand if such heterogeneity is related to the findings, and this source of diversity was explicitly taken into account in the meta-regression analysis. On the other hand, we will do it partially, because in the specification of the meta-regression models we imposed that the impact of all the other covariates was not a function of the particular measure of health. Dividing the sample into as many subsamples as the five different measures of

[14] See Table B.11 in Lipsey and Wilson (2001).

health would result in small sample size problems for some of them.

The graph in Figure 3.3, known as funnel plot (Light and Pillemer, 1984), shows the scatter plot of the partial correlation coefficient and its precision, measured by the inverse of its standard error as defined in Equation (3.2). In the absence of publication bias, the partial correlation coefficient should vary randomly around its average, which is an estimate of the true effect. Hence, the symmetry of the funnel around the average effect is of help in graphically visualizing a possible publication bias (Stanley, 2005). The funnel plot shows a mild asymmetry, given the longer tail to the right of the average partial correlation coefficient. It is not easy to reach a conclusion about publication bias by means of this graphical approach. Indeed, it relies on the assumption that there is a single 'true' effect common to all empirical studies. Hence, if there is heterogeneity among articles due to different datasets, time spans, countries or methodologies, it may cause the funnel's skewness. In this case, the funnel plot seems to suggest that there is not an evident publication bias. However, in the next section, on the basis of Meta-Regression Analysis (MRA), we will formally test for the presence of publication bias.

Table 3.2: Summary statistics of partial correlations, t-statistics, and number of observations by type of health outcome

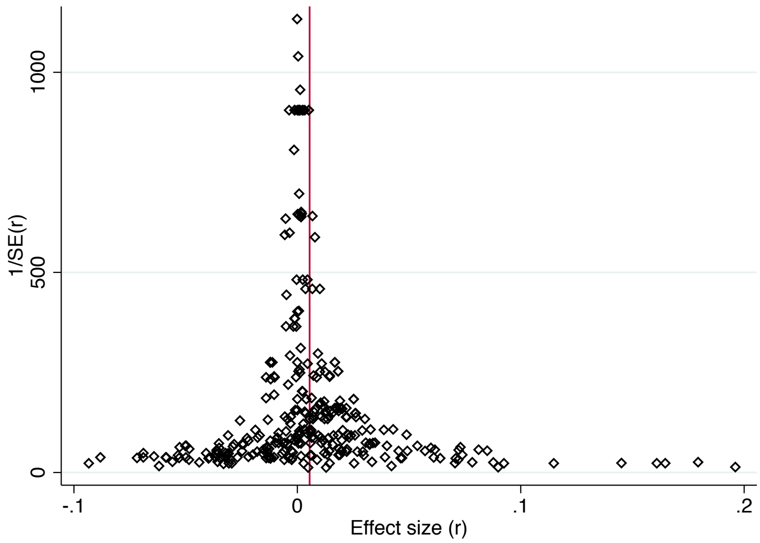
Outcome variables used as health measures	Number of studies	Number of results	Relative frequency of results (%)	Average partial correlation (r)	Average t-statistic[a]	Average sample size[a]
Mental health	47	105	34.1	0.0095	0.8611	12,568
Physical health	30	71	23.1	0.0069	0.0334	47,394
General and self-reported health	32	46	14.9	0.0091	1.0967	17,178
Healthcare utilization	15	44	14.3	-0.0048	0.6342	289,704
Mortality	19	42	13.7	0.0004	-0.5500	290,393
Total	85[b]	308	100.0	0.0055	0.4807	98,761

[a] These averages are computed before the winsorization.
[b] This amount is not the sum of the absolute frequencies reported in this column, because the same study could have focused on multiple health dimensions and therefore could count in multiple lines of the same column.

3.3 Testing for publication bias

To formally assess the relevance of publication bias and to eventually remove it from the estimate of the genuine retirement effect on health, we used the "Funnel Asymmetry Test – Precision Effect Test" (FAT-PET) (Egger et al., 1997; Stanley, 2005, 2008), which is a standard model to assess the presence of publication bias. Used since the end of the 1990s

Figure 3.3: Funnel plot of precision ($1/SE(r)$) versus effect size (r)



Notes: The number of observations is 308. The vertical line is the average of the partial correlation coefficients r (0.0055).

in the economic literature (Card and Krueger, 1995; Ashenfelter et al., 1999; Görg and Strobl, 2001), it is based on a simple regression of the i-th effect size on a constant and its standard error:

$$r_i = \gamma_1 + \gamma_0 SE(r_i) + \varepsilon_i, \tag{3.3}$$

where ε_i is the idiosyncratic error terms and γ_0 will be equal to zero when the effect size r_i varies randomly around the precision effect γ_1, meaning no publication bias. Publication bias is proportional to the inverse of the square root of the sample size, which in turn is proportional to the standard error (Begg and Berlin, 1988). The Funnel Asymmetry Test (FAT) tests the hypothesis of no publication bias (Egger et al., 1997), i.e. $H_0 : \gamma_0 = 0$, and is therefore also a test of funnel asymmetry (Sutton et al., 2000). If the null hypothesis is rejected, a publication bias is affecting this strand of the literature, potentially posing a serious problem for interpretation of the scientific research (Begg and Berlin, 1988). The Precision Effect Test (PET) tests the null hypothesis $H_0 : \gamma_1 = 0$. The rejection of the null hypothesis can be interpreted as the presence of an authentic empirical effect, corrected for publication selection: when the sample size goes to infinity and the standard error goes to 0, the observed effects goes to γ_1 (Stanley, 2008).

Table 3.3 displays the results of different estimation and specifications of Equation (3.3). Model (1) reports the ordinary least squares (OLS) estimates of Equation (3.3), without taking advantage of the known form of heteroskedasticity affecting the distribution of r_i, as seen in Equation (3.2). This knowledge is instead exploited in Model (2), which displays the results when Equation (3.3) is estimated by Weighted Least Squares (WLS-FE) using $1/SE(r_i)^2$ as weights. Models (3) and (4) are robustness checks. In Model (3) we replicate our simple FAT-PET estimates by replacing $SE(r_i)$ with the inverse of the square root of the sample size as an alternative precision measure. Because the sample size is not subject to estimation error, it avoids an errors-in-variables bias that could instead affect $SE(r_i)$. If $SE(r_i)$ is endogenous in Models (1) and (2) because it is affected by measurement error, we may solve the problem by using an IV approach, instrumenting $SE(r_i)$ with the square root of the number of observations, which is strongly correlated to the standard error but should not be able to explain the estimated effect once we control for the standard error. This is called the Funnel Asymmetry Instrumental Variable Estimator (FAIVE) by Stanley (2005). Finally, in Model (5) we report the results if in Equation (3.3) we replace $SE(r_i)$ with its square to capture eventual non-linearities: this is the Precision Effect Estimate with Standard Error (PEESE) model, which is a meta-regression method to be preferred in correcting for publication bias when there is a genuine nonzero effect (Stanley and Doucouliagos, 2012, 2014).

From the five models reported in Table 3.3, we find weak evidence of publication bias only in the FAT-PET model estimated by WLS-FE. Furthermore, the FAT-PET point estimates of γ_0, ranging from 0.282 to 0.487, suggest that, if it exists, the publication bias is positive and small.

The precision coefficient is equal to 0.001 and significant only in the PEESE model. Hence, the mean effect of retirement on health is positive. However, it is extremely low, considering that, according to Cohen (1988), a partial correlation coefficient of 0.1 is to be considered as "small", and in the analysis of Doucouliagos (2011), who focused on economic results, it should be at least 0.07 to be considered as "small".[15]

The recognition of publication bias as a threat to the reliability of the scientific knowledge has taken place at different times in different disciplines. For example, psychological and medical research has acknowledged it since the end of the 1950s (Sterling, 1959; Rosenthal, 1979; Begg and Berlin, 1988). The economic research has taken instead some more years, until the 1990s (see e.g. Card and Krueger, 1995; Ashenfelter

[15] In Doucouliagos (2011), 0.17 is the threshold for "moderate" and 0.33 for "large".

Table 3.3: FAT-PET and PEESE tests and corrections for publication bias

	FAT-PET								PEESE[c]	
	(1) OLS		(2) WLS-FE		(3) WLS-FE[a]		(4) FAIVE[b]		(5) WLS-FE	
Publication bias	0.487	(0.384)	0.409*	(0.227)	0.414*	(0.233)	0.282	(0.269)	10.145	(7.768)
Precision effect	-0.002	(0.004)	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)	0.001**	(0.001)
R^2	0.034		0.023		0.023		0.028		0.009	

Standard errors robust heteroskedasticity and within-study correlation are in parentheses. The number of observations (studies) is 308 (85).
[a] The inverse of the square root of the sample size is used instead of $SE(r_i)$ as a precision measure.
[b] The F-statistic for the power of the excluded instrument is 32.97.
[c] PEESE gives a less biased correction for publication bias (Stanley and Doucouliagos, 2012, 2014).

et al., 1999). Therefore, one might wonder whether researchers and journal editors have different sensitivities to the problem across different disciplines, resulting in publication bias being limited only to some disciplines. To check whether this might be the case, we distinguished the study results into three broad subject areas: medicine/psychology, economics/business, and a residual category. Then, we generalized Equation (3.3) by having one constant per each subject area and the standard error interacted with the subject area indicator. We found that publication bias does not arise in any of the separate subject areas. We report the tests for publication bias by subject area in the Appendix.

To sum up, after a battery of tests, we concluded that publication bias is not importantly affecting this strand of the literature, and that the mean effect of retirement on health is positive but very close to zero. The next meta-regressions reported were aimed at understanding possible heterogeneity among studies in the retirement effect on health. We kept the PEESE specification as the benchmark model, so as to correct for publication bias when multiple covariates were included in the model specification.

3.4 Multiple meta-regressions

To detect possible sources of heterogeneous effects of retirement on health, we included in the PEESE specification a series of covariates: measures of health, methods to identify the effect, institutional contexts, geographical areas, gender, year of publication, SJR index, the average number of Google scholar citations per year, type of previous occupation, birth cohorts, and the way in which we derived the t-statistics. We employed the PEESE specification because its quadratic form of the standard errors has been proven to be less biased and often more efficient than the FAT-PET specification when there is a nonzero genuine effect (Stanley and Doucouliagos, 2014).[16]

Formally, we estimated by WLS-FE the following equation for our effect size

$$r_i = \gamma_1 + \gamma_0 SE(r_i)^2 + \boldsymbol{\beta}_1 \mathbf{x}_i + \varepsilon_i, \tag{3.4}$$

which is equivalent to estimating by OLS the transformed model

$$\frac{r_i}{SE(r_i)} = \gamma_1 \frac{1}{SE(r_i)} + \gamma_0 SE(r_i) + \boldsymbol{\beta}_1 \frac{\mathbf{x}_i}{SE(r_i)} + \frac{\varepsilon_i}{SE(r_i)}, \tag{3.5}$$

where $\mathbf{x}_i$ is the vector of result characteristics.

A problem in estimating Equation (3.5) is related to the model uncertainty about which variables should be included. We overcame it by employing one of the most commonly used tools in meta-analysis, Bayesian Model Averaging (BMA). BMA takes into account all possible models by running many regressions with different subsets of control variables and computing the weighted averages of the estimated coefficients. The weights are Posterior Model Probabilities (PMP) and are related to the goodness of fit of each model. The sum of PMPs indicates the Posterior Inclusion Probability (PIP) for each regressor, which provides the information on the likelihood of the regressor belonging to the true specification. A PIP above 0.5 for a given regressor is usually used as a rule of thumb to include it in the final model (Eicher et al., 2011). For each covariate, BMA returns the posterior coefficient distribution, which yields the posterior mean (PM) of the regression coefficient and the posterior standard deviation (PSD).

We used the BMA estimator discussed by Magnus et al. (2010), who introduced the distinction between two subsets of explanatory variables. The first subset is the set of "focus" regressors, which are those wanted in the model for theoretical (or other) reasons. In our case, the focus variables were those capturing the publication bias and the precision effect. The second subset is the set of "auxiliary" regressors, which are additional covariates that may be relevant to explaining the estimated effect, but this is not certain. Since we had 25 auxiliary covariates, the number of possible models to be considered was 2^{25}. BMA proceeds by applying conventional non-informative priors on the focus variables and the error variance σ^2, and an informative multivariate Gaussian prior on the auxiliary variables.

[16] Table A3.3 in the Appendix displays the results of the FAT-PET specification. The results are very similar to the ones from the PEESE model.

In a subsequent step, we performed a model-average procedure by using the Weighted Average Least Squares (WALS) (Magnus et al., 2010). WALS occupies an intermediate position between the Bayesian approach of BMA and the frequentist model-averaging procedure. In fact, it is a Bayesian combination of frequentist estimators (Magnus and De Luca, 2016). WALS uses conventional non-informative priors on the focus regressors and the error variance σ^2 and a distribution with zero mean for the independent and identically distributed elements of the t-ratios associated with linear combinations of the auxiliary regressors.[17] Unlike BMA, WALS relies on preliminary orthogonal transformations of the auxiliary regressors and their parameters, which reduce the computational burden from 2^{25} to 25. For this reason, WALS does not allow computation of the PIPs. An auxiliary covariate is considered to be robustly correlated with the outcome variable if the t-ratio of its coefficient is greater than 1 in absolute value or, equivalently, if the corresponding one-standard error band does not include zero (De Luca and Magnus, 2011). The advantage of WALS over BMA is that it does not impose an *ad hoc* assumption on the prior on the model space (in general BMA uses a uniform prior assigning equal probability to each model), but it is theoretically based (Magnus and De Luca, 2016).

Finally, like Havranek et al. (2015) and Xue et al. (2021), we conducted a frequentist check by estimating Equation (3.5) by OLS after restricting the set of regressors to those with PIP > 0.5 according to BMA. We ran the same frequentist check after the WALS estimates.

Table 3.4 reports the estimation results. For the BMA, we show the estimated posterior means, the posterior standard deviations, and the posterior inclusion probabilities of each regressor. For the WALS, we include the results deriving from two different assumptions about the model prior distributions. In the last columns of Table 3.4, we present the findings of the frequentist checks.

As regards the focus regressors, whilst for these variables the Posterior Inclusion Probabilities from BMA model are not informative, OLS estimates reveal no publication bias, even after controlling for a set of covariates. According to BMA results, there are 6 auxiliary covariates which are significant in explaining the heterogeneous effects of retirement on health (PIP > 0.5): two measures of health outcomes, fixed-effects/first-difference estimator, mandatory or involuntary retirement, year of publication and the dummy for the birth cohort. WALS results are quite similar, although some further covariates seem to

[17] The prior distribution of the t ratios can be either a neutral Laplace prior (Magnus et al., 2010), or a neutral Subbotin prior distribution (Einmahl et al., 2011).

be important: physical health and healthcare utilization, postponed retirement, the SJR indicator, estimates not distinguishing between males and females, RD design, and PSM estimator.

All the models reveal that the studies which used general and self-reported health indicators or mental health measures were the ones most likely to report positive effects of retirement on health. The analyses focusing on physical health or healthcare utilization were more likely to find positive effects than those dealing with mortality, although the difference in terms of correlation points was negligible. These findings reflect the results of some earlier systematic surveys in this field: as pointed out by Bassanini and Caroli (2015) or suggested by Nishimura et al. (2018) after re-estimating previous analyses, most of the evidence concerning the health effects of retirement shifts towards a positive impact on physical and mental dimensions of health, better self-assessed health, and lower healthcare utilization.

The results for the identification strategy suggest that the heterogeneity across this dimension is not particularly important in explaining different findings. We find that only those studies using a fixed-effects or a first-differences approach are more likely to report negative effects on health. This finding contrasts with the one reported by Nishimura et al. (2018), who instead showed that the choice of the estimation strategy is one of the key factors in explaining why the estimated results of the retirement effect on health differ.

An important factor in explaining heterogeneous estimated effects of retirement on health is the institutional context and the retirement scheme: mandatory or involuntary retirement has a PIP close to 1 and the greatest negative effect in magnitude. In the WALS results and, although with a lower magnitude, studies focusing on postponed retirement are also associated with a lower chance of detecting positive retirement effects than are studies dealing with early or statutory retirement. These findings confirm the conclusions of Bassanini and Caroli (2015), who showed that being forced to work while preferring to retire and, symmetrically, being forced to stop working because employees have no control on the retirement and work decisions have a health damaging effect. Similar results are reported by Pabón-Carrasco et al. (2020) and Li et al. (2021), but only on the effects on depressive symptoms. Moreover, the negative impact of postponed retirement on health, compared to statutory retirement, may reflect the consequences of being stuck in employment while one had planned to retire, for example due to pension reforms which raise the retirement age or the length of the contribution period required for entitlement to a pension (see e.g. Blake and Garrouste, 2019; Shai, 2018).

Table 3.4: Heterogeneity in the estimated effects of retirement on health

	Bayesian Model Averaging[(a)]			Weighted-Average Least Square $(q = 1)$[(b)]		Weighted-Average Least Square $(q = 0.5)$[(b)]		OLS check after BMA[(c)]			OLS check after WALS[(d)]		
	PM	PSD	PIP	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Focus Regressors													
Publication bias	**8.635**	**5.761**	**1.000**	**8.274**	**5.657**	**8.131**	**5.682**	9.536		8.962	8.809		8.246
Precision effect	**0.004**	**0.003**	**1.000**	**0.013**	**0.004**	**0.013**	**0.004**	0.002	***	0.001	0.008	**	0.004
Auxiliary Regressors													
Google Scholar citations per year	0.000	0.000	0.220	0.000	0.000	0.000	0.000	–		–	–		–
Scimago Journal Ranking	-0.000	0.001	0.320	**-0.002**	**0.001**	**-0.002**	**0.001**	–		–	-0.002	***	0.001
Year of publication	**0.001**	**0.000**	**0.550**	**0.001**	**0.000**	**0.001**	**0.000**	0.000	***	0.000	0.001	***	0.000
Scimago subject areas (Reference category: Multi-area journals)													
Economics/Business	-0.000	0.000	0.060	0.000	0.002	0.000	0.002	–		–	–		–
Medicine/Psychology	-0.000	0.000	0.070	0.001	0.002	0.000	0.002	–		–	–		–
Health outcomes (Reference category: Mortality)													
General and self-reported health	**0.012**	**0.002**	**1.000**	**0.010**	**0.002**	**0.010**	**0.002**	0.011	***	0.002	0.012	***	0.002
Physical health	0.001	0.001	0.160	**0.002**	**0.001**	**0.001**	**0.001**	–		–	0.002	*	0.001
Mental health	**0.009**	**0.002**	**1.000**	**0.007**	**0.002**	**0.007**	**0.002**	0.009	***	0.002	0.008	***	0.003
Healthcare utilization	0.001	0.002	0.400	**0.001**	**0.001**	**0.001**	**0.001**	–		–	0.002	**	0.001
Identification strategies (Reference category: Other methods)													
Regression discontinuity design (RDD)	0.000	0.001	0.120	**-0.006**	**0.004**	**-0.005**	**0.004**	–		–	-0.002	**	0.001
Instrumental variables (IV)	0.000	0.001	0.070	-0.004	0.005	-0.003	0.005	–		–	–		–
Difference-in-differences (DiD)	0.000	0.001	0.050	-0.001	0.004	-0.001	0.004	–		–	–		–
Propensity score matching (PSM)	-0.000	0.002	0.050	**-0.010**	**0.009**	**-0.010**	**0.009**	–		–	-0.008		0.009
Fixed-effects/First-differences	**-0.013**	**0.003**	**1.000**	**-0.014**	**0.005**	**-0.014**	**0.005**	-0.013	***	0.004	-0.013	***	0.003
Institutional contexts (Reference category: Statutory retirement)													
Mandatory or involuntary retirement	**-0.025**	**0.009**	**0.960**	**-0.021**	**0.007**	**-0.022**	**0.007**	-0.026	***	0.008	-0.026	***	0.007
Early retirement	-0.000	0.001	0.110	-0.001	0.001	-0.001	0.001	–		–	–		–
Postponed retirement	-0.001	0.002	0.230	**-0.005**	**0.002**	**-0.006**	**0.002**	–		–	-0.006	***	0.002
Geographical areas (Reference category: Multi-country analyses)													
Europe	-0.000	0.001	0.060	**-0.002**	**0.002**	**-0.002**	**0.002**	–		–	-0.002		0.003
Extra-European countries	-0.000	0.001	0.060	**-0.003**	**0.002**	**-0.003**	**0.002**	–		–	-0.003		0.003
Sex (Reference category: Males)													
Females	0.000	0.000	0.050	0.000	0.001	0.000	0.001	–		–	–		–
Males+Females	-0.000	0.001	0.070	**-0.002**	**0.002**	**-0.002**	**0.002**	–		–	-0.002		0.002
Calculation of t-statistic (Reference category: from 95% CI)													
t-statistic from $\widehat{\beta}_i / SE_i$	-0.001	0.002	0.230	-0.002	0.003	-0.002	0.003	–		–	–		–
Type of previous occupation (Reference category: White collars)													
Blue-collar	0.000	0.004	0.040	0.001	0.002	0.001	0.002	–		–	–		–
Not specified	0.000	0.000	0.050	0.001	0.001	0.001	0.002	–		–	–		–
Birth cohorts (Reference category: Others)													
Birth cohorts $\leq$ 1950	**-0.002**	**0.002**	**0.650**	-0.001	0.001	**-0.002**	**0.001**	-0.003	***	0.001	–		–

Notes: The results are from the PEESE specification using the inverse of the SE_i^2 as weights. PM = Posterior Mean of the coefficient; PSD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability. The number of observations (studies) is 308 (85). Auxiliary variables for which the PIP is above 0.5 in BMA or the corresponding one-standard error band does not include zero in WALS are in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

[(a)] In the BMA, we use the uniform distribution for model priors, the Zellner's g prior for the distributions of the coefficients and a Markov Chain Monte Carlo algorithm to search over the model space, by distinguishing between focus and auxiliary regressors.

[(b)] $q = 1$ indicates the Laplace model prior distribution; $q = 0.5$ implies the Subbotin model prior distribution.

[(c)] The model specification under "OLS" includes those variables which have a PIP > 0.5 in BMA ($R^2 = 0.29$).

[(d)] The second model specification under "OLS" includes those variables which are relevant according to WALS ($R^2 = 0.35$).

Regarding the publication year, we find that the estimated effects of retirement on health tend to be more and more positive over time: the year of publication presents a PIP $= 0.55$ and a positive and significant coefficient. As regards study-quality measures, WALS estimates reveal that the SJR indicator is negatively correlated to the partial correlation coefficient, meaning that the more positive the detected relation between retirement and health, the lower the SJR index of the journal where the result was published. Finally, the health effects of retirement are independent of geographical area, gender, and the previous type of occupation. Concerning this last, it should be taken into account that the number of study results distinguishing between blue- and white-collar workers is fairly low. Hence, our meta-analysis is not endowed with the statistical power to shed light on this particular source of heterogeneity.

Finally, the coefficient of the dummy for study results coming from individuals who were born before 1950 is significant and negative. This means that when studies include more recent cohorts in their samples, the retirement effect on health is more likely to be positive. Nevertheless, the difference is very small.

The results presented in Table 3.4 suggest sources of heterogeneity in the study results. However, it is not easy to visualise from it if for particular combinations of study features the expected retirement effect is significantly positive or significantly negative. To be more informative from this point of view, we used the OLS estimates from the frequentist check after BMA and computed the expected partial correlation coefficients for all the combinations of the covariates, after fixing the publication year to the median and setting γ_0 to zero, so as to mimic the absence of publication bias.

Table 3.5 displays the expected partial correlation coefficients for all the combinations of the explanatory variables. We find that, for the most frequent combination involving mental health as outcome variable (third line of Table 3.5), retirement has a positive and highly significant impact, with a partial correlation coefficient equal to 0.010. The covariate profile with the largest positive predicted partial correlation coefficient (0.013) has general and self-reported health as outcome variable. According to the classifications in Cohen (1988) or Doucouliagos (2011), which set to 0.1 and 0.07, respectively, the size of the partial correlation coefficient to be considered as "small", the detected magnitudes are very modest. In the case of physical health or healthcare utilization or mortality, the predicted average effect for the chosen combinations of covariates is even closer to zero. Finally, regardless the health outcome, when a study focuses on mandatory or involuntary retirement, we predict an expected negative effect between -0.029 and -0.013.

Table 3.5: Expected partial correlation coefficients of the health effect of retirement for all the combinations of covariates (covariates not mentioned in each line are fixed at the reference)

	Coeff.		Std. Err.	p-value	Frequencies Abs.	Rel. (%)
Physical health or healthcare utilization or mortality	0.002	***	0.001	0.001	73	23.70
Physical health or healthcare utilization or mortality + birth cohorts $\leq$ 1950	-0.001	*	0.001	0.064	66	21.43
Mental health	0.010	***	0.002	0.000	65	21.10
General and self-reported health	0.013	***	0.002	0.001	28	9.09
Mental health + mandatory or involuntary retirement	-0.016	**	0.008	0.045	17	5.52
Mental health + birth cohorts $\leq$ 1950	0.007	***	0.002	0.001	15	4.87
General and self-reported health + birth cohorts $\leq$ 1950	0.010	***	0.002	0.000	11	3.57
Physical health or healthcare utilization or mortality + fixed-effects/first-differences	-0.011	***	0.003	0.001	10	3.25
Physical health or healthcare utilization or mortality + mandatory or involuntary retirement	-0.024	***	0.008	0.002	6	1.95
Mental health + fixed-effects/first-differences	-0.003		0.004	0.528	5	1.62
General and self-reported health + mandatory or involuntary retirement	-0.013	*	0.008	0.093	4	1.30
General and self-reported health + fixed-effects/first-differences	0.000		0.004	0.995	3	0.97
Mental health + mandatory or involuntary retirement + birth cohorts $\leq$ 1950	-0.019	**	0.008	0.019	2	0.65
Physical health or healthcare utilization or mortality + mandatory or involuntary retirement + birth cohorts $\leq$ 1950	-0.027	***	0.008	0.001	2	0.65
Mental health + mandatory or involuntary retirement + fixed-effects/first-differences$\leq$ 1950	-0.029	***	0.009	0.001	1	0.33

Notes: *** Significant at 1%, ** significant at 5%, * significant at 10%. Year of publication is normalized at its median value and γ_0 is set to zero.

3.5 Conclusions

We summarized the literature on the impact of retirement on health using meta-analytic techniques. Our meta-sample consisted of 308 observations from 85 articles published in peer-reviewed journals in the period 2000-2021. Among these findings, 28% supported the hypothesis according to which retirement improves health; 60% provided no statistically significant effects; and only 12% reported evidence in favour of a worsening health status after retirement.

In a first step, we checked for the presence of publication bias under the assumption of a common effect and by using a battery of meta-regression based techniques. We did not find evidence for publication bias. The average retirement effect is extremely small, considering the figures suggested by Cohen (1988) or Doucouliagos (2011) to value the size of a partial correlation coefficient as "small".

We then used model averaging strategies to explore possible sources of effect heterogeneity across several study characteristics, like research design, estimation strategy, institutional context, and type of previous occupation. Our results suggest that the different reported estimates are linked to the differences in health outcomes used by studies. The identification/estimation strategy does not appear to be particularly important for explaining heterogeneous findings, although studies which opted for fixed-effects or first-differences tended to report more negative estimated effects. Finally, a further source of heterogeneity is the type of retirement scheme. Compared to standard retirement, mandatory/involuntary retirement and, to a lesser extent, postponed retirement are associated with more negative health outcomes.

These findings have important implications for public policy, especially because many OECD countries still adopt mandatory retirement ages (OECD, 2017, Section 2.4) and are rising further their retirement age (OECD, 2019). Although we find that the effect of retirement on health outcomes is in general very small in magnitude, the predicted precision effects for different combinations of covariates displayed in Table 3.5 suggest that having no choice about the timing of retirement and being involuntarily retired (i.e. the category that we called "mandatory or involuntary retirement") may have health damaging implications. Policy-makers should consider not only the financial sustainability of the pension system, but also the raising healthcare spending due to the negative impact of mandatory or involuntary retirement. Optimal welfare pension policies should ensure workers a greater degree of freedom in choosing whether to retire and the timing of their retirement.

Finally, as suggested by Kuhn (2018), there are reasons to suspect that the health effects of retirement are heterogeneous across dimensions, such as different types of prior occupation (e.g. blue- vs. white-collar workers), different types of physically/mentally demanding previous jobs, time horizons or health behaviours, which are only partially investigated in our paper. We have tried to shed light on whether retirement differently affects blue- and white-collar workers. However, only a very limited number of the studies surveyed distinguished between blue- and white-collar workers. Hence, our meta-analysis has very limited statistical capacity to provide answers on this issue. We therefore conclude with a research suggestion: future research should take these further dimensions into account to gain a clearer picture of the multifaceted nature of the effects of retirement on health.

Chapter 4

Should I stay or should I go? How the timing of retirement affects mortality

4.1 Introduction

In recent decades, pension reforms represent one of the main and most frequent public policy interventions, since most of the OECD countries have carried out reforms that increased the standard retirement age and are considering rising further in the future in order to guarantee the financial sustainability of their pension systems (OECD, 2019; Boeri and van Ours, 2021). In particular, in most European countries the intensity of pension reforms has been particularly strong since the 2000s, with changes in eligibility criteria like the retirement age, the required contributory period and the pension calculation scheme (Carone et al., 2016). Although the empirical literature related to the health effects of retirement provides a significant degree of heterogeneity that could be due to different health outcomes, or differences in the institutional context, understanding the health consequences of retirement is critical to provide policy makers a clearer picture for the design of pension policies that are welfare improving.

About this issue, Filomena and Picchio (2022b) suggested that retirement seems to slightly improve self-reported health and mental health, but having no choice about the timing of retirement, being involuntarily retired or being forced to continue working due to policy reforms which postpone the time of retirement might have some health damaging implications.[1] Moreover, the identification of the causal health effects of retirement

[1] On the role played by choice vs. constraint in shaping the health impact of work and retirement, see

involves methodological issues that are not easy to deal with (Kuhn, 2018). First, omitted variables biases might be induced by differences in unobserved individual characteristics that influence both health and retirement decisions (e.g. subjective life expectancy). Unobserved heterogeneity could be time-constant but also time-varying. To control for unobserved time-constant individual heterogeneity, researchers typically use individual fixed effects panel data models (Eibich, 2015). Second, estimation biases due to reverse causality might arise, because causality not only could run from retirement to health, but it is also likely to go from health to retirement decisions. To deliver credibly estimates of the causal effect of retirement on health, more recent studies address endogeneity issues adopting instrumental variables methods with eligibility ages for retirement as instrument, or regression discontinuity design (RDD) exploiting increase in the retirement probability when a worker attain the eligibility age. Third, estimation biases could be due to measurement errors when researchers adopt subjective health measures as outcome variables. Indeed, the decision to retire early might influence the reporting subjective answers of the interviewees, because they could assess their own health differently after retirement. Self-reported health measures are at risk of two kinds of measurement error: i) self-assessed health might not be comparable across individuals ("classical measurement error"); ii) individual who do not work might justify their labor market status by their ill health ("justification bias"). It refers to retirees' tendencies to exaggerate their poor health conditions in order to provide socially acceptable justification for their retirement and observed health would be understated for retirees (Behncke, 2012; Insler, 2014).

Otherwise, mortality is the most objective health measure and so the easiest to interpret, although only few studies focus on its relationship with retirement. Garrouste and Perdrix (2022) reviewed the health effects of retirement and did not highlight any impacts on mortality nor on the likelihood to develop pathologies, although heterogeneous effects arise for some sub-population analyses. Sewdas et al. (2020) provided a meta-analysis with 25 studies and a focus limited to the link between mortality and early and on-time retirement. Their results show that early retirement, compared to continued working, is not associated with higher risk of mortality, while on-time retirement, compared to continued working, is associated with a higher mortality risk. Finally, van Ours (2022) concluded that on average retirement seems to improve mental health, to deteriorate cognitive skills but to not affect mortality. However, the range of outcomes is related to heterogeneity in terms of personal traits, type of previous job and institutional arrangements. We discuss

also Bassanini and Caroli (2015).

the available empirical evidence in more detail in the next section, where we present a brief overview of previous studies on the relationship between retirement and mortality. These findings are also summarized in Table A4.1, where we distinguish them according to some study-related features.

Our article contributes to this literature studying the effects of retirement and its timing on mortality focusing on a sample of private-sector workers from the AD-SILC database (2004-2010), which is obtained by matching the IT-SILC database and data from the National Social Insurance Agency (INPS). We add to the debate an empirical innovation by adopting a factor analytic model with dynamic selection into treatment to evaluate the causal impact of retirement and its timing on mortality, in which workers differ in unobserved characteristics jointly affecting selection into retirement and subsequent health outcomes. These unobserved traits, such as labor force attachment, liquidity constraints, different health problems or behaviors, may affect the retirement decision and subsequent health outcomes and make difficult the identification of the causal effect of retirement. We perform a factor-analytic dynamic model (FADM) (Carneiro et al., 2003; Heckman and Navarro, 2007) in which we achieve the nonparametric identification of the treatment effect by (i) imposing a factor structure on the unobserved characteristics; (ii) retrieving a complete working history for each individual thanks to the longitudinal structure of the dataset which provides multiple observations over time of the endogenous variables; (iii) resorting to selection-free measures of the latent factor. Because of the dynamic selection into treatment, we take into account selection on the time-varying unobservables jointly affecting retirement decision and health outcomes by the factor structure with a latent trait and time-varying factor loadings.

Furthermore, during the 1990s Italy has experienced a series of pension reforms, that first introduced and then progressively increased minimum retirement age, years of contributions needed and modified benefit calculation schemes. These reforms introduced quasi-experimental variations in pensionable age between individuals depending on their birth date. We exploit the 1992 pension reform in the empirical analysis as exogenous shocks through a dummy variable equal to 1 if the individual was affected by the institutional change and 0 otherwise. The main effect of such reform was the increase of the normal retirement age (NRA) from 60 and 55 to 65 and 60 for men and women, respectively. Because we are interested on the effect of retirement and its timing, the reform works as a further exclusion restriction in our treatment equation. The starting point of retirement decision is the year in which individuals reach ages 50 and the impact of retirement is allowed to vary according to the timing since that year. The evaluation of the health outcomes is estimated at different years after retirement. Finally, we investigate heterogeneities by focusing on gender, educational attainment, marital status, and previous kind of job. These features allow us to identify the effects of retirement and to draw policy implications of its timing.

The set-up of the article is as follows. Section 4.2 presents the literature review on previous studies focused on the effects of retirement on mortality. Section 4.3 describes the institutional framework, data and sample. Section 4.4 presents the econometric model and the identification assumptions to perform our factor analysis. Section 4.5 reports and comments on the main estimation results. Section 4.6 concludes.

4.2 Previous studies on retirement and mortality

As pointed out by Garrouste and Perdrix (2022), the empirical literature mainly shows that switching from employment to retirement has no significant effect on mortality, although some mixed results for different sub-groups. Hernaes et al. (2013) found that a retirement reform in Norway induced some workers to indeed retire early, but their mortality was not affected. Even Grøtting and Lillebø (2020) found no effects of retirement on acute hospitalizations or mortality in Norway, while Rose (2020) estimated a not significant increase in mortality in the UK. Bozio et al. (2021) investigated the impact of delaying retirement on mortality among the French population in the private sector. They showed that an exogenous increase of one year in the claiming age has no significant impact on the probability to die. Hagen (2018) used a reform that raised the age at which broad categories of Swedish local government workers were entitled to retire with full pension benefits from 63 to 65. The results show no evidence that the reform impacted mortality or health care utilization. Hult et al. (2010) indicated that there are no general differences in mortality depending on timing of retirement, but rather on poor health before early retirement. Similar findings are provided by Litwin (2007), where heterogeneity is driven by education, gender and previous diagnosis of major illnesses. Kalwij et al. (2013) found no increased mortality risk during retirement among older individuals who have been early retired or unemployed between the ages 58 and 65. Nielsen (2019) estimated the causal effect of retirement using both a reform induced change in the old age pension age and a large discontinuity in retirement take-up at age 60. The results show that neither early retirement nor statutory retirement has any effects on mortality in Denmark. Coe and Lindeboom (2008) used unexpected early retirement window offers to instrument for retirement behavior and found no negative effects of early retirement on mortality. Eyjólfsdóttir et al. (2019) examined the effect of prolonging working life beyond age 65 on mortality and a series of indicators of late-life physical health in a representative sample of the Swedish population, finding no significant effect of working to age 66 or above.

Different findings arise when studies focus on particular sub-populations: Hallberg et al. (2015) found that a retirement reform for Swedish army personnel increased early retirement and reduced mortality. Bloemen et al. (2017), using a temporary change in the rules for early retirement of older civil servants in the Netherlands, found that early retirement reduces mortality. Lalive and Staubli (2015) estimated that delayed labor force exit increased mortality among Swiss women affected by a pension reform that increase the full retirement age. In contrast, Fitzpatrick and Moore (2018) found that early retirement in the U.S. increased male mortality, especially for unmarried males and with low education levels, but the same does not matter for women. The authors attributed such effect to retirement-associated changes in unhealthy behaviors. Kuhn et al. (2020) used Austrian administrative data finding that retirement increased mortality only for blue-collar men. Brockmann et al. (2009) found a significant higher mortality risk among pensioners with reduced earning capacities than among old-age pensioners who either left the labor market earlier. Healthy people who retire early do not experience shorter long-term survival than those who retire late. On the contrary, taking into account the hospitalization prior to retirement, early retirement significantly lowers mortality risks. Zulkarnain and Rutledge (2018) showed that delayed retirement reduces 5-year mortality rate for men but not for women in the Netherlands.

4.3 Data and sample

4.3.1 Sample selection criteria

To investigate the impact of retirement and its timing on mortality we made use of the AD-SILC database. It is obtained by merging the survey data from the IT-SILC database provided by the Italian National Institute of Statistics (ISTAT) with the administrative data on labor market histories gathered by the National Social Insurance Agency (INPS). We extracted data on Italian workers from each wave from 2004 to 2010. Moreover, the

AD-SILC is matched with the regional time series of unemployment, employment, real GDP growth rates and hospital beds per 1,000 inhabitants as a proxy for public spending on health from ISTAT, which we used as time-varying controls in our empirical analysis.

We aim to analyze the effect of retirement on health outcomes for both men and women. However, the longer the time horizon we want to analyze, the fewer individuals we can follow until 2013, so the number of individuals we can follow for a longer time span is decreasing with the size of the time window considered after retirement.[2]

We restricted the subsequent empirical analysis focusing on a sample of males and females born in the birth cohort 1929-1944. Firstly, since the IT-SILC database contains survey data rather than administrative data, it is not so easy to study the effect of retirement on mortality. As pointed out by Brugiavini and Peracchi (2012), the historical data on mortality trends document an important increase in life expectancy and longevity. The main reasons may be a better quality of health services and a more responsible behavior at the individual level in terms of lifestyle, and the increase in longevity experienced by the Italian population in the past decades is also correlated with increasing exits from the labor force of older workers. Indeed, in this case the more we reduced the minimum age of the individuals in the sample, the more we would have an oversizing of individuals in good health, and an undersizing of individuals in poor health or dead in the time span of the analysis. Therefore, we had a real sample selection, considering the birth cohort 1929 as the maximum age. Secondly, the youngest cohort is set to be 1944 since mortality is not a common issue for younger cohorts. For the same reason, we estimated the impact of the timing of retirement on mortality starting from 72 years old (see footnote 4.3.1) up to 78 years old after grouping the number of observations by periods of 3 years.[3]

The starting sample of 356,739 observations contained personal information on all individuals. However, each individual in the IT-SILC dataset is interviewed for up to 4 consecutive years. Since only time-invariant information are obtained from this dataset,

[2]Note that the individuals who died do not exit from the sample of the subsequent years of evaluation. The only reason because the total number of observations decreases is the right-censoring, i.e. we can follow individuals only up to 2013. Thus, the sooner an individual reaches 50 years old, the longer we can keep him/her in the sample.

[3]Consequently, we were not interested in evaluating the impact of pension reforms introduced after that one in 1992. The subsequent 1995 reform changed the pension formula only for those with less than 18 years of contributions in 1995 (tipically workers born between 1955 and 1965), and in practice it left the cohorts born before 1945 unaffected (Ardito et al., 2020). The Amato reform had instead a major effect on retirement behavior as it was the first signal of a coherent redesigning of the social security system (Brugiavini, 1999), and this is one of the main reasons why we focus only on it. A detailed discussion of the pension rules in Italy before and after the 1990s pension reforms is presented in the Appendix B.

we decided to consider only the first interview of each individual, thus avoiding "re-call" biases. This restriction reduced the sample to 147,436 observations. The following match with data on county of births allowed us to exclude from the analysis 8,118 individuals born abroad, since if they worked in their country of origin before to move in Italy, they may be subject to different eligibility criteria to claim the pension age. Table 4.1 reports in more detail the selection criteria that reduced the sample to individuals for whom we re-built labor market histories such as wages, participation, and year of retirement thanks to the INPS administrative data. Since 31,782 individuals not included in the INPS database are dropped because self-employed or inactive, our analysis focuses only on salaried employees.

Table 4.1: Sample size across selection criteria

	Men		Women	
	Left in the sample	Removed	Left in the sample	Removed
Individuals in IT-SILC (2004-2010)	172,899	–	183,840	–
After taking the first interview only	71,509	101,390	75,927	107,913
After removing individuals with errors on gender	71,505	4	75,912	15
After removing individuals born abroad	67,930	3,575	71,369	4,543
After removing individuals not included in the INPS database	55,263	12,667	52,254	19,115
After selecting only individuals born in 1929-1944	10,705	44,558	11,321	40,933
After removing individuals due to incorrect information related to working periods	10,690	15	11,309	12
After removing individuals receiving only other kinds of pension benefits	9,905	785	8,820	2,489
After removing individuals retired before 50 years old	9,721	184	8,490	330
After removing individuals who died before reaching 50 years old	9,720	1	8,490	–
After removing public employees and professionists	9,507	213	8,149	341
After removing individuals with errors on the date of death	9,503	4	8,146	3
After removing individuals not included in the INPS database for more than 5 years before ages 50	9,464	39	7,386	760
After removing not retired men (women) who died within 65 (60)	9,443	21	7,381	5
After removing individuals with 5-year average previous earning greater than €700,000 (outliers)	9,442	1	7,380	1
After removing individuals not observed at least at 72 years old	7,653	1,789	5,923	1,457
After removing individuals interviewed at 81 years old and observable only at 84	7,621	32	5,897	26
After removing individuals never retired during the period under analysis	7,532	89	5,744	153
After removing individuals retired at more than 70 years old	7,471	61	5,704	40
After removing individuals observable only at 81 years old	7,120	351	5,414	290
Final sample	7,120	165,779	5,414	179,690

Because we are interested in studying the effect of retirement and its timing, we focused on the transitions from work to retirement and excluded 3,274 cases of transitions occurring through disability or other retirement options. We also excluded 514 cases of early retirement occurring at less than 50 years of age. These criteria allowed us to distinguish between forced retirement due to health reasons or involuntary job loss and statutory retirement because they might entail different health effects.[4]

[4]Moreover, we focused only on private-sector employees, due to the different pension rules that cover public-sector workers and professionists. We also excluded 799 individuals with only sporadic episodes of subordinate employment during their career (e.g. housewives or individuals who became self-employed). Furthermore, in order to model the retirement decision, we excluded further 21 (5) men (women) who died before reaching 65 (60) and not yet retired.

To avoid selection issues we can estimate the probability of survival at a certain age after retirement decision only if individuals are interviewed before reaching that age. There were 3,246 individuals that we might follow only up to 69 years old, and we decided to exclude them from the analysis because the event of death is too rare at that age.[5] Thus, at this point the sample was composed only by individuals for which the age at the interview is lower than the age at which we observe the health outcome (see Table 4.2 for more details).[6]

After applying the aforementioned selection criteria, our final sample is made up of 12,534 observations, of which 7,120 males and 5,414 females.

4.3.2 Descriptive statistics

In what follows we presented some preliminary statistics. Table 4.2 shows the number of observations from 72 to 78 years old grouped by periods of 3 years, with descriptive statistics related to time-invariant characteristics that are provided after splitting the sample between men and women. The main differences among gender appear to be related to the average labor earnings and labor market participation during the last 5 years before reaching ages 50. This reflects that the labor market functioning is gender sensitive. Especially in Italy, the traditional family gender roles and family functions translated into the male breadwinner model and the mother caretaker model (Saraceno, 1994). Further differences concern education: older cohorts of women rarely attained secondary and tertiary education levels.

Table 4.3 reports complete information on our outcome variable, that is a binary variable of value 1 if the individual is alive during the observed time window and 0 otherwise, by different ages at retirement and in each specific time window. We note that the probability of survival is higher in individuals who switch into retirement "on time", that is at the normal retirement age according to the rules before the reform. Further details on age at death across different eligibility cohorts and on age at retirement are provided in

[5]The samples evaluated at 69 years old were composed by 5,286 men and 4,074 women, in which the cases of non-survival were only 248 (4%) and 85 (2%), respectively. Thus, we decided to evaluate the health outcome starting from 72 years old.

[6]As a consequence, 58 individuals were interviewed at 81 years old, so we could not observe their health outcome later and we removed them from the final sample. Finally, 242 individuals never retired during the period we followed them, whereas further 101 individuals got pension after 70 years old. We removed these individuals to avoid error measurements in the original dataset or to exclude the possibility that these individuals retired under different pension schemes.

Table 4.2: Sample composition at different years of outcome evaluation

Age at which the outcome is observed	Year of birth	Age at the interview	Age at retirement	% Primary education	% Secondary or tertiary education	5-year average labor earnings	% 5-year average fraction of days at work	Observations
	Males							
72	1938	67	59	77	23	18,210	58	4,652
75	1936	70	60	81	19	17,049	58	4,288
78	1934	72	60	83	17	16,481	56	3,658
	Females							
Age at which the outcome is observed	Year of birth	Age at the interview	Age at retirement	% Primary education	% Secondary or tertiary education	5-year average labor earnings	% 5-year average fraction of days at work	Observations
72	1938	68	58	86	14	8,064	39	3,394
75	1935	70	57	89	11	8,272	42	3,176
78	1933	73	58	89	11	8,715	40	2,594

Notes: The table shows the mean values of the age at retirement, the age at the interview, and other time-invariant characteristics predetermined with respect to the treatment for the number of individuals analysed at three different ages after retirement.
[(a)] Yearly labor earnings used to calculate the average labor earnings before ages 50 are in 2014 prices and deflated by the ISTAT consumer price index.
[(b)] The fraction of sick leave is computed as the ratio between days in sick leave and total contract days in the same year.

the supporting information (see Table C4.1). On average, men switched into retirement at 59 years old, while women postponed retirement of about 3 years with respect to the age 55. This means that men usually experience longer and uninterrupted career profiles which allow them for seniority pension. By keeping the sample separated by gender, we observe that 1,396 males died during the period under analysis (20%), with an average age at death of 74.6. Otherwise, less than 11% of women died in the same time window at an average age of 75.

Table 4.3: Survival at different ages by the timing of retirement

		Males				Females			
		Survival		Observations		Survival		Observations	
Survival at age	Timing of retirement	Mean (%)	Std. Dev.	Absolute	Relative (%)	Mean (%)	Std. Dev.	Absolute	Relative (%)
72	Early retirement	88.8	31.5	2,491	53.55	96.3	19.0	322	9.49
	Retirement at NRA	94.1	23.5	375	8.06	97.2	16.6	781	23.01
	Postponed retirement	90.9	28.8	1,786	38.39	94.7	22.4	2,291	67.50
	Total	90.0	29.9	4,652	100.00	95.4	20.9	3,394	100.00
75	Early retirement	79.2	40.6	1,963	45.78	90.5	29.4	231	7.27
	Retirement at NRA	90.6	29.2	606	14.13	95.4	21.0	1,190	37.47
	Postponed retirement	83.8	36.9	1,719	40.09	89.5	30.7	1,755	55.26
	Total	82.6	37.9	4,288	100.00	91.8	27.5	3,176	100.00
78	Early retirement	62.4	48.5	1,461	39.94	81.1	39.3	169	6.52
	Retirement at NRA	82.9	37.6	709	19.38	87.1	33.6	921	35.51
	Postponed retirement	72.8	44.5	1,488	40.68	82.4	38.1	1,504	57.98
	Total	70.6	45.6	3,658	100.00	84.0	36.7	2,594	100.00

Notes: NRA is the normal retirement age according to the rules in force before the 1992 pension reform (60 years old for men and 55 years old for women).

Furthermore, Table C4.2 in the Appendix reports preliminary correlations between the timing of retirement and mortality. These first preliminary findings are also graphically displayed in Figure C4.3 and show that the impact of 1 year delay of retirement on the probability of survival is fairly nil for both men and women. However, at this stage the estimation results cannot be interpreted as a causal relationship. Although our outcome variable is mortality and we avoid reverse causality and measurement errors in the health indicator issues, unobservables factors correlating with both health and retirement are likely to affect the results. For instance, health problems and behaviors which we are not able to observe might jointly affect both selection into retirement and subsequent health outcomes. Furthermore, attachment to the labor market might influence retirement decision and subsequent health outcomes. The next sections explain the econometric model aimed at disentangling the true causal effect of retirement (occurring at different time across individuals) on mortality from the spourious one induced by unobserved traits of workers with different labor market histories, health problems and behaviors individuals may have experienced during their life.

4.4 Econometric model

4.4.1 General framework

We outlined a model with multiple time periods and treatment effect that is heterogeneous over the timing of retirement. Let $i = 1, ..., n$ index an individual and t the age at which the health outcome is evaluated. To keep the model tractable in estimation and have a limited number of equations, we restricted the set of the age index t to {72, 75, 78}.

We denoted as Y_{it} the health outcome, that is an indicator variable of value 1 if the individual is still alive in the observed time window and 0 otherwise. Furthermore, we defined a dummy variable D_{ir} that takes value 1 if an individual is retired at time r, with $r = 0, ..., 2$, where $r = 0$ indicates early retirement, $r = 1$ is for retirement at the NRA before the 1992 pension reform, $r = 2$ for postponed retirement over NRA, and the effect of retirement is observed if it occurs before t. For each individual i the observed health outcome at time t can be written as

$$P(Y_{it} = 1|D_{ir}, X_{it}) = \Lambda(\sum_{r=0}^{2} \beta_t D_{ir} + \mu_t(X_{it})) \quad (4.1)$$

where β_t is the effect of the treatment variable D at time r on mortality at ages t, μ_t is a function of observed covariates X_{it} and $\Lambda(\cdot)$ is the standard logistic cdf of the logit specification. Because we are interested in the effect of the timing of retirement on mortality, our treatment is a categorical variable corresponding to the distance between the year in which the individual switches into retirement and the year in which the same individual was 50. Therefore, there is no single effect of retirement, but different effects of retirement depending on its timing. We modeled selection at time r after reaching 50 years old as a function of a treatment-time specific index

$$V_{ir} = v(Z_i) + u_{ir} \tag{4.2}$$

where v is a function of a vector of covariates Z_{ir}, u_{ir} denotes the individual and treatment-time unobserved heterogeneity, $r \in \{0, 1, 2\}$, and treatment time is selected according to

$$D_{ir} = 1(R_i = r | D_{ir'} = 0, r' < r) \tag{4.3}$$

where $1\{\cdot\}$ is an indicator function and $R_i = r$ if V_{ir} crosses some threshold. The selection into treatment is dynamic and treatment time r can only be selected if treatment has not been taken before. Thus, we estimated the impact of retirement and its timing on mortality at 72, 75, and 78 years old. In summary, we estimated the parameters for 6 outcomes separated by sex, along those entering selection and two measurement equations. Hence, our econometric framework is like to the one in Fruehwirth et al. (2016), where the effect of the timing r of one treatment is related to the outcome of interest over t.

4.4.2 Identification strategy

The identification of the effect of the timing of retirement on subsequent health outcomes requires to take into account unobserved heterogeneity across individuals. Firstly, this might be related to differences in unobserved factors correlating with both health outcomes and the time of retirement (Ardito et al., 2020), such as labor force attachment, financial constraints, previous health problems, and different health behaviors. For instance, workers with high labor force attachment may be more likely to postpone their retirement decision or, at the same time, a long working career in physically/mentally demanding jobs may induce workers to switch into retirement at the normal retirement age. Moreover, workers with financial liquidity constraints may be less likely to switch into retirement and more likely to keep working. A further example of unobserved traits may concern different health behaviors of the individuals. Secondly, unobserved heterogeneity is likely to change over time, affecting the retirement decision. For instance, health shocks may affect the worker's health conditions or those of his/her partner at some point and modify the preference of the workers towards the choice to retire from work.

Thus, we need to specify the joint distribution of the unobserved components determining both the outcomes and the selection into treatment. In doing so, we performed a factor analytic dynamic model where the unobserved terms of outcomes and selection into treatment equations are composed of a latent factor θ. It collects the unobserved differences among workers that determine the selection into treatments and the effect of the timing of retirement on subsequent mortality, and the error terms are conditionally independent given the factor. In order to account for the unobserved heterogeneity, we recovered the joint distribution of the unobservables in the selection (u_i) and outcome equations (ϵ_{it}) by imposing a factor structure (Carneiro et al., 2003; Fruehwirth et al., 2016). Thus, we have

$$\epsilon_{it} = \alpha_t \theta_{it} + \varepsilon_{it} \tag{4.4}$$

$$u_{ir} = \lambda_r \theta_{ir} + v_i \tag{4.5}$$

where θ_{it} is a latent factor in $\theta_i = (\theta_{i1}, ..., \theta_{iT_i})$ with a multivariate distribution with $cov(\theta_{it}, \theta_{it'}) \neq 0$, for all $t \neq t'$. It is a vector of mutually independent factors, as well as the error terms. In summary, the unobserved terms in the outcome and treatment equations are made of a latent factor θ which collects unobserved differences among individuals, and random components ε_{it} and v_i. Unobserved heterogeneity varies over time because of the factor distribution and a linear combination of the factor with time-varying coefficients that are called *factor loadings* and denoted as α_t and λ_r. Thus, we can see how selection into retirement and the treatment effect of retirement vary by unobserved individual charachteristics and the timing of retirement.

Factor analytic dynamic models have previously been used in the literature on education and labor economics: Carneiro et al. (2003) studied the impact of different schooling levels on future returns; Fruehwirth et al. (2016) and Cockx et al. (2019) estimated how grade retention affects subsequent school performances; Picchio et al. (2021) studied the effect of childbirth and its timing on female labor market outcomes in Italy. We are the first in performing a factor analytic model with dynamic selection into treatment in the literature on the effect of retirement and its timing on health. In our case, as in Picchio et al. (2021), unobservables are all included in a single latent factor θ, instead of differencing by several sources of unobserved heterogeneity. Thus, our framework differs from Fruehwirth et al. (2016), where latent variable is composed by multidimensional unobservables abilities. In order to identify the distribution of θ, we applied the normalization $\alpha_t = 1$.

Following Carneiro et al. (2003), we relied on selection-free measurements to control for the unobservables that jointly determine selection into treatment and its effect, and to reduce the degree of arbitrariness of factor analysis. We made use of predetermined information with respect to the moment of reaching 50 years old and therefore to the selection into treatment. Thus, we specified our additional measures as

$$M_i^l = \omega^l(S_i^l) + \xi^l \theta_{i72} + e_i^l \tag{4.6}$$

with $l = 1, 2$ and where M_i^l are predetermined information with respect to reaching ages 50 and the realization of the treatment intensity. ω^l consists in a linear combination of observed covariates S_i^l, ξ^l is a factor with time-varying coefficients and e_i^l is a zero-mean error term independent of both S_i^l and θ_{i72}.

We have two additional measurement equations which contain predetermined characteristics of each individual. These latent variables are crucial in order to model the unobserved heterogeneity due to persistent differences in unobservables characteristics such as labor market attachment, health behaviors and/or unobservables persistent shocks that could simultaneously affect both selection into treatment and the outcome of the treatment. The first measure M_i^1 is a variable which corresponds to the average fraction of days spent at work during the 5 years preceding the achievement of age 50. The second measure M_i^2 is the average labor earning the worker received in the same time window. These two continuous variables are specified as follows:

$$M_i^1 = s_i' \zeta^1 + \xi^1 \theta_{i72} + e_i^1 \tag{4.7}$$

$$M_i^2 = s_i' \zeta^2 + \xi^2 \theta_{i72} + e_i^2 \tag{4.8}$$

This measures are likely to be determined by unobserved traits like labor force attachment and liquidity constraints. For instance, workers with financial constraints might need to work longer to make ends meet (Hofäcker et al., 2015), and opt for postpone the timing of retirement. Career attachment has been shown to be critical for the retirement process (Wang and Shi, 2014): as suggested by Adams et al. (2002), workers' career attachment is negatively related to the decision to retire. Otherwise, workers with higher experience of nonemployment may opt to retire earlier because of health problems or lower labor force attachment. Moreover, unemployment, employment and financial constraints may affect workers in different class positions in different ways (Radl, 2013). As noted by Hofäcker and Naumann (2015), higher educated and high-skilled workers show higher labor market attachment because of larger non-material resources such as social contacts. Otherwise, lower educated and blue collar workers may rather be driven by a financial need to remain employed, albeit in physically demanding working conditions and often resulting in health impairments later. At the same time, their economic disadvantage may be probably correlated with bad health conditions or unhealthy behaviors, according to the socioeconomic gradient in health (Blane, 2006). Such unobserved charachteristics should be relevant in explaining both the retirement decision and subsequent health outcomes. In these measurement equations the explanatory variables s_i are independent of θ_{i72}. Once the latent factor is specified as time-varying, our measurements will be functions of θ_{i72} entering both the treatment and the outcome equations at age 72.

Our selection into treatment equation D_{ir} takes value 1 if the individual retire from work at r, according to the latent index V_{ir}. It can be written as follows:

$$V_{ir} = M_i' \gamma_M + z_i' \gamma_z + \lambda \theta_{it} + v_i \tag{4.9}$$

where we included the effect of the two measures described above (M_i') and individual characteristics of worker i (z_i), which are independent of θ_{it}. Since our treatment is a categorical variable consisting in the time distance between the year of retirement and the year of reaching ages 50, we model it through an ordered logit model with dependent variable R_i, where $R \in \{0,\ 1,\ 2\}$. Because of the dynamic structure of the model, estimation has to be based on treatment-time specific probabilities. Considering v_i the linear index of the previous equation and containing only combinations of observables explanatory variables, then the probability that individual i switches into retirement at r can be written as

$$\begin{aligned}
P(r = 0|V_{ir}, \theta_{it}) &= \Lambda(\delta_0 - v_i - \lambda\theta_{it}), \\
P(r = 1|V_{ir}, \theta_{it}) &= \Lambda(\delta_1 - v_i - \lambda\theta_{it}) - \Lambda(\delta_0 - v_i - \lambda\theta_{it}), \\
P(r = 2|V_{ir}, \theta_{it}) &= 1 - \Lambda(\delta_1 - v_i - \lambda\theta_{it}),
\end{aligned} \tag{4.10}$$

where $r = 0$ indicates early retirement, $r = 1$ is for retirement at the NRA before the 1992 pension reform, $r = 2$ for postponed retirement over NRA, and $\Lambda(\cdot)$ is the standard logistic cdf.

At this point, we re-writed our outcome equations concerning the probability of survival at ages $t \in \{72, 75, 78\}$, which are jointly analysed in our framework. We can write it as follows:

$$P(Y_{it} = 1|D_{ir}, X_{it}, \theta_{it}) = \Lambda(\beta_{tr} D_{ir} + M_i^l \pi + x_{it}' \pi_x + \alpha_t \theta_{it}) \tag{4.11}$$

where β_{tr} corresponds to the treatment effect, x_{it} is a vector of individuals time-constant and time-varying characteristics and it is independent of latent factor θ_{it}.

Assuming that the regularity conditions (A-1 and A-2) in Carneiro et al. (2003) hold, the nonparametric identification of the previous distribution is obtained as in Heckman and Smith (1998) and work as follows: we identify the joint distribution of (ϵ_i, u_i, v_i), with $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iT})$, $v = (v_i^1, v_i^2)$, with $v_i^l = \xi^l \theta_{i72} + e_i^l$, and $l = 1, 2$. Moreover, we included some continuous variables among the set of observed determinants of one outcome but excluded from the others, as shown in Table C4.3. These variables are the regional employment rate, the regional unemployment rate, the regional GDP growth rate, and the regional number of hospital bed per 1,000 inhabitants. More in detail, we used: i) the average values during the 5 years before ages 50 in $\omega^l(S_i^l)$ for $l = 1, 2$ and in $v(Z_i)$; ii) the average values for the years between 50 years old and each ages t in $\mu_t(X_{it})$. In Carneiro et al. (2003), this is enough to satisfy the support condition (A-3) necessary to prove the nonparametric identification. These regional rates in the set of covariates provide additional identification conditions and are of help to identify the causal effects of endogenous variables in a dynamic discrete time panel data model (Bhargava, 1991; Mroz and Savage, 2006).

4.4.3 Likelihood function

Let include all the parameters for our measurement, treatment and outcome equations in $\phi = (\tau^1, \tau^2, \psi, \varphi)$. The likelihood for individual i is the joint density of (M_i^l, V_i, Y_i) conditional on observable and unobservable characteristics, so the individual contribution to the likelihood function can be written as

$$\mathcal{L}_i(\phi \mid M_i^l, V_i, Y_i, S_i^l, Z_i, X_i, \theta_i) = g(M_i^l \mid S_i^l, \theta_{i72}, \tau^l)h(V_{ir} \mid Z_i, \theta_{i72}, \varphi)f(Y_i \mid M_{it}, D_{ir}, X_{it}, \theta_{it}, \psi) \tag{4.12}$$

where all the sets of covariates contain the constant, g, h and f are logistic density functions.

In order to account for the presence of individual time-varying unobserved heterogeneity, the vector of latent factor $\theta_i = (\theta_{i72}, ..., \theta_{i78})$ follows a multivariate discrete distribution with H support points. Thus, θ_i takes values θ^h, with $h = 1, ..., H$, following a multi-logit parametrization

$$p^h = Pr(\theta_i = \theta^h) = \frac{exp(p^h)}{\sum_{u=1}^{H} exp(p^h)} \tag{4.13}$$

with normalization $\theta^1 = 0$ and $p^H = 0$. The i-th contribution to the likelihood becomes

$$\mathcal{L}_i(\phi, \rho, \Theta \mid M_i^l, V_i, Y_i, S_i^l, Z_i, X_i) = \sum_{h=1}^{H} p^h \mathcal{L}_{ih}(\phi, p^h \mid M_i^l, V_i, Y_i, S_i^l, Z_i, X_i, \theta_i = \theta^h) \tag{4.14}$$

that is the likelihood in Equations (4.12), conditional on θ_i taking value θ^h. The matrix Θ contains the vectors of support points $(\theta^1, ..., \theta^H)$, whereas the vector ρ collects the weights determining the H masses of probabilities.

The sample log-likelihood is the sum across the natural logarithm of the individuals contributions in Equation (4.14):

$$ln(\mathcal{L}) = \sum_{i=1}^{N} ln[\mathcal{L}_i(\phi, \rho, \Theta \mid M_i^l, V_i, Y_i, S_i^l, Z_i, X_i)] \tag{4.15}$$

which is maximized with respect to its parameters using analytical derivatives. Moreover, we imposed that the coefficients of the covariates for the health outcome equation not vary with t to save in the number of parameters to estimate.

4.5 Estimation results

4.5.1 Main findings

In order to estimate the abovementioned model, we make use of three different assumptions as concern the latent factor structure. In particular, we firstly estimate our model without unobserved heterogeneity; and then we introduce a time-constant and a time-varying latent factor with discrete distribution. The differences across these three specifications are reported in Table 4.4, which shows post-estimates statistics such as the log-likelihood values and the Akaike and Bayesian Information Criteria (AIC and BIC, respectively).

Table 4.4: Summary statistics on the estimated models across different assumptions on the latent factor

	Males			Females		
	Without unobserved heterogeneity	Time-constant unobserved heterogeneity	Time-varying unobserved heterogeneity	Without unobserved heterogeneity	Time-constant unobserved heterogeneity	Time-varying unobserved heterogeneity
Number of parameters	60	69	71	60	69	71
Log-likelihood	28,429	25,120	25,118	17,229	12,185	12,181
AIC	56,979	50,377	50,378	34,578	24,507	24,505
BIC	57,391	50,851	50,866	34,974	24,962	24,973
Distribution of the latent factor	–	Discrete	Discrete	–	Discrete	Discrete
Number of support points	–	3	3	–	3	3

Notes: AIC = Akaike information criterion; BIC = Bayesian information criterion.

In choosing the number of support points, we follow the recommendation provided by Gaure et al. (2007) and adopted by Picchio et al. (2021), that is to reach the number of support points which minimizes the Akaike Information Criterion. In doing so, when we assume time-constant and time-varying unobserved heterogeneity we made use of $H = 3$ support points and we obtain a continuous improvement in terms of information criteria. We stopped at $H = 3$ support points because of numerical problems going on. Nevertheless, the estimated coefficients of the treatment dummies concerning the timing of retirement had became very stable along the last specifications and not affected by further increases in the number of support points. As we can see from Table 4.4, the time-varying specification of the latent factor yield the best results in terms of information criteria, both for men and women, although these are very close to the time-invariant specifications. In what follows, we report and comment on the effects of the timing of retirement across the two different specifications, whereas sections D and E in the Appendix report the full set of estimation results for all the models.

The main objective of the analysis is to evaluate whether retirement has an effect on future health outcomes that we measure through the probability of survival at different ages, and if such effect is different according to the timing of retirement. Table 4.5 shows the estimated logit coefficients of both early and postponed retirement (with respect to the NRA before the 1992 pension reform) on the probability of survival evaluated at different t (72, 75, 78) without unobserved heterogeneity and assuming time-constant and time-varying unobserved heterogeneity, respectively. The results are also graphically displayed in the supporting information.

Table 4.5: Estimated (logit) coefficients of the timing of retirement on the probability of survival

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Without unobserved heterogeneity						
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.397	-0.117	0.217			
	(0.282)	(0.203)	(0.153)			
Postponed retirement $\in 61, R$	-0.269	0.017	0.514***			
	(0.290)	(0.206)	(0.154)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.669	-0.372	0.055
				(0.465)	(0.411)	(0.327)
Postponed retirement $\in 56, R$				0.382*	-0.541**	0.039
				(0.318)	(0.233)	(0.171)
b) With time-constant unobserved heterogeneity						
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.422	-0.120	0.242			
	(0.283)	(0.203)	(0.156)			
Postponed retirement $\in 61, R$	-0.255	0.019	0.504***			
	(0.291)	(0.206)	(0.154)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.675	-0.425	-0.031
				(0.467)	(0.425)	(0.340)
Postponed retirement $\in 56, R$				0.362	-0.553**	-0.002
				(0.317)	(0.236)	(0.175)
c) With time-varying unobserved heterogeneity						
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.413	-0.119	0.225			
	(0.283)	(0.204)	(0.156)			
Postponed retirement $\in 61, R$	-0.263	0.018	0.506***			
	(0.291)	(0.207)	(0.155)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.748	-0.397	-0.054
				(0.467)	(0.425)	(0.343)
Postponed retirement $\in 56, R$				0.411	-0.613***	-0.041
				(0.318)	(0.243)	(0.176)
Observations	4,652	4,288	3,658	3,394	3,176	2,594

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

It clearly emerges that the estimated effects are mostly the same along the three different latent factor structures employed in our econometric framework. On the one hand, early retirement has no effect on the probability of survival for both men and women at different t. On the other hand, delaying retirement after the pre-1992 NRA increases the probability of survival at 78 by 0.506 for men, but also reveals a higher probability of death at 75 by -0.613 for women. Although these results are in line with previous empirical studies which detect, on average, a null effect of retirement on mortality (see e.g. Coe and Lindeboom, 2008; Hernaes et al., 2013; Bozio et al., 2021), or a different effect by gender, with men negatively affected by retirement (Fitzpatrick and Moore, 2018; Zulkarnain and Rutledge, 2018), the most relevant finding is the lack of unobserved heterogeneity in our framework, highlighted by very similar coefficients across the three main models, by the very similar values of the log-likelihood and information criteria between the last two different assumption on the latent factor, and by the absence of any correlations across the 6 equations, as shown in Tables F4.5 and F4.6. These findings suggest that mortality is such an extreme event and then is hardly affected both by the timing of retirement and by unobserved traits across individuals. In contrast, these unobserved characteristics are likely to have an impact on further health conditions which characterize individuals at different ages after retirement, that is on morbidity. Therefore, we suspect that our model may prove even more useful in identifying the effects of retirement and its timing on health outcomes other than mortality.

4.5.2 Sensitivity analysis

We run several sensitivity checks to assess the robustness of our findings and to check any heterogeneous dimensions. Firsly, we modify the definition of the treatment in the benchmark model using the distance between 50 and the year of retirement to evaluate the impact of 1 year delayed retirement. The results are shown in Table G4.1 and are in line with those obtained using the benchmark definition of the treatment; the timing of retirement has no effect on the probability of survival, although delaying retirement of 1 year is positively associated to survival at 78 years old for men.

Secondly, we use a different combination of exclusion restrictions to test if they play a relevant role on our results. For instance, geographical area, local labor market conditions, business cycle and public health investments during the last years of the working career may affect not only our predetermined measures and the selection into retirement, but also impact on future health outcomes. Thus, we proceed by testing the main findings under a different combination of these exclusion restrictions, that is we include the average regional unemployment, employment, GDP growth rates and number of hospital beds per 1,000 inhabitants during the last 5 years before reaching ages 50 in the future health outcome equations. The findings from these alternative specification are reported in Table G4.2 and are in line with the benchmark ones.

Thirdly, in order to check the presence of heterogeneous findings across sub-populations we run several further checks by splitting the sample according to the previous kind of job, education and marital status. Indeed, Hofäcker and Naumann (2015) suggest that lower educated and blue collars are employed in physically demanding working conditions and therefore are more exposed to subsequent health impairments. Thus, we perform heterogeneity analyses by distinguishing between: i) blue collar vs. white collar workers; ii) low educated vs. higher educated workers; iii) married vs. unmarried (at the time of the interview).[7] Tables from G4.3 to G4.6 show the related findings. We note that some heterogeneities arise in terms of previous kind of occupations. While we do not detect any effects of retirement on mortality on the reduced sample of white collar workers, postponing retirement after the initial NRA has an asymmetric health effect on blue collar workers, that is positive for men at 78 and negative for women at 75. The same findings concern different educational achievements for males and females. However, these results should be taken with caution. Indeed, although the literature suggests that higher educated or white collar workers may benefit from continuing to work because of their higher labor market attachment and social networks, the number of individuals making up this subsample is extremely small. Second, blue collar and low educated workers are commonly employed in more physically-demanding jobs and then remaining employed after the NRA is supposed to determine health impairments. The analysis on women is in line with this hypothesis, but the results on men are in contrast with previous evidence and remain quite puzzling. Finally, as for marital status at the moment of the interview, the positive effect on the probability of survival at 78 of delaying retirement is greater for unmarried men; this suggests that they benefit more from continuing to work rather than retiring and living alone in old age, whereas getting pension after 55 reduces the same probability at 75 for married women.

[7] First, in order to define a worker as blue or white collar we used the mode of his/her qualification, measured on the last 10 years the individuals are present in the INPS dataset as employees. Where the mode was missing, we exploited the information about the overall career and therefore a worker was classified as blue collar if he spent more years as blue rather than white collar. However, 2,283 individuals had not these information so this heterogeneity analysis is performed on a reduced sample. Second, we classified as lower educated workers those ones with primary education level, while individuals who atteined at least the high school diploma were classified as higher educated. Third, unmarried means divorced, widowed or single.

4.6 Final remarks

We estimate the effect of retirement and its timing on mortality in Italy on a sample of individuals born between 1929 and 1944 and previously occupied in the private sector. The empirical analysis is carried out separately for men and women and estimates the impact at 72, 75, and 78 years old.

We use a factor analytic model with dynamic selection into treatment, taking into account time-varying unobserved heterogeneity jointly affecting the retirement decision and subsequent health outcomes. In line with previous empirical literature, we find that retirement does not significantly affect mortality; we only detect an asymmetrical effect of delaying retirement on men and women in the long-term. Furthermore, we shed light on further sources of heterogeneity. On the one hand, higher educated and white collar workers are not affected by the timing of retirement. On the other hand, the positive effect on survival at 78 for men seems to be driven by blue collar, lower educated and unmarried individuals, whereas the negative effect at 75 of delaying retirement arises in particular for married women.

Our results make difficult to draw policy recommendations, but although the effect of retirement on mortality is on average nil, we cannot conclude that the cessation of the working career has no impact on individuals' health. In particular, the most relevant finding is the lack of unobserved heterogeneity in our framework, highlighted by very similar coefficients across the three main models, and by the absence of any correlations across the 6 equations. We may speculate that mortality is such an extreme event which is hardly affected both by the timing of retirement and by unobserved traits of the individuals. In contrast, these unobserved traits are likely to impact several health conditions which characterize individuals at different ages after retirement. This issue deserves future investigations by researchers, who will therefore focus on the effects of retirement on morbidity, i.e. what health conditions individuals experience once they leave the labor market and according to the timing of retirement.

Our findings also suggest future research to explore heterogeneity of the health effects in more detail. Because health effects of retirement are often ambiguous or heterogeneous, researchers should make more efforts to reveal that heterogeneity, which is helpful for the interpretation of the health consequences of later life events such as retirement. For instance, the availability of information on the industry or task in which workers were employed in their life may be particularly useful to drawing conclusions about the effects of retirement and its timing according to different mentally and physically-demanding job activities.

CONCLUSIONS

This thesis investigates the causal effects of nonemployment on future career dynamics and the impact of retirement on health outcomes. While the starting point consisted in approaching the related empirical fields through two extensive overviews and meta-analyses of the previous published studies, the empirical chapters made use of the AD-SILC database that allow to rebuild all the labor market history of the individuals included in the two econometric analyses. From a methodological point of view, the use of factor analytic dynamic models is aimed at i) disentangling the true causal effect of periods of nonemployment from the spurious one induced by systematic differences across individuals with different labor market histories, leading to the identification of the causal effects of youth nonemployment by taking into account a series of individual and time-varying unobserved factors related to personal characteristics and the socio-economic context; ii) estimating the causal health impact of retirement, occurring at different time across individuals, exploiting the exogenous shock of a pension reform which aims at increase the normal retirement age.

The thesis consists of two chapters related to the labor market entry issues, and further two chapters concerning the health implications of labor market exits through different timing of retirement. Thus, the first part provides i) an extensive overview and meta-analysis of empirical studies which adopt causal estimation strategies in evaluating the unemployment scarring effects; ii) an empirical analysis strictly focused on the short- and long-term effects of youth nonemployment experienced after high school completion on subsequent labor market performances. First, previous empirical evidence reveals significant earning losses and lower probabilities of employment following unemployment spells. These penalties are greater in case of youth nonemployment or lay-offs, but further sources of heterogeneities in the magnitude of the scarring effects concern, for example, age, tenure, and education level. Second, the negative effects of youth nonemployment after school completion are very persistent in terms of earnings; they are still sizable and statistically significant 25 years after diploma. As for labor market participation, measured as the fraction of days spent at work in a year, it is negatively affected by early nonemployment for a shorter span, as it disappears for both men and women by the 10th year after the school completion. These findings suggest that those individuals who randomly experienced nonemployment after school completion were able to get reintegrated after a while, but in a downgraded track; given that people experiencing early nonemployment send a worse signal, accumulate less human capital relatively to their employed peers, and are more likely to face liquidity constraints, they could lower their reservation wage and be more likely to accept worse jobs, characterized by a career track of lower profile, which traps them in lower wages and lower chances of subsequent promotions. Furthermore, both chapters highlight the importance of the identification strategy: while the meta-analysis results reveal that the penalties are larger in case of identification strategies based on selection on observables, the empirical findings from the factor analytic model show that the model which does not control for time-varying unobserved traits tends to overestimate future labor market penalties both in terms of duration and magnitude.

These findings have a number of policy implications. Given that the exposure to early nonemployment generates persistent earnings scars and participation penalties shorter-lasting but still present, favoring work experience after school completion is a very urgent socioeconomic goal. The policy maker could confine these negative consequences operating through different channels. First, whereas the intensive use of particular form of temporary contracts, especially during bad economic times, should be discouraged to avoid any precariousness trap (Filomena and Picchio, 2022a), apprenticeships are effective ways for Italian younger workers to increase the probability of promotion to an open-ended contract (Picchio and Staffolani, 2019). Thus, the policy maker could favor training programs and apprenticeships for those who were exposed to early nonemployment so as to facilitate the recoup of general human capital, but even for older workers (Picchio and van Ours, 2013), to avoid the greater penalties faced by laid-off older employees who are characterized by more firm specific skills and less recent tasks. Second, the policy maker could intervene facilitating the match between employers and the youth who suffered early nonemployment, for example by *ad hoc* subsidies for hiring school-leavers with difficulties in making the school-to-work transition. Finally, to limit the lowering of the reservation wage and the acceptance of bad jobs in downgraded tracks, the welfare state could play a role: benefits and, to circumscribe moral hazard, monitoring job search behaviors, so as to guide the school leavers exposed to nonemployment towards more efficient and better quality job matches.

As mentioned above, the second part of this thesis focuses on the health effect of retirement and contributes to this strand of the economic and health literature by i) filling the gap due to the lack of a rigorous and extensive meta-analysis on the subject; ii) evaluating the causal effect of retirement and its timing on mortality on a sample of private employees in Italy. Using meta-regression techniques and after checking for the presence of publication bias, the average effect of retirement on health outcomes is very small and barely significant, under the assumption of a common true effect. Furthermore, findings from model averaging strategies suggest that retirement seems to slightly improve self-reported health and mental health, but having no choice about the timing of retirement, being involuntarily retired or being forced to continue working due to policy reforms which postpone the time of retirement might have some health damaging implications. These findings have important implications for public policy. Although the effect of retirement on health outcomes is in general very small in magnitude, policy makers should consider not only the financial sustainability of the pension system, but also the raising healthcare spending due to the negative impact of mandatory or involuntary retirement. Optimal welfare pension policies should ensure workers a greater degree of freedom in choosing whether to retire and the timing of their retirement.

In the fourth and last chapter I used a factor analytic model with dynamic selection into treatment, taking into account time-varying unobserved heterogeneity jointly affecting the retirement decision and subsequent health outcomes. I found that retirement does not significantly affect mortality; I only detected an asymmetrical effect of delaying retirement on men and women in the long-term. Although the effect of retirement on mortality is on average nil, we cannot conclude that the cessation of the working career has no impact on individuals' health. In particular, the most relevant finding from our analysis is the lack of unobserved heterogeneity. This may suggest that mortality is such an extreme event and then is not significantly affected both by the timing of retirement and by unobserved traits of the individuals. In contrast, these unobserved traits are likely to impact several health conditions which characterize individuals at different ages after retirement. This issue deserves future investigations by researchers, who will therefore focus on the effects of retirement on morbidity, i.e. what health conditions individuals experience once they leave the labor market and according to the timing of retirement.

Finally, as suggested by Kuhn (2018), there are reasons to suspect that the health effects of retirement are heterogeneous across several dimensions, such as different types of prior occupation (e.g. blue- vs. white-collar workers), different types of physically/mentally demanding previous jobs, time horizons, or health behaviours, which are only partially investigated in our paper. The thesis tried to shed light on whether retirement differently affects blue- and white-collar workers, as well as lower- and higher-educated ones. Therefore, it concludes with a research suggestion: future research should take these further dimensions into account to gain a clearer picture of the multifaceted nature of the effects of retirement on mortality and further measures of health.

Bibliography

Abebe, D. S. and C. Hyggen (2019). Moderators of unemployment and wage scarring during the transition to young adulthood: evidence from Norway. In *Negotiating Early Job Insecurity*. Edward Elgar Publishing.

Abolhassani, M. and R. Alessie (2013). Subjective well-being around retirement. *De Economist 161*(3), 349–366.

Adams, G. A., J. Prescher, T. A. Beehr, and L. Lepisto (2002). Applying work-role attachment theory to retirement decision-making. *The International Journal of Aging and Human Development 54*(2), 125–137.

Adascalitei, D. and C. P. Morano (2016). Drivers and effects of labour market reforms: Evidence from a novel policy compendium. *IZA Journal of Labor Policy 5*(1), 1–32.

Ahmad, N. (2014). State dependence in unemployment. *International Journal of Economics and Financial Issues 4*(1), 93.

Alderman, H. and E. M. King (1998). Gender differences in parental investment in education. *Structural Change and Economic Dynamics 9*(4), 453–468.

Allen, J. and F. M. Alpass (2020). Trajectories of material living standards, physical health and mental health under a universal pension. *Journal of Epidemiology and Community Health 74*(4), 362–368.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2016). Cashier or consultant? Entry labor market conditions, field of study, and career success. *Journal of Labor Economics 34*(S1), S361–S401.

Apouey, B. H., C. Guven, and C. Senik (2019). Retirement and unexpected health shocks. *Economics & Human Biology 33*, 116–123.

Ardito, C., R. Leombruni, D. Blane, and A. d'Errico (2020). To work or not to work? The effect of higher pension age on cardiovascular health. *Industrial Relations: A Journal of Economy and Society 59*(3), 399–434.

Arranz, J. M., M. A. Davia, and C. García-Serrano (2005). Labour market transitions and wage dynamics in Europe. Technical report, ISER Working Paper Series.

Arranz, J. M. and C. García-Serrano (2003). Non-employment and subsequent wage losses. Technical report, Instituto de Estudios Fiscales.

Arulampalam, W. (2001). Is unemployment really scarring? Effects of unemployment experiences on wages. *The Economic Journal 111*(475), 585–606.

Arulampalam, W., A. L. Booth, and M. P. Taylor (2000). Unemployment persistence. *Oxford Economic Papers 52*(1), 24–50.

Arulampalam, W., P. Gregg, and M. Gregory (2001). Introduction: unemployment scarring. *The Economic Journal 111*(475), F577–F584.

Ashenfelter, O., C. Harmon, and H. Oosterbeek (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics 6*(4), 453–470.

Åslund, O. and H. Grönqvist (2010). Family size and child outcomes: is there really no trade-off? *Labour Economics 17*(1), 130–139.

Atalay, K. and G. F. Barrett (2014). The causal effect of retirement on health: new evidence from Australian pension reform. *Economics Letters 125*(3), 392–395.

Atalay, K., G. F. Barrett, and A. Staneva (2019). The effect of retirement on elderly cognitive functioning. *Journal of Health Economics 66*, 37–53.

Atalay, K. and R. Zhu (2018). The effect of a wife's retirement on her husband's mental health. *Applied Economics 50*(43), 4606–4616.

Ayllón, S. (2013). Unemployment persistence: not only stigma but discouragement too. *Applied Economics Letters 20*(1), 67–71.

Ayllón, S., J. Valbuena, and A. Plum (2021). Youth unemployment and stigmatization over the business cycle in Europe. *Oxford Bulletin of Economics and Statistics*.

Baert, S. and D. Verhaest (2019). Unemployment or overeducation: which is a worse signal to employers? *De Economist 167*(1), 1–21.

Bamia, C., A. Trichopoulou, and D. Trichopoulos (2008). Age at retirement and mortality in a general population sample: the Greek EPIC study. *American Journal of Epidemiology 167*(5), 561–569.

Barban, N., X. De Luna, E. Lundholm, I. Svensson, and F. C. Billari (2020). Causal effects of the timing of life-course events: age at retirement and subsequent health. *Sociological Methods & Research 49*(1), 216–249.

Barrett, G. F. and M. Kecmanovic (2013). Changes in subjective well-being with retirement: assessing savings adequacy. *Applied Economics 45*(35), 4883–4893.

Bassanini, A. and E. Caroli (2015). Is work bad for health? The role of constraint versus choice. *Annals of Economics and Statistics 119/120*, 13–37.

Bauer, A. B. and R. Eichenberger (2018). Worsening workers' health by lowering retirement age: the malign consequences of a benign reform. *Journal of the Economics of Ageing 18*, 100296.

Baumann, I. (2016). *The debate about the consequences of job displacement*, pp. 1–33. Cham: Springer International Publishing.

Becker, G. (1975). *Human capital: a theoretical and empirical analysis, with special reference to education, Second Edition*. National Bureau of Economic Research, Inc.

Begg, C. and J. Berlin (1988). Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 151*(3), 419–463.

Behncke, S. (2012). Does retirement trigger ill health? *Health Economics 21*(3), 282–300.

Belloni, M., E. Meschi, and G. Pasini (2016). The effect on mental health of retiring during the economic crisis. *Health Economics 25*(S2), 126–140.

Bender, K. A. (2012). An analysis of well-being in retirement: the role of pensions, health, and 'voluntariness' of retirement. *The Journal of Socio-Economics 41*(4), 424–433.

Bertoni, M. and G. Brunello (2017). Pappa ante portas: the effect of the husband's retirement on the wife's mental health in Japan. *Social Science & Medicine 175*, 135–142.

Bhargava, A. (1991). Identification and panel data models with endogenous regressors. *The Review of Economic Studies 58*(1), 129–140.

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

Biewen, M. and S. Steffes (2010). Unemployment persistence: is there evidence for stigma effects? *Economics Letters 106*(3), 188–190.

Binh Tran, D. and V. Zikos (2019). The causal effect of retirement on health: understanding the mechanisms. *Australian Economic Review 52*(4), 427–446.

Birkelund, G. E., K. Heggebø, and J. Rogstad (2017). Additive or multiplicative disadvantage? The scarring effects of unemployment for ethnic minorities. *European Sociological Review 33*(1), 17–29.

Blake, H. and C. Garrouste (2019). Collateral effects of a pension reform in France. *Annals of Economics and Statistics 133*, 57–86.

Blake, J. (1981). Family size and the quality of children. *Demography 18*(4), 421–442.

Blake, J. (1989). Number of siblings and educational attainment. *Science 245*(4913), 32–36.

Blane, D. (2006). The life course, the social gradient, and health. In M. Marmot and R. Wilkinson (Eds.), *Social determinants of health*, pp. 54–77. Oxford: Oxford University Press.

Bloemen, H., S. Hochguertel, and J. Zweerink (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *Health Economics 26*(12), e204–e218.

Bloemen, H., S. Hochguertel, and J. Zweerink (2019). The effect of incentive-induced retirement on spousal retirement rates: evidence from a natural experiment. *Economic Inquiry 57*(2), 910–930.

Boeri, T. and J. van Ours (2021). *The economics of imperfect labor markets*. Princeton University Press.

Böheim, R. and M. P. Taylor (2002). The search for success: do the unemployed find stable employment? *Labour Economics 9*(6), 717–735.

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

Bonsang, E. and T. J. Klein (2012). Retirement and subjective well-being. *Journal of Economic Behavior & Organization 83*(3), 311–329.

Borland, J. (2020). Scarring effects: a review of Australian and international literature. *Australian Journal of Labour Economics 23*(2), 173–188.

Börsch-Supan, A. and M. Schuth (2014). Early retirement, mental health, and social networks. In *Discoveries in the Economics of Aging*, NBER Chapters, pp. 225–250. University of Chicago Press.

Bound, J. and T. Waidmann (2007). Estimating the health effects of retirement. Working Papers wp168, University of Michigan, Michigan Retirement Research Center.

Bozio, A., C. Garrouste, and E. Perdrix (2021). Impact of later retirement on mortality: evidence from France. *Health Economics 30*(5), 1178–1199.

Bratberg, E. and O. A. Nilsen (2000). Transitions from school to work and the early labour market experience. *Oxford Bulletin of Economics and Statistics 62*, 909–929.

Brockmann, H., R. Müller, and U. Helmert (2009). Time to retire–time to die? A prospective cohort study of the effects of early retirement on long-term survival. *Social Science & Medicine 69*(2), 160–164.

Brodeur, A., N. Cook, and A. Heyes (2020). Methods matter: p-hacking and publication bias in causal analysis in economics. *American Economic Review 110*(11), 3634–60.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg (2016). Star wars: the empirics strike back. *American Economic Journal: Applied Economics 8*(1), 1–32.

Brugiavini, A. (1999). Social security and retirement in Italy. In J. Gruber and D. A. Wise (Eds.), *Social security and retirement around the world*, pp. 181–237. University of Chicago Press.

Brugiavini, A. and F. Peracchi (2012). Health status, welfare programs participation, and labor force activity in Italy. In D. A. Wise (Ed.), *Social security programs and retirement around the world: historical trends in mortality and health, employment, and disability insurance participation and reforms*, pp. 175–216. University of Chicago Press.

Brunello, G. and S. Comi (2015). The side effect of pension reforms on the training of older workers. Evidence from Italy. *The Journal of the Economics of Ageing 6*, 113–122.

Burda, M. C. and A. Mertens (2001). Estimating wage losses of displaced workers in Germany. *Labour Economics 8*(1), 15–41.

Burdett, K. (1978). A theory of employee job search and quit rates. *The American Economic Review 68*(1), 212–220.

Burgess, S., C. Propper, H. Rees, and A. Shearer (2003). The class of 1981: the effects of early career unemployment on subsequent unemployment experiences. *Labour Economics 10*(3), 291–309.

Butterworth, P., S. C. Gill, B. Rodgers, K. J. Anstey, E. Villamil, and D. Melzer (2006). Retirement and mental health: analysis of the Australian national survey of mental health and well-being. *Social Science & Medicine 62*(5), 1179–1191.

Calvo, E., N. Sarkisian, and C. R. Tamborini (2013). Causal effects of retirement timing on subjective physical and emotional health. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences 68*(1), 73–84.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008a). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008b). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics 90*(3), 414–427.

Card, D. and A. B. Krueger (1995). Time-series minimum-wage studies: a meta-analysis. *The American Economic Review 85*(2), 238–243.

Carlsson, S., T. Andersson, K. Michaëlsson, D. Vågerö, and A. Ahlbom (2012). Late retirement is not associated with increased mortality, results based on all Swedish retirements 1991–2007. *European Journal of Epidemiology 27*(6), 483–486.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carone, G., P. Eckefeldt, L. Giamboni, V. Laine, and S. Pamies (2016). Pension reforms in the EU since the early 2000's: achievements and challenges ahead. European Economy Discussion Paper No. 42, Brussels.

Carrino, L., K. Glaser, and M. Avendano (2020). Later retirement, job strain, and health: evidence from the new state pension age in the United Kingdom. *Health Economics 29*(8), 891–912.

Celidoni, M., C. Dal Bianco, and G. Weber (2017). Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of Health Economics 56*, 113–125.

Celidoni, M. and V. Rebba (2017). Healthier lifestyles after retirement in Europe? Evidence from SHARE. *The European Journal of Health Economics 18*(7), 805–830.

Chamberlain, G. (1984). Panel data. *Handbook of Econometrics 2*, 1247–1318.

Charles, K. K. (2004). Is retirement depressing?: Labor force inactivity and psychological well-being in later life. In S. E. Polachek (Ed.), *Accounting for worker well-being*, Volume 23 of *Research in Labor Economics*, pp. 269–299. Emerald Group Publishing Limited.

Che, Y. and X. Li (2018). Retirement and health: evidence from China. *China Economic Review 49*, 84–95.

Chung, S., M. E. Domino, and S. C. Stearns (2009). The effect of retirement on weight. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences 64*(5), 656–665.

Churchill, A. and V. Mishra (2018). Returns to education in China: a meta-analysis. *Applied Economics 50*(54), 5903–5919.

Clark, A. E. and A. Lepinteur (2019). The causes and consequences of early-adult unemployment: evidence from cohort data. *Journal of Economic Behavior & Organization 166*, 107–124.

Cockx, B. (2016). Do youths graduating in a recession incur permanent losses? *IZA World of Labor*.

Cockx, B. and M. Picchio (2012). Are short-lived jobs stepping stones to long-lasting jobs? *Oxford Bulletin of Economics and Statistics 74*(5), 646–675.

Cockx, B. and M. Picchio (2013). Scarring effects of remaining unemployed for long-term unemployed school-leavers. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 176*(4), 951–980.

Cockx, B., M. Picchio, and S. Baert (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics 34*(3), 403–424.

Coe, N. B. and M. Lindeboom (2008). Does retirement kill you? Evidence from early retirement windows. IZA Discussion Paper No. 3817, Institute of Labor Economics (IZA), Bonn, Germany.

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

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

Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.

Corcoran, M. (1982). The employment and wage consequences of teenage women's nonemployment. In *The youth labor market problem: its nature, causes, and consequences*, pp. 391–426. University of Chicago Press.

Couch, K. A. (2001). Earnings losses and unemployment of displaced workers in Germany. *ILR Review 54*(3), 559–572.

Cullen, M. (1999). Personal reflections on occupational health in the twentieth century: spiraling to the future. *Annual Review of Public Health 20*, 1–13.

Dave, D., R. I. Rashad, and J. Spasojevic (2008). The effects of retirement on physical and mental health outcomes. *Southern Economic Journal 75*(2), 497–523.

Dayaram, K. and A. McGuire (2019). Retirement reforms: occupational strain and health. *Industrial Relations: A Journal of Economy and Society 58*(3), 522–542.

De Fraja, G., S. Lemos, and J. Rockey (2021). The wounds that do not heal: The lifetime scar of youth unemployment. *Economica 88*(352), 896–941.

De Luca, G. and J. R. Magnus (2011). Bayesian model averaging and weighted-average least squares: equivariance, stability, and numerical issues. *Stata Journal 11*(4), 518–544.

Deelen, A., M. de Graaf-Zijl, and W. van den Berge (2018). Labour market effects of job displacement for prime-age and older workers. *IZA Journal of Labor Economics 7*(1), 3.

Dieckhoff, M. (2011). The effect of unemployment on subsequent job quality in Europe: a comparative study of four countries. *Acta Sociologica 54*(3), 233–249.

Doiron, D. and T. Gørgens (2008). State dependence in youth labor market experiences, and the evaluation of policy interventions. *Journal of Econometrics 145*(1-2), 81–97.

Dorsett, R. and P. Lucchino (2018). Young people's labour market transitions: the role of early experiences. *Labour Economics 54*, 29–46.

Doucouliagos, C. (1995). Worker participation and productivity in labor-managed and participatory capitalist firms: a meta-analysis. *ILR Review 49*(1), 58–77.

Doucouliagos, H. (2011). How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Working Papers SWP 2011/5, Deakin University, Department of Economics.

Doucouliagos, H. and P. Laroche (2009). Unions and profits: a meta-regression analysis. *Industrial Relations: A Journal of Economy and Society 48*(1), 146–184.

Dufouil, C., E. Pereira, G. Chêne, M. M. Glymour, A. Alpérovitch, E. Saubusse, M. Risse-Fleury, B. Heuls, J.-C. Salord, M.-A. Brieu, et al. (2014). Older age at retirement is associated with decreased risk of dementia. *European Journal of Epidemiology 29*(5), 353–361.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal 315*(7109), 629–634.

Eibich, P. (2015). Understanding the effect of retirement on health: mechanisms and heterogeneity. *Journal of Health Economics 43*, 1–12.

Eicher, T. S., C. Papageorgiou, and A. E. Raftery (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics 26*(1), 30–55.

Einmahl, J., J. Magnus, and K. Kumar (2011). On the choice of prior in Bayesian model averaging. Discussion Paper 2011-003, Tilburg University, Center for Economic Research.

Eliason, M. and D. Storrie (2006). Lasting or latent scars? Swedish evidence on the long-term effects of job displacement. *Journal of Labor Economics 24*(4), 831–856.

Ellwood, D. T. (1982). Teenage unemployment: permanent scars or temporary blemishes? In *The youth labor market problem: its nature, causes, and consequences*, pp. 349–390. University of Chicago Press.

Eriksson, S. and D.-O. Rooth (2014). Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *The American Economic Review 104*(3), 1014–39.

Evenson, K. R., W. D. Rosamond, J. Cai, A. V. Diez-Roux, and F. L. Brancati for the Atherosclerosis Risk in Communities Study Investigators (2002). Influence of retirement on leisure-time physical activity: the atherosclerosis risk in communities study. *American Journal of Epidemiology 155*(8), 692–699.

Eyjólfsdóttir, H. S., I. Baumann, N. Agahi, J. Fritzell, and C. Lennartsson (2019). Prolongation of working life and its effect on mortality and health in older adults: propensity score matching. *Social Science & Medicine 226*, 77–86.

Fallick, B. C. (1996). A review of the recent empirical literature on displaced workers. *ILR Review 50*(1), 5–16.

Farber, H. S., C. M. Herbst, D. Silverman, and T. Von Wachter (2019). Whom do employers want? The role of recent employment and unemployment status and age. *Journal of Labor Economics 37*(2), 323–349.

Farber, H. S., D. Silverman, and T. Von Wachter (2016). Determinants of callbacks to job applications: An audit study. *American Economic Review 106*(5), 314–18.

Farber, H. S., D. Silverman, and T. M. Von Wachter (2017). Factors determining callbacks to job applications by the unemployed: An audit study. *RSF: The Russell Sage Foundation Journal of the Social Sciences 3*(3), 168–201.

Fé, E. and B. P. Hollingsworth (2016). Short-and long-run estimates of the local effects of retirement on health. *Journal of the Royal Statistical Society: Series A Statistics in Society 179*(4), 1051–1067.

Feng, J., Q. Li, and J. P. Smith (2020). Retirement effect on health status and health behaviors in urban China. *World Development 126*, 104702.

Filomena, M. and M. Picchio (2022a). Are temporary jobs stepping stones or dead ends? A systematic review of the literature. *International Journal of Manpower 43*(9), 60–74.

Filomena, M. and M. Picchio (2022b). Retirement and health outcomes in a meta-analytical framework. *Journal of Economic Surveys* (forthcoming).

Finkel, D., R. Andel, M. Gatz, and N. L. Pedersen (2009). The role of occupational complexity in trajectories of cognitive aging before and after retirement. *Psychology and Aging 24*(3), 563–573.

Fisher, G. G., A. Stachowski, F. J. Infurna, J. D. Faul, J. Grosch, and L. E. Tetrick (2014). Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of Occupational Health Psychology 19*(2), 231–242.

Fitzpatrick, M. D. and T. J. Moore (2018). The mortality effects of retirement: evidence from social security eligibility at age 62. *Journal of Public Economics 157*, 121–137.

Fonseca, R., A. Kapteyn, J. Lee, G. Zamarro, and K. Feeney (2014). A longitudinal study of well-being of older Europeans: does retirement matter? *Journal of Population Ageing 7*(1), 21–41.

Frimmel, W. and G. J. Pruckner (2020). Retirement and healthcare utilization. *Journal of Public Economics 184*, 104146.

Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics 34*(4), 979–1021.

Gangji, A. and R. Plasman (2007). The matthew effect of unemployment: how does it affect wages in Belgium. DULBEA Working Papers 07-19.RS, ULB – Universite Libre de Bruxelles.

Gangji, A. and R. Plasman (2008). Microeconomic analysis of unemployment persistence in Belgium. *International Journal of Manpower 29*(3), 280–298.

Gangl, M. (2004). Welfare states and the scar effects of unemployment: a comparative analysis of the United States and West Germany. *American Journal of Sociology 109*(6), 1319–1364.

Gangl, M. (2006). Scar effects of unemployment: an assessment of institutional complementarities. *American Sociological Review 71*(6), 986–1013.

Garrouste, C. and E. Perdrix (2022). Is there a consensus on the health consequences of retirement? A literature review. *Journal of Economic Surveys 36*(4), 841–879.

Gartell, M. (2009). Unemployment and subsequent earnings for Swedish college graduates. A study of scarring effects. Arbetsrapport 2009:2, Institute for Futures Studies.

Gaure, S., K. Røed, and T. Zhang (2007). Time and causality: a Monte Carlo assessment of the timing-of-events approach. *Journal of Econometrics 141*(2), 1159–1195.

Gaure, S., K. Røed, and L. Westlie (2008). The impacts of labor market policies on job search behavior and post-unemployment job quality. IZA Discussion Papers 3802, Institute of Labor Economics (IZA).

Ghirelli, C. (2015). Scars of early non-employment for low educated youth: evidence and policy lessons from Belgium. *IZA Journal of European Labor Studies 4*(1), 20.

Gibbons, R. and L. F. Katz (1991). Layoffs and lemons. *Journal of Labor Economics 9*(4), 351–380.

Gill, S. C., P. Butterworth, B. Rodgers, K. J. Anstey, E. Villamil, and D. Melzer (2006). Mental health and the timing of men's retirement. *Social Psychiatry and Psychiatric Epidemiology 41*(7), 515–522.

Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. *Journal of Health Economics 45*, 27–46.

Gorry, A., D. Gorry, and S. N. Slavov (2018). Does retirement improve health and life satisfaction? *Health Economics 27*(12), 2067–2086.

Gregg, P. (2001). The impact of youth unemployment on adult unemployment in the NCDS. *The Economic Journal 111*(475), F626–F653.

Gregg, P. and E. Tominey (2005). The wage scar from male youth unemployment. *Labour Economics 12*(4), 487–509.

Gregory, M. and R. Jukes (2001). Unemployment and subsequent earnings: estimating scarring among British men 1984–94. *Economic Journal 111*(475), 607–625.

Grip, A. d., M. Lindeboom, and R. Montizaan (2012). Shattered dreams: the effects of changing the pension system late in the game. *Economic Journal 122*(559), 1–25.

Grøtting, M. W. and O. S. Lillebø (2020). Health effects of retirement: evidence from survey and register data. *Journal of Population Economics 33*(2), 671–704.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2017). Heterogeneous scarring effects of full-year nonemployment. *American Economic Review 107*(5), 369–73.

Görg, H. and E. Strobl (2001). Multinational companies and productivity spillovers: a meta-analysis. *The Economic Journal 111*(475), F723–F739.

Hagen, J. (2018). The effects of increasing the normal retirement age on health care utilization and mortality. *Journal of Population Economics 31*(1), 193–234.

Hallberg, D., P. Johansson, and M. Josephson (2015). Is an early retirement offer good for your health? Quasi-experimental evidence from the army. *Journal of Health Economics 44*, 274–285.

Hällsten, M. (2017). Is education a risky investment? The scarring effect of university dropout in Sweden. *European Sociological Review 33*(2), 169–181.

Hämäläinen, K. (2003). Education and unemployment: state dependence in unemployment among young people in the 1990s.

Hamermesh, D. S. (1989). What do we know about worker displacement in the US? *Industrial Relations: A Journal of Economy and Society 28*(1), 51–59.

Harrison, J. and L. Dawson (2016). Occupational health: meeting the challenges of the next 20 years. *Safety and Health at Work 7*(2), 143–149.

Havranek, T., R. Horvath, Z. Irsova, and M. Rusnak (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics 96*(1), 100–118.

Havránek, T., T. D. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, and R. C. M. van Aert (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys 34*(3), 469–475.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, 153–161.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Heckman, J. J. and J. A. Smith (1998). Evaluating the welfare state. Working Paper 6542, National Bureau of Economic Research.

Helbling, L. A. and S. Sacchi (2014). Scarring effects of early unemployment among young workers with vocational credentials in Switzerland. *Empirical Research in Vocational Education and Training 6*(1), 12.

Heller-Sahlgren, G. (2017). Retirement blues. *Journal of Health Economics 54*, 66–78.

Henkens, K., H. van Solinge, and W. T. Gallo (2008). Effects of retirement voluntariness on changes in smoking, drinking and physical activity among Dutch older workers. *The European Journal of Public Health 18*(6), 644–649.

Hernaes, E., S. Markussen, J. Piggott, and O. L. Vestad (2013). Does retirement age impact mortality? *Journal of Health Economics 32*(3), 586–598.

Hessel, P. (2016). Does retirement (really) lead to worse health among European men and women across all educational levels? *Social Science & Medicine 151*, 19–26.

Heylen, V. (2011). Scarring, the effects of early career unemployment.

Hofäcker, D., M. Hess, and E. Naumann (2015). Changing retirement transitions in times of paradigmatic political change: toward growing inequalities? In C. Torp (Ed.), *Challenges of aging*, pp. 205–226. Houndmills, UK: Palgrave Macmillan.

Hofäcker, D. and E. Naumann (2015). The emerging trend of work beyond retirement age in Germany. *Zeitschrift für Gerontologie und Geriatrie 48*(5), 473–479.

Horner, E. M. (2014). Subjective well-being and retirement: analysis and policy recommendations. *Journal of Happiness Studies 15*(1), 125–144.

Horner, E. M. and M. R. Cullen (2016). The impact of retirement on health: quasi-experimental methods using administrative data. *BMC Health Services Research 16*(1), 1–9.

Hult, C., M. Stattin, U. Janlert, and B. Järvholm (2010). Timing of retirement and mortality-a cohort study of Swedish construction workers. *Social Science & Medicine 70*(10), 1480–1486.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources 49*(1), 195–233.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review*, 685–709.

Johnston, D. W. and W.-S. Lee (2009). Retiring to the good life? The short-term effects of retirement on health. *Economics Letters 103*(1), 8–11.

Jokela, M., J. E. Ferrie, D. Gimeno, T. Chandola, M. J. Shipley, J. Head, J. Vahtera, H. Westerlund, M. G. Marmot, and M. Kivimäki (2010). From midlife to early old age: health trajectories associated with retirement. *Epidemiology (Cambridge, Mass.) 21*(3), 284–290.

Jovanovic, B. (1979a). Firm-specific capital and turnover. *Journal of Political Economy 87*(6), 1246–1260.

Jovanovic, B. (1979b). Job matching and the theory of turnover. *Journal of Political Economy 87*(5), 972–990.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303–316.

Kajitani, S., K. Sakata, and C. McKenzie (2017). Occupation, retirement and cognitive functioning. *Ageing and Society 37*(8), 1568–1596.

Kalwij, A., R. Alessie, and M. Knoef (2013). Pathways to retirement and mortality risk in the Netherlands. *European Journal of Population/Revue Européenne de Démographie 29*(2), 221–238.

Karhula, A., H. Lehti, and J. Erola (2017). Intergenerational scars? The long-term effects of parental unemployment during a depression. *Research on Finnish Society 10*, 87–99.

Kawaguchi, D. and T. Murao (2014). Labor-market institutions and long-term effects of youth unemployment. *Journal of Money, Credit and Banking 46*(S2), 95–116.

Keef, S. P. and L. A. Roberts (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology 57*(1), 97–129.

Kesavayuth, D., R. E. Rosenman, and V. Zikos (2016). Retirement, personality, and well-being. *Economic Inquiry 54*(2), 733–750.

Kim, J. E. and P. Moen (2002). Retirement transitions, gender, and psychological well-being: a life-course, ecological model. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 57*(3), P212–P222.

Kim, S. and K. Koh (2020). Does early access to pension wealth improve health? *Economic Inquiry 58*(4), 1783–1794.

Kletzer, L. G. (1998). Job displacement. *Journal of Economic Perspectives 12*(1), 115–136.

Kletzer, L. G. and R. W. Fairlie (2003). The long-term costs of job displacement for young adult workers. *ILR Review 56*(4), 682–698.

Knights, S., M. N. Harris, and J. Loundes (2002). Dynamic relationships in the Australian labour market: heterogeneity and state dependence. *Economic Record 78*(242), 284–298.

Kolodziej, I. W. and P. García-Gómez (2019). Saved by retirement: beyond the mean effect on mental health. *Social Science & Medicine 225*, 85–97.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). Duration dependence and labor market conditions: evidence from a field experiment. *The Quarterly Journal of Economics 128*(3), 1123–1167.

Kuchibhotla, M., P. F. Orazem, and S. Ravi (2020). The scarring effects of youth joblessness in Sri Lanka. *Review of Development Economics 24*(1), 269–287.

Kuhn, A. (2018). The complex effects of retirement on health. *IZA World of Labor 430*.

Kuhn, A., S. Staubli, J.-P. Wuellrich, and J. Zweimüller (2020). Fatal attraction? Extended unemployment benefits, labor force exits, and mortality. *Journal of Public Economics 191*, 104087.

Kühntopf, S. and T. Tivig (2012). Early retirement and mortality in Germany. *European Journal of Epidemiology 27*(2), 85–89.

Kuusi, T., P. Martikainen, and T. Valkonen (2020). The influence of old-age retirement on health: causal evidence from the Finnish register data. *Journal of the Economics of Ageing 17*, 100257.

Lalive, R. and S. Staubli (2015). How does raising women's full retirement age affect labor supply, income, and mortality? Evidence from Switzerland. Working Paper NB 14–09, NBER Retirement Research Center.

Latif, E. (2011). The impact of retirement on psychological well-being in Canada. *Journal of Socio-Economics 40*(4), 373–380.

Lazear, E. (1986). Raids and offer-matching. In R. Ehrenberg (Ed.), *Research in Labor Economics*. JAI Press, Greenwich, CT.

Lee, J. and J. P. Smith (2009). Work, retirement, and depression. *Journal of Population Ageing 2*(1-2), 57–71.

Lei, X. and H. Liu (2018). Gender difference in the impact of retirement on cognitive abilities: evidence from urban China. *Journal of Comparative Economics 46*(4), 1425–1446.

Li, H. and M. Hiwatari (2020). Family size and educational attainment: the case of China. Faculty of Economics and Business, Hokkaido University, Discussion Paper Series A, 353.

Li, W., X. Ye, D. Zhu, and P. He (2021). The longitudinal association between retirement and depression: a systematic review and meta-analysis. *American Journal of Epidemiology*.

Light, R. and D. Pillemer (1984). *Summing up. The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lipsey, M. W. and D. B. Wilson (2001). *Practical meta-analysis*. SAGE publications, Inc.

Litwin, H. (2007). Does early retirement lead to longer life? *Ageing and Society 27*(5), 739.

Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *The Review of Economic Studies 58*(4), 733–753.

Lucifora, C. and D. Vigani (2018). Health care utilization at retirement: the role of the opportunity cost of time. *Health Economics 27*(12), 2030–2050.

Lupi, C., P. Ordine, et al. (2002). Unemployment scarring in high unemployment regions. *Economics Bulletin 10*(2), 1–8.

Magnus, J. R. and G. De Luca (2016). Weighted-average least squares (WALS): a survey. *Journal of Economic Surveys 30*(1), 117–148.

Magnus, J. R., O. Powell, and P. Prüfer (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics 154*(2), 139–153.

Mandal, B. and B. E. Roe (2007). Job loss, retirement and the mental health of older Americans. *Journal of Mental Health Policy and Economics 11*(4), 167–176.

Manzoni, A. and I. Mooi-Reci (2011). Early unemployment and subsequent career complexity: a sequence-based perspective. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften 131*(2), 339–348.

Mavromaras, K., P. Sloane, and Z. Wei (2015). The scarring effects of unemployment, low pay and skills under-utilization in Australia compared. *Applied Economics 47*(23), 2413–2429.

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

Mazzonna, F. and F. Peracchi (2017). Unhealthy retirement? *Journal of Human Resources 52*(1), 128–151.

Mein, G., P. Martikainen, H. Hemingway, S. Stansfeld, and M. Marmot (2003). Is retirement good or bad for mental and physical health functioning? Whitehall II longitudinal study of civil servants. *Journal of Epidemiology and Community Health 57*(1), 46–49.

Messe, P.-J. and F.-C. Wolff (2019a). Healthier when retiring earlier? Evidence from France. *Applied Economics 51*(47), 5122–5143.

Messe, P.-J. and F.-C. Wolff (2019b). The short-term effects of retirement on health within couples: evidence from France. *Social Science & Medicine 221*, 27–39.

Mincer, J. (1974). *Schooling, experience, and earnings*. National Bureau of Economic Research, Inc.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Medicine 6*(7), 1–6.

Mojon-Azzi, S., A. Sousa-Poza, and R. Widmer (2007). The effect of retirement on health: a panel analysis using data from the Swiss Household Panel. *Swiss Medical Weekly 137*(4142), 581–585.

Möller, J. and M. Umkehrer (2015). Are there long-term earnings scars from youth unemployment in Germany? *Jahrbücher für Nationalökonomie und Statistik 235*(4-5), 474–498.

Mooi-Reci, I. and H. B. Ganzeboom (2015). Unemployment scarring by gender: human capital depreciation or stigmatization? Longitudinal evidence from the Netherlands, 1980–2000. *Social Science Research 52*, 642–658.

Mortensen, D. (1987). Job search and labor market analysis. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 2, Chapter 15, pp. 849–919.

Mortensen, D. T. (1988). Wages, separations, and job tenure: on-the-job specific training or matching? *Journal of Labor Economics 6*(4), 445–471.

Mosca, I. and A. Barrett (2016). The impact of voluntary and involuntary retirement on mental health: Evidence from older Irish adults. *Journal of Mental Health Policy and Economics 19*(1), 33–44.

Motegi, H., Y. Nishimura, and M. Oikawa (2020). Retirement and health investment behaviors: an international comparison. *Journal of the Economics of Ageing 16*, 100267.

Mousteri, V., M. Daly, and L. Delaney (2018). The scarring effect of unemployment on psychological well-being across Europe. *Social Science Research 72*, 146–169.

Mroz, T. A. and T. H. Savage (2006). The long-term effects of youth unemployment. *Journal of Human Resources 41*(2), 259–293.

Müller, T. and M. Shaikh (2018). Your retirement and my health behavior: evidence on retirement externalities from a fuzzy regression discontinuity design. *Journal of Health Economics 57*, 45–59.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, 69–85.

Naticchioni, P., M. Raitano, and C. Vittori (2016). La meglio gioventù: earnings gaps across generations and skills in Italy. *Economia Politica 33*(2), 233–264.

Nekoei, A. and A. Weber (2017). Does extending unemployment benefits improve job quality? *American Economic Review 107*(2), 527–61.

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

Nickell, S., P. Jones, and G. Quintini (2002). A picture of job insecurity facing British men. *The Economic Journal 112*(476), 1–27.

Nielsen, N. F. (2019). Sick of retirement? *Journal of Health Economics 65*, 133–152.

Nilsen, Ø. A. and K. H. Reiso (2014). Scarring effects of early-career unemployment. *Nordic Economic Policy Review 1*(2014), 13–46.

Nishimura, Y., M. Oikawa, and H. Motegi (2018). What explains the difference in the effect of retirement on health? Evidence from global aging data. *Journal of Economic Surveys 32*(3), 792–847.

Nordström Skans, O. (2011). Scarring effects of the first labor market experience. IZA Discussion Papers 5565, Institute of Labor Economics (IZA).

Nunley, J. M., A. Pugh, N. Romero, and R. A. Seals (2017). The effects of unemployment and underemployment on employment opportunities: Results from a correspondence audit of the labor market for college graduates. *ILR Review 70*(3), 642–669.

Nüß, P. (2018). Duration dependence as an unemployment stigma: Evidence from a field experiment in Germany. Technical report, Economics Working Paper.

Oberholzer-Gee, F. (2008). Nonemployment stigma as rational herding: a field experiment. *Journal of Economic Behavior & Organization 65*(1), 30–40.

OECD (2017). *Pensions at a Glance 2017: OECD and G20 Indicators*. Paris: OECD Publishing.

OECD (2019). *Pensions at a Glance 2019: OECD and G20 Indicators*. Paris: OECD Publishing.

Oksanen, T., J. Vahtera, H. Westerlund, J. Pentti, N. Sjösten, M. Virtanen, I. Kawachi, and M. Kivimäki (2011). Is retirement beneficial for mental health? Antidepressant use before and after retirement. *Epidemiology 22*(4), 553–559.

Olesen, K., R. Rugulies, N. H. Rod, and J. P. Bonde (2014). Does retirement reduce the risk of myocardial infarction? A prospective registry linkage study of 617 511 Danish workers. *International Journal of Epidemiology 43*(1), 160–167.

Omori, Y. (1997). Stigma effects of nonemployment. *Economic Inquiry 35*(2), 394–416.

Ordine, P. and G. Rose (2015). Educational mismatch and unemployment scarring. *International Journal of Manpower*.

Oreopoulos, P., T. Von Wachter, and A. Heisz (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Oshio, T. and M. Kan (2017). The dynamic impact of retirement on health: evidence from a nationwide ten-year panel survey in Japan. *Preventive Medicine 100*, 287–293.

Pabón-Carrasco, M., L. Ramirez-Baena, R. López Sánchez, I. Rodríguez-Gallego, N. Suleiman-Martos, and J. L. Gómez-Urquiza (2020). Prevalence of depression in retirees: a meta-analysis. *Healthcare 8*(3).

Pastore, F., C. Quintano, and A. Rocca (2020). Stuck at a crossroads? The duration of the Italian school-to-work transition. *International Journal of Manpower*.

Pastore, F., C. Quintano, and A. Rocca (2021). Some young people have all the luck! The duration dependence of the school-to-work transition in Europe. *Labour Economics 70*, 101982.

Patrinos, H. A. and G. Psacharopoulos (1997). Family size, schooling and child labor in Peru: an empirical analysis. *Journal of Population Economics 10*(4), 387–405.

Petreski, M., N. Mojsoska-Blazevski, and M. Bergolo (2017). Labor-market scars when youth unemployment is extremely high: evidence from Macedonia. *Eastern European Economics 55*(2), 168–196.

Picchio, M. (2022). Meta-analysis. In *Handbook of Labor, Human Resources and Population Economics, Program Evaluation*. Springer Cham.

Picchio, M., C. Pigini, S. Staffolani, and A. Verashchagina (2021). If not now, when? The timing of childbirth and labor market outcomes. *Journal of Applied Econometrics 36*(6), 663–685.

Picchio, M. and S. Staffolani (2019). Does apprenticeship improve job opportunities? a regression discontinuity approach. *Empirical Economics 56*(1), 23–60.

Picchio, M. and J. C. van Ours (2013). Retaining through training even for older workers. *Economics of Education Review 32*, 29–48.

Picchio, M. and J. C. van Ours (2020). Mental health effects of retirement. *De Economist 168*(3), 419–452.

Pilipiec, P., W. Groot, and M. Pavlova (2020). The effect of an increase of the retirement age on the health, well-being, and labor force participation of older workers: a systematic literature review. *Journal of Population Ageing*, forthcoming.

Pissarides, C. A. (1992). Loss of skill during unemployment and the persistence of employment shocks. *The Quarterly Journal of Economics 107*(4), 1371–1391.

Plum, A. and S. Ayllón (2015). Heterogeneity in unemployment state dependence. *Economics Letters 136*, 85–87.

Quaade, T., G. Engholm, A. M. T. Johansen, and H. Møller (2002). Mortality in relation to early retirement in Denmark: a population-based study. *Scandinavian Journal of Public Health 30*(3), 216–222.

Raaum, O. and K. Røed (2006). Do business cycle conditions at the time of labor market entry affect future employment prospects? *The Review of Economics and Statistics 88*(2), 193–210.

Radl, J. (2013). Labour market exit and social stratification in Western Europe: the effects of social class and gender on the timing of retirement. *European Sociological Review 29*(3), 654–668.

Raitano, M. and M. Fana (2019). Labour market deregulation and workers' outcomes at the beginning of the career: evidence from Italy. *Structural Change and Economic Dynamics 51*, 301–310.

Reed, W. (2020). A note on the use of partial correlation coefficients in meta-analyses. Working paper no. 8/2020, Department of Economics and Finance, University of Canterbury, New Zealand.

Rijs, K. J., R. Cozijnsen, and D. J. Deeg (2012). The effect of retirement and age at retirement on self-perceived health after three years of follow-up in Dutch 55-64-year-olds. *Ageing and Society 32*(2), 281–306.

Roberts, B. A., R. Fuhrer, M. Marmot, and M. Richards (2011). Does retirement influence cognitive performance? The Whitehall II study. *Journal of Epidemiology and Community Health 65*(11), 958–963.

Rodríguez, J., J. Colston, Z. Wu, and Z. Chen (2020). Graduating during a recession: a literature review of the effects of recessions for college graduates. Technical report, Center for Research on CollegeWorkforce Transitions (CCWT), University of Wisconsin Madison. October.

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

Rose, L. (2020). Retirement and health: evidence from England. *Journal of Health Economics 73*, 102352.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin 86*(3), 638–641.

Ruhm, C. J. (1991). Are workers permanently scarred by job displacements? *The American Economic Review 81*(1), 319–324.

Saraceno, C. (1994). The ambivalent familism of the Italian welfare state. *Social Politics 1*(1), 60–82.

Schmillen, A. and M. Umkehrer (2017). The scars of youth: effects of early-career unemployment on future unemployment experience. *International Labour Review 156*(3-4), 465–494.

Sewdas, R., A. De Wind, S. Stenholm, P. Coenen, I. Louwerse, C. Boot, and A. Van Der Beek (2020). Association between retirement and mortality: working longer, living longer? A systematic review and meta-analysis. *Journal of Epidemiology and Community Health 74*(5), 473–480.

Shai, O. (2018). Is retirement good for men's health? Evidence using a change in the retirement age in Israel. *Journal of Health Economics 57*, 15–30.

Shi, L. P. and S. Wang (2021). Demand-side consequences of unemployment and horizontal skill mismatches across national contexts: An employer-based factorial survey experiment. *Social Science Research*, 102668.

Son, C.-H. et al. (2020). The effects of retirement on health-related quality of life of retirement-aged adults for four marital status subgroups. *International Journal of Economic Sciences 9*(1), 179–201.

Spence, A. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Spivey, C. (2005). Time off at what price? The effects of career interruptions on earnings. *ILR Review 59*(1), 119–140.

Stanley, T. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics 70*(1), 103–127.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys 19*(3), 309–345.

Stanley, T. D. and H. Doucouliagos (2012). *Meta-regression analysis in economics and business*. Abingdon: Routledge.

Stanley, T. D. and H. Doucouliagos (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods 5*(1), 60–78.

Stanley, T. D., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, et al. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys 27*(2), 390–394.

Sterling, T. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. *Journal of the American Statistical Association 54*(285), 30–34.

Stevens, A. H. (1997). Persistent effects of job displacement: the importance of multiple job losses. *Journal of Labor Economics 15*(1, Part 1), 165–188.

Stewart, M. B. (2007). The interrelated dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics 22*(3), 511–531.

Strandh, M., A. Winefield, K. Nilsson, and A. Hammarström (2014). Unemployment and mental health scarring during the life course. *European Journal of Public Health 24*(3), 440–445.

Sutton, A. J., K. R. Abrams, D. R. Jones, D. R. Jones, T. A. Sheldon, and F. Song (2000). *Methods for meta-analysis in medical research*. Chichester: Wiley.

Syse, A., M. Veenstra, T. Furunes, R. J. Mykletun, and P. E. Solem (2017). Changes in health and health behavior associated with retirement. *Journal of Aging and Health 29*(1), 99–127.

Tanzi, G. M. (2022). Scars of youth non-employment and labour market conditions. *Italian Economic Journal* (forthcoming), 1–25.

Tatsiramos, K. (2009). Unemployment insurance in Europe: unemployment duration and subsequent employment stability. *Journal of the European Economic Association 7*(6), 1225–1260.

Tsai, S. P., J. K. Wendt, R. P. Donnelly, G. De Jong, and F. S. Ahmed (2005). Age at retirement and long term survival of an industrial population: prospective cohort study. *British Medical Journal 331*(7523), 995.

Tumino, A. (2015). The scarring effect of unemployment from the early'90s to the great recession. Technical report, ISER Working Paper Series.

Ugur, M. (2014). Corruption's direct effects on per-capita income growth: a meta-analysis. *Journal of Economic Surveys 28*(3), 472–490.

van der Heide, I., R. M. van Rijn, S. J. Robroek, A. Burdorf, and K. I. Proper (2013). Is retirement good for your health? A systematic review of longitudinal studies. *BMC Public Health 13*(1), 1–11.

van Mourik, C. (2020). A meta-analysis into the causal effect of retirement on health. Netspar Academic Series, Thesis for the Master Public Administration, Leiden University.

van Ours, J. C. (2022). How retirement affects mental health, cognitive skills and mortality; an overview of recent empirical evidence. *De Economist 170*(3), 375–400.

Verho, J. (2008). Scars of recession: the long-term costs of the Finnish economic crisis. Working Paper Series 2008:9, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Viechtbauer, W. and M. Cheung (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Method 1*(2), 112–125.

Vishwanath, T. (1989). Job search, stigma effect, and escape rate from unemployment. *Journal of Labor Economics 7*(4), 487–502.

Von Wachter, T. (2020). The persistent effects of initial labor market conditions for young adults and their sources. *Journal of Economic Perspectives 34*(4), 168–94.

Vooren, M., C. Haelermans, W. Groot, and H. van den Brink (2019). The effectiveness of active labor market policies: a meta-analysis. *Journal of Economic Surveys 33*(1), 125–149.

Waldron, H. (2001). Links between early retirement and mortality. ORES Working Paper Series No. 93, Social Security Administration, Office of Policy, Washington.

Wang, M. and J. Shi (2014). Psychological research on retirement. *Annu. Rev. Psychol 65*, 209–33.

Webb, M. D. (2014). Reworking wild bootstrap based inference for clustered errors. Queen's Economics Department Working Paper No. 1315, Kingston, Canada.

Westerlund, H., M. Kivimäki, A. Singh-Manoux, M. Melchior, J. E. Ferrie, J. Pentti, M. Jokela, C. Leineweber, M. Goldberg, M. Zins, et al. (2009). Self-rated health before and after retirement in France (GAZEL): a cohort study. *The Lancet 374*(9705), 1889–1896.

Wijanarko, D. W. and I. D. G. K. Wisana (2019). Does the number of siblings affect adult income? An Indonesian case study. *International Journal of Business and Society 20*(S1), 29–41.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*(1), 39–54.

Wu, C., M. C. Odden, G. G. Fisher, and R. S. Stawski (2016). Association of retirement age with mortality: a population-based longitudinal study among older adults in the USA. *Journal of Epidemiology and Community Health 70*(9), 917–923.

Xue, X., M. Cheng, and W. Zhang (2021). Does education really improve health? A meta-analysis. *Journal of Economic Surveys 35*(1), 71–105.

Xue, X., R. R. Reed, and A. Menclova (2020). Social capital and health: A meta-analysis. *Journal of Health Economics 72*, 102317.

Zhang, Y., M. Salm, and A. van Soest (2018). The effect of retirement on healthcare utilization: evidence from China. *Journal of Health Economics 62*, 165–177.

Zhao, M., Y. Konishi, and H. Noguchi (2017). Retiring for better health? Evidence from health investment behaviors in Japan. *Japan and the World Economy 42*, 56–63.

Zhu, R. (2016). Retirement and its consequences for women's health in Australia. *Social Science & Medicine 163*, 117–125.

Zulka, L. E., I. Hansson, and L. B. Hassing (2019). Impact of retirement on cognitive function: a literature review. *Journal of Gerontopsychology and Geriatric Psychiatry 32*(4), 187–203.

Zulkarnain, A. and M. S. Rutledge (2018). How does delayed retirement affect mortality and health? CRR Working Paper 2018-11, Center for Retirement Research at Boston College.

Appendix (Ch. 1)

Table A1.1: Effect on future employment

Author(s)	Country	Data (Time span)	Sample size	Method	Reason of unemployment	Heterogeneity in magnitude/duration
Arulampalam et al. (2000)	UK	BHPS (1991-95)	10,402	DREP	General	Age
Bratberg and Nilsen (2000)	*NOR*	*KIRUT (1989-94)*	*11,239*	*ToE*	*Youth unemployment*	*Stigma effect*
Couch (2001)	GER	GSOEP (1988-96)	8,416	FE-DiD	Job displacement	–
Gregg (2001)	UK	NCDS (1983-85)	8,255	IV	Youth unemployment	Gender
Böheim and Taylor (2002)	UK	BHPS (1991-99)	4,582	ToE	General	–
Knights et al. (2002)	AUS	ALS (1985-88)	9,792	DREP	Lagged employment	–
Hämäläinen (2003)	*FIN*	*Adm. Data (1987-98)*	*5,095*	*DREP*	*Youth unemployment*	*Education level*
Eliason and Storrie (2006)	*SWE*	*EE Data (1987-99)*	*120,093*	*PSM*	*Job displacement*	*Age*
Mroz and Savage (2006)	USA	NLSY (1979-94)	3,731	Other	Youth unemployment	–
Stewart (2007)	UK	BHPS (1991-96)	4,739	DREP	General	Wage level
Doiron and Gørgens (2008)	*AUS*	*AYS (1989-94)*	*1,363*	*ToE*	*Youth unemployment*	–
Gangji and Plasman (2008)	BEL	PSBH (1994-02)	3,352	DREP	General	–
Gaure et al. (2008)	*NOR*	*Adm. Data (1993-01)*	*373,065*	*ToE*	*General*	–
Oberholzer-Gee (2008)	SWI	Exp. Data (1999)	628	Field Exp.	General	–
Verho (2008)	FIN	Adm. Data (1990s)	22,474	PSM	Plant closure	Wage level
Biewen and Steffes (2010)	GER	GSOEP (1991-04)	4,967	DREP	General	–
Dieckhoff (2011)	4 EU	ECHP (1994-01)	9,124	DiD	General	Country
Heylen (2011)	BEL	VDAB (1995-09)	41,784	CF	Youth unemployment	Business cycle at graduation
Manzoni and Mooi-Reci (2011)	GER	GSOEP (1984-05)	9,653	DREP	General	–
Nordström Skans (2011)	SWE	IFAU (1991-94)	17,978	FE-DiD	Youth unemployment	Duration of unemployment spell
Ayllón (2013)	SPA	ECHP (1994-01)	4,160	DREP	General	Cyclical unemployment rate
Cockx and Picchio (2013)	BEL	CBSS (1998-02)	14,660	ToE	Youth unemployment	–
Kroft et al. (2013)	USA	Exp. Data (2008-11)	12,054	Field Exp.	General	Unemployment rate, job features
Ahmad (2014)	DEN	Adm. Data (1994-03)	6,797	DREP	General	–
Eriksson and Rooth (2014)	SWE	Exp. Data (2007)	8,466	Field Exp.	General, youth	Duration of unemployment spell, type of job
Helbling and Sacchi (2014)	SWI	TREE (2003-07)	1,269	PSM	Youth unemployment	–
Nilsen and Reiso (2014)	*NOR*	*Adm. Data (1990-98)*	*29,356*	*PSM*	*Youth unemployment*	–
Ghirelli (2015)	BEL	SONAR (1994-02)	1,902	IV	Youth unemployment	–
Mavromaras et al. (2015)	Australia	HILDA (2001-10)	41,615	DREP	General	–
Plum and Ayllón (2015)	10 EU	ECHP (1994-01)	103,576	DREP	General	Country
Tumino (2015)	UK	BHPS (1991-12)	13,033	DREP	General	Business cycle
Farber et al. (2016)	USA	Exp. Data	6,072	Field Exp.	General	Age, low-skilled
Birkelund et al. (2017)	NOR	Exp. Data (2011-13)	1,188	Field Exp.	General	Gender
Farber et al. (2017)	USA	Exp. Data (2012-14)	12,224	Field Exp.	General	Age
Nunley et al. (2017)	USA	Exp. Data (2013)	9,396	Field Exp.	Youth unemployment	–
Petreski et al. (2017)	MAC	SWTS (2012)	1,044	IV	Youth unemployment	Age, gender, experience
Schmillen and Umkehrer (2017)	GER	IEB, BHP	697,580	IV	Youth unemployment	Number of unemployment spells
Deelen et al. (2018)	NET	Adm. Data (2000-11)	4,100,016	DiD	Job displacement	Age, tenure, education level, local labor market
Dorsett and Lucchino (2018)	UK	BHPS (1991-08)	8,279	ToE	Youth unemployment	Duration of unemployment spell
Nüß (2018)	GER	Exp. Data (2016)	3,124	Field Exp.	General	Firm characteristics
Abebe and Hyggen (2019)	NOR	YiN (2003-07)	2,123	PSM	Youth unemployment	Age, gender, education level
Baert and Verhaest (2019)	BEL	Exp. Data (2013-14)	1,620	Field Exp.	Youth unemployment	–
Farber et al. (2019)	USA	Exp. Data (2017)	8,488	Field Exp.	General	Age
Kuchibhotla et al. (2020)	Sri Lanka	Survey Data (2006)	609	PSM	Youth unemployment	–
Ayllón et al. (2021)	12 EU	EU-SILC (2004-15)	257,823	DREP	Youth unemployment	Country, gender, business cycle
Shi and Wang (2021)	SWI, GRE	Exp. Data (2016)	9,837	Field Exp.	General	Country
Tanzi (2022)	ITA	CO (2009-10)	67,949	IV	Youth unemployment	Regional labor market conditions

Notes: CF = Control Function; PSM = Propensity Score Matching; FE-DiD = Panel Fixed-Effects/Difference-in-Differences; IV = Instrumental Variables; ToE = Timing of Events; DREP = Dynamic Random-Effects Probit models; Field Exp. = Field Experiment; Other = Other methods, such as Discrete Factor Maximum Likelihood, Exclusion restrictions, etc.
For "General" reasons of unemployment we mean experiences of unemployment during recession and unemployment episodes for which the reason is not clearly specified.
Articles in italics are those ones not included in the meta-regression analysis.

Table A1.2: Effect on future labor earnings

Author(s)	Country	Data (Time span)	Sample size	Method	Reason of unemployment	Heterogeneity in magnitude/duration
Bratberg and Nilsen (2000)	*NOR*	*KIRUT (1989-94)*	*11,239*	*ToE*	*Youth unemployment*	*Stigma effect*
Arulampalam (2001)	UK	BHPS (1991-97)	2,092	FE-DiD	General	Duration of previous unemployment
Burda and Mertens (2001)	GER	GSOEP, IAB (1985-94)	2,185	Other	Job displacement	Wage level
Couch (2001)	GER	GSOEP (1988-96)	8,416	FE-DiD	Job displacement	–
Gregory and Jukes (2001)	UK	NESPD, JUVOS (1984-94)	66,000	FE-DiD	General	Age, wage level
Lupi et al. (2002)	ITA	SHIW (1993-95)	1,112	FE	General	Regional unemployment rate
Nickell et al. (2002)	UK	NES, JUVOS (1982-97)		FE-DiD	General	Age, high-skilled, duration of unemployment
Arranz and García-Serrano (2003)	SPA	HSIPRE (1987-97)	65,340	FE-DiD	Job displacement	Duration of unemployment, age, tenure, occupation
Kletzer and Fairlie (2003)	USA	NLSY (1984-93)	12,686	FE-DiD	Job displacement	Age
Arranz et al. (2005)	6 EU	ECHP (1995-01)	9,205	FE-DiD	General, job displacement	Duration of unemployment, country, age
Gregg and Tominey (2005)	UK	NCDS	4,449	IV	Youth unemployment	Number of unemployment spells
Spivey (2005)	USA	NLSY (1979-00)	6,111	FE-DiD	General	Gender, duration of unemployment
Eliason and Storrie (2006)	*SWE*	*EE. Data (1987-99)*	*120,093*	*PSM*	*Job displacement*	*Age*
Gangl (2006)	USA, 12 EU	SIPP, ECHP (1994-01)	6,260	DiD	General	Country, wage level, age, gender
Mroz and Savage (2006)	USA	NLSY (1979-94)	3,731	Other	Youth unemployment	–
Gangji and Plasman (2007)	BEL	PSBH (1994-02)	2,521	FE-DiD	General	Duration of unemployment
Gaure et al. (2008)	*NOR*	*Adm. Data (1993-01)*	*373,065*	*ToE*	*General*	–
Verho (2008)	FIN	Adm. Data (1990s)	22,474	PSM	Plant closure	Wage level
Gartell (2009)	SWE	IFAU (1991-99)	36,422	CF	Youth unemployment	Duration of unemployment, local unemployment rate
Nordström Skans (2011)	SWE	IFAU (1991-94)	17,978	FE-DiD	Youth unemployment	Duration of unemployment spell
Cockx and Picchio (2013)	BEL	CBSS (1998-02)	14,660	ToE	Youth unemployment	–
Helbling and Sacchi (2014)	SWI	TREE (2003-07)	1,269	PSM	Youth unemployment	–
Ghirelli (2015)	BEL	SONAR (1994-02)	1,902	IV	Youth unemployment	–
Möller and Umkehrer (2015)	GER	Adm. Data (1978-02)	728,841	IV	Youth unemployment	Wage level
Mooi-Reci and Ganzeboom (2015)	NET	OSA (1985-00)	4,815	FE-DiD	General, plant closure	–
Ordine and Rose (2015)	*ITA*	*IITVW (2005-06)*	*1,537*	*PSM*	*General, overeducation*	–
Guvenen et al. (2017)	*USA*	*MEF (1978-10)*		*DiD*	*General*	*Income level*
Petreski et al. (2017)	MAC	SWTS (2012)	1,044	IV	Youth unemployment	Age, gender, experience
Deelen et al. (2018)	NET	Adm. Data (2000-11)	4,100,016	DiD	Job displacement	Age, tenure, education level, local labor market
Abebe and Hyggen (2019)	NOR	YiN (2003-07)	2,123	PSM	Youth unemployment	Age, gender, education level
De Fraja et al. (2021)	UK	LLMDB (bc 1960-67)	26,273	FE; IV	General, youth unemployment	Low-skilled

Notes: See Table A3.1.

Appendix (Ch. 2)

A. Descriptive statistics

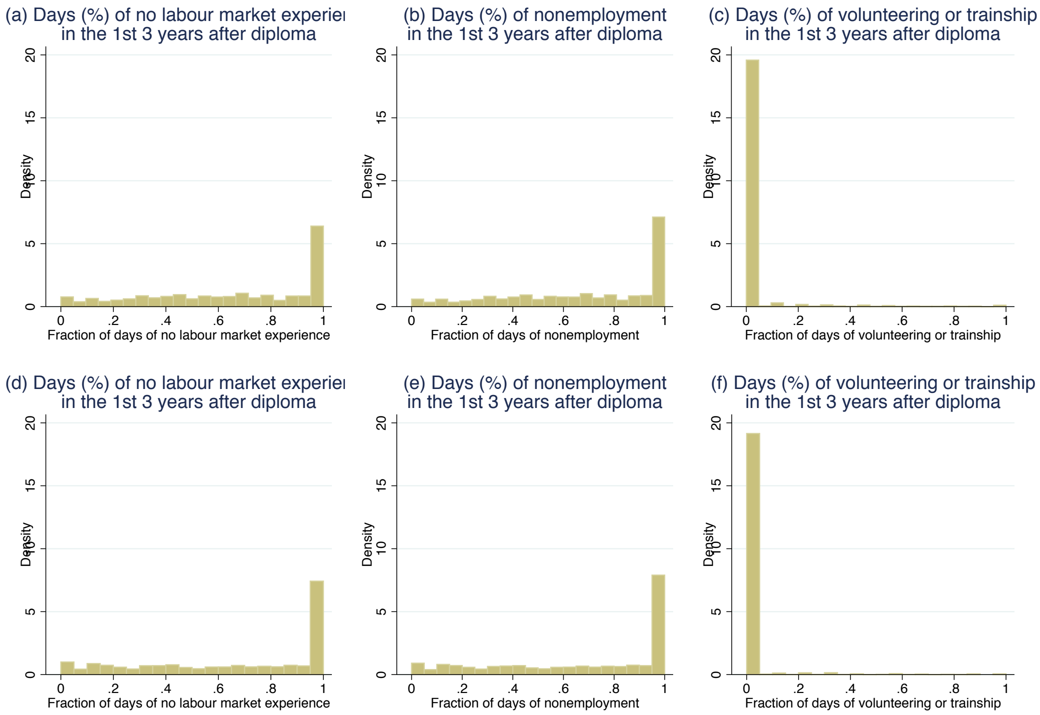
In this section we provide further descriptive statistics. Figure A2.1 shows the fraction of days of nonemployment in the first 3 years after school completion, and then distinguishes between fraction of days in nonemployment only considering paid jobs, and the fraction of days spent in volunteering, stage or trainship. Indeed, it is built as the percentage of days of no labor market experience during the first 3 years after diploma, and then divided between days of nonemployment and days of volunteering, stage or trainship experiences since school completion. The main treatment variable is the fraction of days spent in nonemployment or trainship during the first 3 years, our sample counts 3,467 individuals with no days in employment in this time window, of which 3,178 had neither a job nor a stage or an internship. Figure A2.2 and A2.3 show the distribution of individuals across the age at school completion for the samples observed at different moments after diploma and Table A2.1 shows the distribution of the age at school completion by birth cohorts. There is a very homogeneous distribution in the average age at school completion, and a decrease in standard deviation, between those born in the 1960s and those born in the next decades. As we can see, each percentile of the distribution has remained stable across birth decades: the 25th percentile at 18 years, the median at 19 years and the 90th percentile at 20 years of age.

Table A2.1: Distribution of the age at school completion by birth cohorts

	Age at school completion					
	Born in 1960s		Born in 1970s		Born in 1980s	
	Males	Females	Males	Females	Males	Females
Mean	18.782	18.551	18.809	18.690	18.742	18.773
Std. Dev.	1.222	1.110	1.113	1.024	1.141	0.921
10th percentile	17	17	17	17	17	18
25th percentile	18	18	18	18	18	18
50th percentile	19	19	19	19	19	19
75th percentile	20	19	19	19	19	19
90th percentile	20	20	20	20	20	20
Observations	2,724	2,570	2,176	1,968	496	361

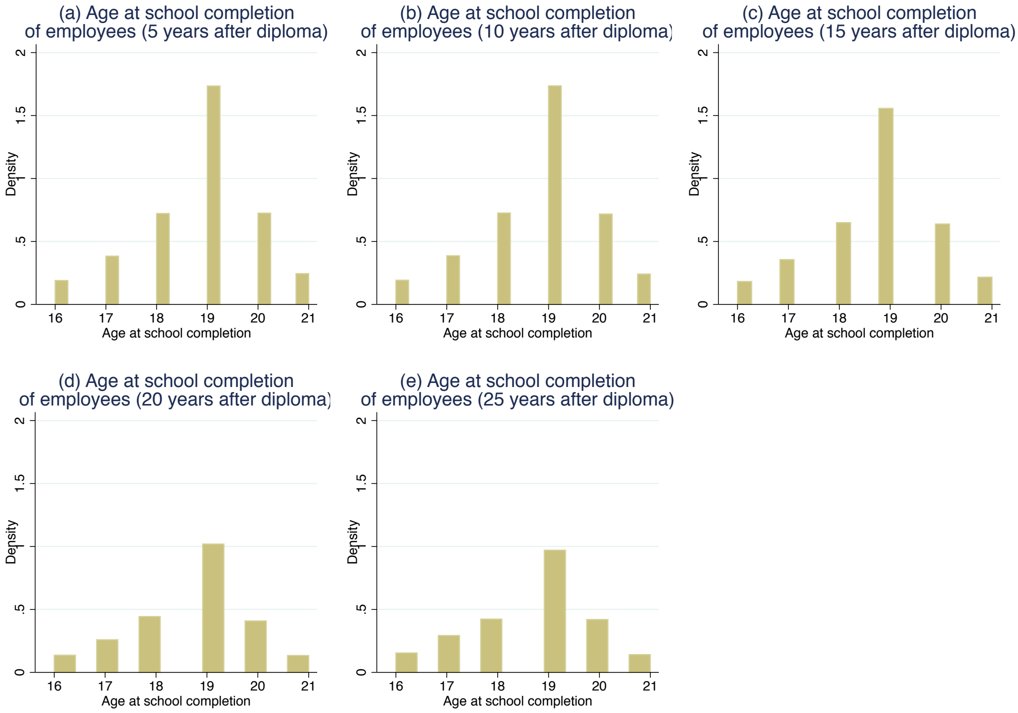
Table A2.2 displays more detailed descriptive statistics on outcome variables than those reported and discussed in section 2.3. The first type of outcome is observed through yearly labor earnings, daily wages and annual total income (i.e. including any subsidies in addiction to wages). Subsequent participation in the labor market is measured as fraction

Figure A2.1: Treatment variables



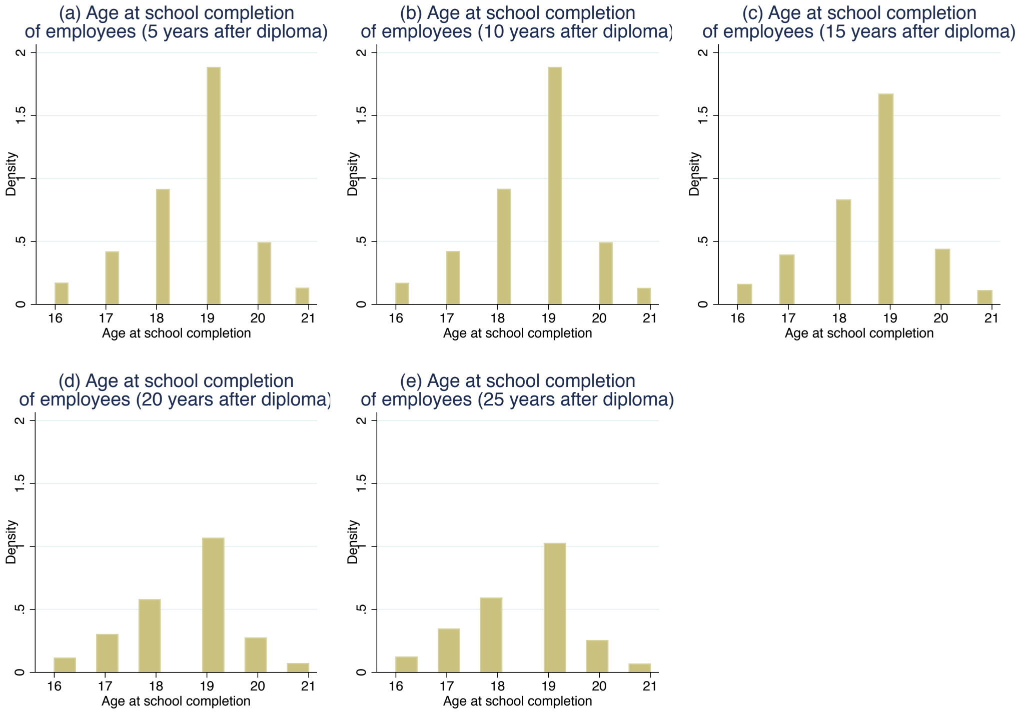
Notes: The histograms display the distribution of employees across the treatment variables, evaluated in a 3-years window after diploma. Graph (a) is drawn using fraction of days of no labor market experience in the first 3 years after school completion for males. Graph (b) and (c) distinguish between the fraction of days of nonemployment and the time spent in volunteering, stage or trainship in the first 3 years. Graph (d) counts the fraction of days of no labor market experience for females in the same time window, whereas graph (e) and (f) distinguish between fraction of days of nonemployment and of days in stage, trainship or volunteering.

Figure A2.2: The age at school completion (Males)



Notes: The histograms display the distribution of individuals across the age at diploma for the samples observed at different moments after school completion. Graph (a) is drawn using the 5,396 individuals of which we can observe the labour market outcomes 5 years after diploma. Graphs (b), (c) and (d) are drawn using 5,310, 4,864, 3,947 and 2,792 observations for whom we can observe the labor market outcomes 10, 15, 20 and 25 years since school exit, respectively.

Figure A2.3: The age at school completion (Females)



Notes: The histograms display the distribution of individuals across the age at diploma for the samples observed at different moments after school completion. Graph (a) is drawn using the 4,899 females of which we can observe the labor market outcomes 5 years after diploma. Graphs (b), (c) and (d) are drawn using 4,722, 4,235, 3,383 and 2,423 observations for whom we can observe the labour market outcomes 10, 15, 20 and 25 years since school exit, respectively.

of total days in employment, and here we also provide separated descriptive statistics among part-time and full-time employment. A notewhorty point is that, while males spent only 2% of days per-year in part-time employment, women pass by 4% to 19% in part-time work along the time span covered by our analysis. At the same time, males spent a fraction of days in full-time employment that is about 35% higher than for females at the 25th year. The standard deviation is strictly decreasing for these kinds of outcomes, with the exception of part-time employment.

Table A2.2: Outcome variables at different years after school completion

Males

Year after school		Yearly labor earnings (€)[(a)]		Daily wages (€)		Total annual income (€)[(b)]	
completion	Observations	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
5	5,396	12,352.83	10,627.46	37.43	29.08	12,485.36	10,742.46
10	5,310	18,734.76	12,607.04	53.20	32.81	19,016.21	12,777.53
15	4,864	22,759.81	14,176.09	63.69	37.06	23,281.36	14,318.02
20	3,947	25,909.90	16,449.95	71.25	41.33	26,657.92	16,539.70
25	2,792	28,344.23	18,118.38	77.25	44.38	29,254.81	18,253.51

Females

Year after school		Yearly labor earnings (€)		Daily wages (€)		Total annual income (€)	
completion	Observations	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
5	4,899	10,190.33	9,658.65	32.17	27.55	10,288.02	9,720.26
10	4,722	13,077.41	11,113.04	40.16	30.66	13,468.62	11,251.32
15	4,235	14,770.19	11,989.04	46.80	32.92	15,483.36	12,222.27
20	3,383	17,242.18	13,109.76	55.09	34.92	17,954.16	13,347.73
25	2,423	19,601.58	13,708.33	62.53	36.31	20,156.64	13,842.76

Males

Year after school		Days in employment[(c)]		Days in part-time		Days in full-time	
completion	Observations	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
5	5,396	0.62	0.45	0.01	0.11	0.60	0.46
10	5,310	0.79	0.38	0.02	0.12	0.77	0.40
15	4,864	0.85	0.34	0.02	0.13	0.83	0.35
20	3,947	0.87	0.31	0.02	0.12	0.86	0.32
25	2,792	0.89	0.28	0.02	0.12	0.88	0.30

Females

Year after school		Days in employment		Days in part-time		Days in full-time	
completion	Observations	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
5	4,899	0.57	0.46	0.04	0.19	0.53	0.47
10	4,722	0.67	0.44	0.08	0.25	0.60	0.47
15	4,235	0.73	0.41	0.14	0.33	0.60	0.47
20	3,383	0.79	0.37	0.17	0.36	0.63	0.46
25	2,423	0.83	0.33	0.19	0.38	0.65	0.45

(a) Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
(b) Total annual income includes any subsidies in addiction to wages.
(c) These outcome variables measure the fraction of days spent in employment.

Further marginal correlation in order to understand relationship between unemployment after graduation and the labor market outcomes are provided running a series of separate OLS regressions for each $t \in \{5, 10, 15, 20, 25\}$

$$Y_{it} = x_{it}\pi_t + \beta_t TR_i + \varepsilon_{it} \tag{A2.1}$$

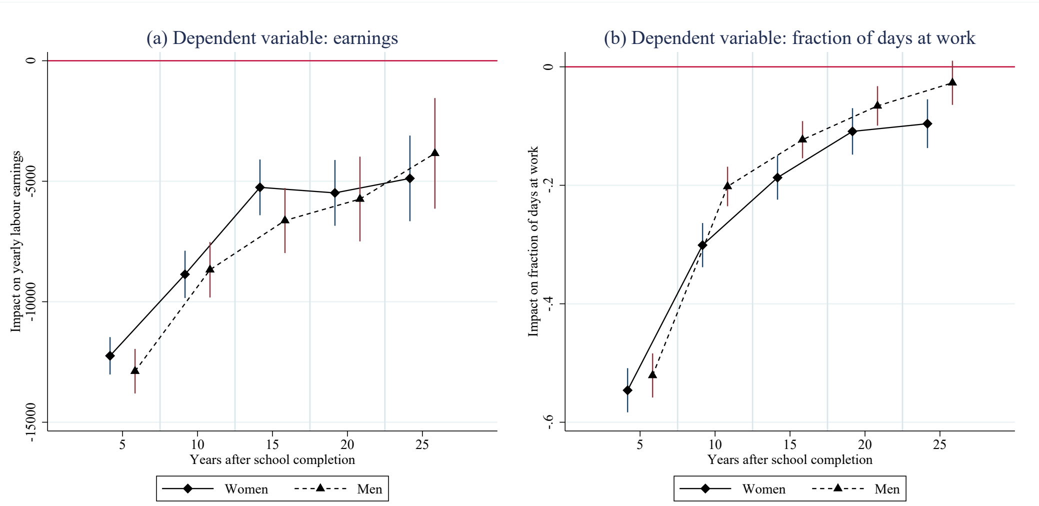
where:

- Y_{it} is either wage or fraction of time spent in employment *t* years after school completion;
- x_{it} is a vector of covariates: the constant term, age at school completion, regional dummies, calendar year dummies, regional unemployment, employment and GDP growth rates in the *t*-th year after school exit, number of kids in t, quarter of birth and further personal information;
- TR_i is the treatment variable, that is the fraction of days of nonemployment during the 3 years after school completion;
- ε_{it} is the error term.

The estimated $\beta_t TR_{it}$ are graphically displayed in Figure A2.4 along with 95% confidence intervals. In particular, Figure A2.4 (a) displays the evolution over time of the wage penalties: the continuous line shows the wage trend for men, the dotted line is the earning penalty for women. Graph (b) in Figure A2.4 focuses instead on the penalty in terms of fraction of time spent in employment. In this case, the fraction of days of nonemployment during the 3 years after school completion seems to generate a higher negative effect in terms of employability for women rather than for men. Tables A2.3 and A2.4 display OLS estimated coefficients of the impact of the three different treatment variables on both labor market outcomes. From 5 to 20 years after school completion, men seem to suffer from a greater wage loss with respect to women, with the exception of 25 years later, when the earning penalty in males sample ($\widehat{\beta} = -3,843.00$) is lower than that in the females one ($\widehat{\beta} = -4,881.49$). Further treatment effects are concerning the fraction of days of nonemployment nor unpaid experiences, and the percentage of days spent in volunteering, trainship or stage during the same period. As concerning the former, our results highlight the same negative effects of our main treatment but with larger magnitude. About the latter, unpaid experiences such as volunteering, stage or trainship during the first 3 years after school completion negatively affects yearly labor earnings 5 years later for both males and females, and participation in labor market until 25 years later for males. However, these kind of experiences seem to be able in improving wages for women in the long-run.

However, these estimation results cannot be given a causal interpretation because of endogeneity of labor market outcomes: time-constant and time-varying unobserved traits jointly determine both the experiences after school completion and the future labor market performances. Our econometric model is aimed at disentangling the true causal effect of nonemployment experiences from the spurious one induced by systematic differences across individuals with different labor market histories, due to both time-varying and time-constant characteristics unobserved by the analyst. Table A2.5 illustrates the exclusion restrictions across our 13 equations.

Figure A2.4: OLS estimated coefficients of the impact of nonemployment during the first 3 years after school completion on yearly labor earnings and fraction of days spent at work



Notes: The vertical segments crossing the dots are 95% confidence intervals.

Table A2.3: OLS estimated coefficients of the impact of three different treatment variables on yearly labor earnings

	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Males					
Nonemployment during the first 3 years after school completion	-12881.81*** (471.41)	-8672.32*** (586.48)	-6630.59*** (688.59)	-5737.47*** (896.69)	-3843.00*** (1170.94)
Nonemployment nor internship during the first 3 years after school completion	-15179.09*** (461.87)	-9666.57*** (598.37)	-7558.83*** (705.19)	-6599.85*** (910.91)	-4868.86*** (1203.01)
Volunteering, stage or trainship during the first 3 years after school completion	-6803.05*** (1016.75)	-1841.83 (1517.44)	-1985.41 (1544.42)	-1726.98 (2012.73)	-2949.59 (2556.87)
b) Females					
Nonemployment during the first 3 years after school completion	-12242.88*** (395.24)	-8862.98*** (500.42)	-5253.49*** (589.27)	-5482.48*** (696.32)	-4881.49*** (906.45)
Nonemployment nor internship during the first 3 years after school completion	-13251.48*** (389.69)	-9048.44*** (503.84)	-4838.30*** (585.02)	-4860.21*** (693.58)	-4051.61*** (893.44)
Volunteering, stage or trainship during the first 3 years after school completion	-6614.63*** (1415.15)	-301.28 (1723.83)	4032.83** (1953.50)	5215.04** (2103.49)	6369.88** (2903.19)
Observations (males)	5396	5310	4864	3947	2792
Observations (females)	4899	4722	4235	3383	2423

Notes: The equations for the labor market outcomes also include age at school completion, regional dummies, calendar year dummies, regional unemployment, regional employment, regional GDP growth, the number of kids, the number of siblings when the individual was 14 years old, predetermined information, quarter of birth, year of birth, and parents' characteristics when the respondent was 14. Their OLS estimated parameters are not reported for the sake of brevity, as well as the case of daily wages and total earnings as dependent variables.
Yearly wages are in 2014 prices and deflated by the ISTAT consumer price index.
*** Significant at 1%, ** significant at 5%, * significant at 10%. Standard errors robust to heteroskedasticity are reported in parentheses.

Table A2.4: OLS estimated coefficients of the impact of three different treatment variables on yearly fraction of days spent at work

	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Males					
Nonemployment during the first 3 years after school completion	-0.521*** (0.019)	-0.202*** (0.017)	-0.123*** (0.016)	-0.066*** (0.017)	-0.027 -0.019
Nonemployment nor internship during the first 3 years after school completion	-0.637*** (0.018)	-0.252*** (0.017)	-0.156*** (0.016)	-0.129*** (0.017)	-0.103*** (0.018)
Volunteering, stage or trainship during the first 3 years after school completion	-0.373*** (0.048)	-0.174*** (0.050)	-0.108** (0.043)	-0.237*** (0.050)	-0.294*** (0.058)
b) Females					
Nonemployment during the first 3 years after school completion	-0.546*** (0.019)	-0.301*** (0.019)	-0.187*** (0.019)	-0.109*** (0.020)	-0.096*** (0.021)
Nonemployment nor internship during the first 3 years after school completion	-0.601*** (0.018)	-0.312*** (0.019)	-0.176*** (0.019)	-0.102*** (0.019)	-0.106*** (0.021)
Volunteering, stage or trainship during the first 3 years after school completion	-0.382*** (0.063)	-0.054 (0.067)	0.116** (0.056)	0.064 (0.052)	-0.062 (0.063)
Observations (males)	5396	5310	4864	3947	2792
Observations (females)	4899	4722	4235	3383	2423

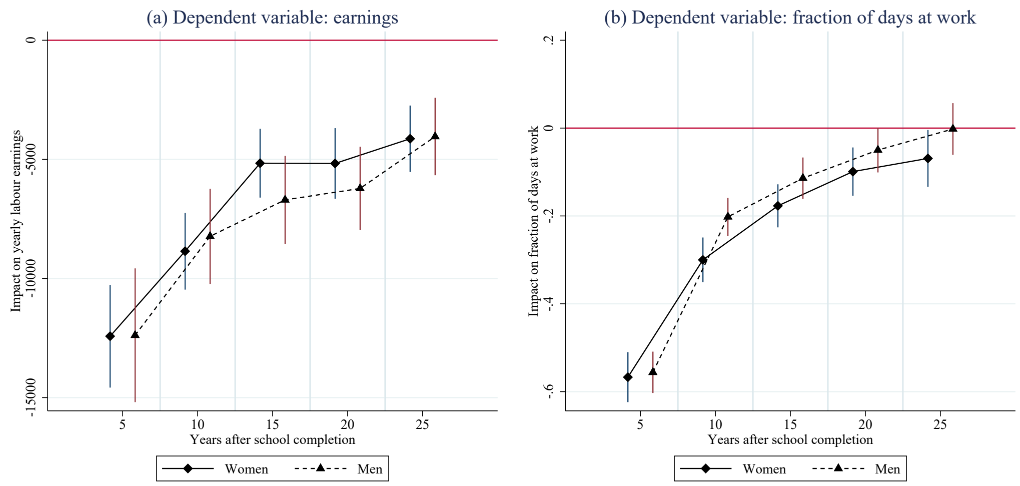
Notes: See Table A2.3.
*** Significant at 1%, ** significant at 5%, * significant at 10%. Standard errors robust to heteroskedasticity are reported in parentheses.

Table A2.5: Observed covariates across equations

Regressors included	Measurement equations: Employment 1 year before school completion	Measurement equations: Number of siblings at 14	Treatment equation: Days (%) of nonemployment during the first 3 years after school completion	Outcomes: Labor market outcomes t years after school completion
Age at school completion	–	–	Yes	Yes
Fraction of time spent at work 1 year before school completion	–	–	Yes	Yes
Mother's age	Yes	Yes	Yes	Yes
Number of siblings at 14	Yes	–	Yes	Yes
Mother's highest education	Yes	Yes	Yes	Yes
Father's highest education	Yes	Yes	Yes	Yes
Mother's employment at 14	Yes	Yes	Yes	Yes
Father's employment at 14	Yes	Yes	Yes	Yes
Respondent lives with both parents at 14	Yes	Yes	Yes	Yes
Quarter of birth	Yes	Yes	Yes	Yes
Year of birth	Yes	Yes	Yes	Yes
Geographical area at birth (5 areas)	Yes	Yes	Yes	–
Geographical area at t (5 areas)	–	–	–	Yes
Regional unemployment rate at birth	Yes	Yes	–	–
Regional employment rate at birth	Yes	Yes	–	–
Regional GDP growth rate at birth	Yes	Yes	–	–
Average regional unemployment rate 3 years after diploma	–	–	Yes	–
Average regional employment rate 3 years after diploma	–	–	Yes	–
Average regional GDP growth rate 3 years after diploma	–	–	Yes	–
Regional unemployment rate at t	–	–	–	Yes
Regional employment rate at t	–	–	–	Yes
Regional GDP growth rate at t	–	–	–	Yes
IT-SILC wave (2005 or 2011)	Yes	Yes	Yes	Yes
Calendar year of observation	Yes	Yes	Yes	Yes
Number of kids	–	–	–	Yes
Average number of kids 3 years after diploma	–	–	Yes	–
Days (%) of nonemployment during the first 3 years after school completion	–	–	–	Yes

B. Full set of estimation results without UH

Figure B2.1: Impact of nonemployment during the first 3 years after school completion on yearly labor earnings and fraction of days spent at work without unobserved heterogeneity



Notes: The vertical segments crossing the dots are 95% confidence intervals.

Table B2.1: Estimated coefficients of the covariates of the labor market outcome equations without unobserved heterogeneity

	Yearly labor earnings (Males)			Yearly fraction of days spent at work (Males)			Yearly labor earnings (Females)			Yearly fraction of days spent at work (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	13.986		15.534	0.000		0.000	5.373		13.857	0.000		0.001
Mother's age at respondent's birth is missing	151.226		494.199	-0.012		0.014	605.540		443.407	-0.009		0.017
Respondent's father has at least secondary education	951.422	***	183.232	-0.022	***	0.005	779.680	***	164.870	-0.006		0.007
Respondent's mother has at least secondary education	26.658		196.572	-0.025	***	0.006	412.113	**	176.057	0.000		0.008
Mother's employment at 14	-495.826	***	165.475	-0.007		0.005	409.957	***	152.685	0.002		0.006
Father's employment at 14	1000.903	***	263.783	0.027	***	0.007	9.044		279.571	0.001		0.011
Respondent lived with both parents at 14	125.569		369.431	-0.006		0.011	422.309		368.693	0.026	*	0.014
Number of siblings at 14 if IT-SILC wave is 2005	-356.875	***	91.883	-0.004	*	0.002	-228.369	***	81.537	-0.006		0.003
Number of siblings at 14 if IT-SILC wave is 2011	-689.988	***	109.041	-0.01	***	0.003	-321.668	***	102.296	-0.016	***	0.004
Quarter of birth - Reference category: October, November, December												
January, February, March	-9.733		205.828	-0.002		0.006	-429.952	**	197.745	-0.005		0.008
April, May, June	-400.657	*	209.006	-0.005		0.006	-332.514		200.968	-0.016	*	0.008
July, August, September	-494.726	**	204.568	-0.002		0.006	141.873		198.403	0.006		0.008
Year of birth/10 (normalized to its minimum)	157.570		482.506	0.020		0.014	-609.114		454.063	0.031		0.018
Geographical area at t - Reference category: North-West												
North-East	-686.557	***	194.280	-0.013	*	0.007	-780.458	***	181.858	-0.013		0.008
Center	-2466.930	***	205.858	-0.015	**	0.007	-1451.317	***	191.579	-0.019	**	0.008
South	-2877.356	***	424.089	-0.044	***	0.012	-1340.293	***	421.772	-0.031	*	0.015
Islands	-1750.283	***	558.850	0.004		0.015	-170.089		589.975	-0.008		0.020
Regional unemployment rate at t	-147.373	***	53.494	-0.009	***	0.001	-313.564	***	626.012	-0.017	***	0.002
Regional employment rate at t	231.668	***	31.988	0.003	***	0.001	86.798	**	37.289	0.000		0.001
Regional growth rate at t	1193.699		3576.819	0.104		0.099	-401.505		3438.835	0.010		0.128
IT-SILC wave 2011	490.063	**	233.883	-0.009		0.007	253.047		214.139	0.009		0.009
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	1284.229	***	307.603	0.026	***	0.009	1225.191	***	299.381	0.030	**	0.012
Between 1986 and 1991	534.515		516.696	0.025	*	0.015	849.786	*	488.748	0.039	**	0.019
Between 1991 and 1996	-119.643		742.364	0.039	*	0.022	814.501		716.969	0.074	***	0.028
After 1996	-1361.657		1037.91	0.020		0.030	682.472		1012.263	0.062	*	0.039
Age at school completion	877.550	***	82.358	-0.001		0.002	766.590	***	834.795	0.010	***	0.003
Number of kids at t	1175.862	***	97.964	0.014	***	0.003	-2103.802	***	88.876	-0.048	***	0.004
Fraction of time spent at work 1 year before diploma	-2229.414	***	321.634	-0.016		0.012	-5460.331	***	534.594	-0.097	***	0.018
Constant at $t = 5$	-5800.698	*	3115.307	0.904	***	0.075	3117.294		3120.362	0.866	***	0.100
Constant at $t = 10$	-2846.612		3065.073	0.832	***	0.076	4030.952		3139.494	0.796	***	0.101
Constant at $t = 15$	-723.747		3105.204	0.823	***	0.077	3690.866		3148.441	0.784	***	0.102
Constant at $t = 20$	1320.226		3065.818	0.797	***	0.077	6377.144	**	3217.662	0.806	***	0.104
Constant at $t = 25$	1625.225		3108.922	0.780	***	0.079	8085.320	**	3207.738	0.842	***	0.105
$\ln(\sigma^2)$	0.533	***	0.006	-2.151	***	0.015	0.128	***	0.006	-1.950	***	0.019

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table B2.2: Estimated coefficients of the measurement equations without unobserved heterogeneity

	Number of siblings at 14 (Males)			Employment 1 year before school completion (Males)			Number of siblings at 14 (Females)			Employment 1 year before school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	-0.013	***	0.003	0.000		0.001	-0.016	***	0.003	-0.001		0.001
Mother's age at respondent's birth is missing	0.273	**	0.107	0.041	**	0.019	-0.381	***	0.105	-0.007		0.018
Respondent's father has at least secondary education	-0.024		0.040	-0.028	***	0.008	-0.122	***	0.041	-0.010		0.007
Respondent's mother has at least secondary education	-0.104	**	0.045	-0.026	***	0.012	-0.088	*	0.046	-0.005		0.008
Mother's employment at 14	-0.136	***	0.040	-0.003		0.007	-0.089	**	0.038	-0.011	*	0.007
Father's employment at 14	0.301	***	0.049	-0.041	***	0.009	0.221	***	0.056	0.005		0.011
Respondent lived with both parents at 14	-0.073		0.066	0.038	***	0.014	0.169	**	0.085	-0.009		0.014
Number of siblings at 14 if IT-SILC wave is 2005	–		–	0.004		0.003	–		–	0.005	*	0.003
Number of siblings at 14 if IT-SILC wave is 2011	–		–	0.012	**	0.004	–		–	0.016	***	0.003
Quarter of birth - Reference category: October, November, December												
January, February, March	-0.028		0.045	-0.008		0.009	0.118	**	0.047	0.014	*	0.008
April, May, June	-0.073		0.047	-0.012		0.009	0.168	***	0.048	0.017	**	0.008
July, August, September	-0.020		0.044	-0.005		0.009	0.078	*	0.048	0.013		0.008
Year of birth/10 (normalized to its minimum)	0.116		0.083	-0.028	*	0.016	-0.105		0.086	-0.025	*	0.014
Geographical area at birth - Reference category: North-West												
North-East	0.127	**	0.052	0.031	***	0.008	0.162	***	0.049	0.022	***	0.007
Center	-0.007		0.059	-0.013		0.011	-0.047		0.061	0.000		0.009
South	0.602	***	0.072	-0.011		0.015	0.712	***	0.076	-0.004		0.015
Islands	0.511	***	0.093	0.013		0.019	0.663	***	0.098	0.003		0.021
Regional unemployment rate at birth	-0.035	**	0.016	-0.010	***	0.003	-0.057	***	0.016	-0.009	***	0.002
Regional employment rate at birth	-0.027	***	0.007	-0.002		0.001	-0.028	***	0.006	-0.001		0.001
Regional growth rate at birth	-1.804	**	0.781	-0.281	*	0.163	-1.047		0.830	-0.195		0.169
IT-SILC wave 2011	-0.498	***	0.036	-0.014		0.010	-0.535	***	0.037	-0.012		0.008
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	-0.173	***	0.061	0.042	***	0.012	-0.073		0.061	-0.011		0.011
Between 1986 and 1991	-0.404	***	0.097	0.051	**	0.020	-0.084		0.094	0.022		0.016
Between 1991 and 1996	-0.527	***	0.132	0.050	*	0.027	-0.193		0.138	0.029		0.023
After 1996	-0.571	***	0.181	0.118	***	0.035	-0.069		0.181	0.066	**	0.030
Constant	3.426	***	0.460	0.239	***	0.089	3.554	***	0.448	0.184	**	0.071
ln(σ^2)	0.179	***	0.010	-3.064	***	0.022	0.140	***	0.014	-3.516	***	0.020

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table B2.3: Estimated coefficients of the treatment equation without unobserved heterogeneity

	Days (%) of nonemployment during the first 3 years after school completion (Males)			Days (%) of nonemployment during the first 3 years after school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	0.001		0.001	0.002	**	0.001
Mother's age at respondent's birth is missing	0.023		0.025	0.054	*	0.029
Respondent's father has at least secondary education	0.045	***	0.010	0.027	**	0.011
Respondent's mother has at least secondary education	0.038	***	0.010	0.056	***	0.012
Mother's employment at 14	0.021	**	0.009	0.008		0.010
Father's employment at 14	-0.014		0.015	-0.014		0.017
Respondent lived with both parents at 14	-0.012		0.021	-0.006		0.023
Number of siblings at 14 if IT-SILC wave is 2005	0.001		0.005	-0.003		0.005
Number of siblings at 14 if IT-SILC wave is 2011	-0.005		0.006	-0.014	**	0.006
Quarter of birth - Reference category: October, November, December						
January, February, March	-0.020	*	0.011	-0.012		0.013
April, May, June	-0.018		0.011	-0.005		0.013
July, August, September	-0.001		0.011	0.004		0.013
Year of birth/10 (normalized to its minimum)	0.007		0.025	0.049	*	0.028
Geographical area at birth - Reference category: North-West						
North-East	-0.031	***	0.011	-0.061	***	0.012
Center	0.040	***	0.012	0.076	***	0.013
South	0.024		0.017	0.071	***	0.019
Islands	-0.020		0.023	0.039		0.029
Average regional unemployment rate 3 years after diploma	0.009	***	0.002	0.014	***	0.003
Average regional employment rate 3 years after diploma	-0.008	***	0.001	-0.007	***	0.002
Average regional growth rate 3 years after diploma	0.142		0.310	0.644	*	0.349
IT-SILC wave 2011	0.023	*	0.013	0.030	**	0.014
Calendar year of t - Reference category: before 1981						
Between 1981 and 1986	-0.011		0.020	-0.050	**	0.021
Between 1986 and 1991	0.033		0.029	-0.044		0.031
Between 1991 and 1996	0.009		0.040	-0.064		0.044
After 1996	-0.027		0.054	-0.096	*	0.061
Average number of kids 3 years after diploma	-0.160	**	0.061	0.028		0.027
Age at school completion	-0.006		0.004	-0.022	***	0.050
Fraction of time spent at work 1 year before diploma	-0.338	***	0.021	-0.390	***	0.030
Constant	1.085	***	0.123	1.195	***	0.141
$\ln(\sigma^2)$	-2.521	***	0.030	-2.435	***	0.032

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

C. Full set of estimation results with UH

In this section we briefly discuss the estimated coefficients associated with the other covariates entering the equations for the yearly labor earnings and the fraction of days spent in employment, but also the selection into treatment and the three equations for the selection-free measurements. A first round of estimates is performed on the model with time-constant unobserved heterogeneity with discrete distribution and 5 support points, while a second set of findings is related to the models with time-varying unobserved heterogeneity. In this case, we stopped at $H = 10$ support points. Table C2.1 shows the estimated coefficients of the explanatory variables entering the outcome equations. Tables C2.2 and C2.3 illustrate the estimated parameters of our two selection-free measurements and of the treatment equation for both males and females. Table C2.4 contains the estimated discrete distribution of the time-varying latent factor with 5 support points for males and females samples. Table C2.5 shows the loading factors connecting this distribution and the error terms of each of the 13 equations in our framework with time-constant unobserved heterogeneity with 5 support points.

Similarly, Tables from C2.6 to C2.10 correspond to the counterpart of the aforementioned estimates for the model with time-varying unobserved heterogeneity. Table C2.6 shows that parents' educational attainment and if they were in employment when the respondent was 14 years old are positively associated with labor earnings. For both males and females, we find that regions matter in explaining earnings variation: individuals in North-West seem to earn more, in particular with respect those ones in the Center, as well as those working in regions with lower unemployment rates. Labor earnings are increasing in the age at which the diploma was obtained, while the number of kids increases earnings for men but reduces those for women. Having worked before diploma determines lower labour earnings but increase participation in the labor market. As concern the fraction of days in employment, individuals living in regions with lower unemployment rates have a larger participation. However, results also suggest that individuals in Southern Italy or living in the Islands spend more time in the labor market.

Table C2.7 reports the estimated parameters for selection-free measurements. We find that the probability of having worked in the year before high school diploma is larger if the number of siblings is higher and for individuals born in North-East Italy. The number of siblings is smaller if respondent's mother was employed and attained higher education levels, and is higher in the Center and in the Southern regions.

Finally, in Table C2.8 we report the estimated coefficients of the selection into treatment equation. Our findings suggest that parents' education is statistically significant in explaining the fraction of days spent in nonemployment after school completion. More interesting, individuals born in North-East regions are less likely to spent time in nonemployment after high school diploma, while school leavers who were born in the Center or in the South are significantly more likely to experience early unemployment. Average regional unemployment rate 3 years after school completion is a further strong predictor of the selection into treatment, whereas the more the individual workerd 1 year before diploma the lower is the probability of experiencing nonemployment after diploma. Finally, the average number of kids 3 years after graduation seems to be negatively associated to nonemployment only for men.

Table C2.9 contains the estimated discrete distribution of the time-varying unobserved heterogeneity with 10 support points once constrained $\theta^h_{20} = \theta^h_{25}$ for each $h = 1, ..., H$. Indeed, we obtain 36 support points (rather than 40) for both males and females. The last columns report the resulting probabilities p^h for each support point and 9 weights for the probability masses. Table C2.10 shows the loading factors connecting the distribution of the latent factor θ and the error terms of the 13 equations included in our framework. In particular, we estimate 2 loading factors for the measurement equations, 1 for the selection into treatment, and 4 for the equations of the participation in the labor market (they would be 5 without the constraint of time-varying unobserved heterogeneity for the last 2 periods). The loading factors entering the yearly labor earnings equations are normalized to 1, so the support points of θ are in 2014 Euro.

Table C2.1: Estimated coefficients of the covariates of the labor market outcome equations with time-constant unobserved heterogeneity

	Yearly labor earnings (Males)			Yearly fraction of days spent at work (Males)			Yearly labor earnings (Females)			Yearly fraction of days spent at work (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	-14.146		21.385	0.001	**	0.000	29.796		18.775	0.000		0.001
Mother's age at respondent's birth is missing	-1008.754		696.388	-0.031	**	0.014	1072.730	*	618.425	0.003		0.017
Respondent's father has at least secondary education	543.159	**	262.219	-0.028	***	0.005	300.084		239.568	-0.018		0.007
Respondent's mother has at least secondary education	-931.101	***	278.276	-0.042	***	0.006	292.566		251.339	-0.003		0.008
Mother's employment at 14	-35.153		228.535	0.000		0.005	351.824		214.264	0.000		0.006
Father's employment at 14	443.752		368.615	0.016	**	0.007	15.072		392.799	0.001		0.010
Respondent lived with both parents at 14	191.037		543.657	0.001		0.010	118.952		510.779	0.018		0.014
Number of siblings at 14 if IT-SILC wave is 2005	-2.565		137.107	0.001		0.002	118.745		138.756	0.003		0.003
Number of siblings at 14 if IT-SILC wave is 2011	-143.024		171.147	-0.001		0.003	346.211	**	159.092	0.001		0.004
Quarter of birth - Reference category: October, November, December												
January, February, March	417.834		292.347	0.004		0.006	-458.132	*	268.451	-0.005		0.008
April, May, June	-51.141		295.902	0.000		0.006	-202.408		281.788	-0.013	*	0.008
July, August, September	83.657		291.587	0.006		0.006	9.249		273.508	0.003		0.008
Year of birth/10 (normalized to its minimum)	617.716		651.431	0.029	**	0.013	-941.106		626.503	0.026		0.017
Geographical area at t - Reference category: North-West												
North-East	-838.572	***	267.465	-0.013	**	0.006	-349.477		240.217	-0.003		0.007
Center	-1541.361	***	284.853	0.007		0.006	-1427.770	***	248.982	-0.018	**	0.007
South	-1536.664	***	484.407	-0.020	**	0.010	-1594.797	***	442.087	-0.037	***	0.012
Islands	-485.220	***	651.284	0.021		0.013	-428.115		629.941	-0.012		0.017
Regional unemployment rate at t	-207.836	***	43.432	-0.009	***	0.001	-316.111	***	417.891	-0.018	***	0.001
Regional employment rate at t	263.789	***	30.884	0.004	***	0.001	119.069	***	28.887	0.001		0.001
Regional growth rate at t	-4028.908		2670.057	0.024		0.084	-2344.243		2645.202	-0.035		0.101
IT-SILC wave 2011	246.732		332.791	-0.012	*	0.007	-153.359		307.608	-0.001		0.009
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	1441.680	***	453.798	0.031	***	0.009	1838.093	***	412.516	0.044	***	0.012
Between 1986 and 1991	854.756		731.313	0.030	**	0.015	1936.745	***	689.412	0.065	***	0.019
Between 1991 and 1996	193.692		1031.498	0.047	**	0.021	2149.440	**	992.597	0.106	***	0.028
After 1996	-1480.031		1386.918	0.022		0.029	1959.121		1338.022	0.091	**	0.038
Age at school completion	7450.237	***	1139.143	-0.003		0.002	566.074	***	115.719	0.005		0.003
Number of kids at t	885.457	***	112.791	0.011	***	0.003	-2699.539	***	95.848	-0.061	***	0.003
Fraction of time spent at work 1 year before diploma	975.960	*	588.968	0.038	***	0.012	-2690.928	***	747.287	-0.026		0.020
Constant at $t = 5$	-1553.102		3132.122	1.077	***	0.070	3974.224		3077.591	0.897	***	0.087
Constant at $t = 10$	5835.035	*	3137.120	1.045	***	0.070	5339.772	*	3106.162	0.836	***	0.088
Constant at $t = 15$	11244.890	***	3192.380	1.017	***	0.072	5171.108	*	3133.992	0.825	***	0.089
Constant at $t = 20$	15905.060	***	3187.798	0.954	***	0.072	7676.434	**	3173.279	0.839	***	0.090
Constant at $t = 25$	17136.460	***	3234.930	0.907	***	0.074	9026.959	***	3190.162	0.870	***	0.091
$\ln(\sigma^2)$	-0.095	***	0.006	-2.347	***	0.014	-0.474	***	0.006	-2.217	***	0.016

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table C2.2: Estimated coefficients of the measurement equations with time-constant unobserved heterogeneity

	Number of siblings at 14 (Males)			Employment 1 year before school completion (Males)			Number of siblings at 14 (Females)			Employment 1 year before school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	-0.012	***	0.003	0.000		0.001	-0.016	***	0.003	-0.001		0.001
Mother's age at respondent's birth is missing	-0.263	**	0.107	0.042	**	0.019	-0.381	***	0.106	-0.008		0.018
Respondent's father has at least secondary education	-0.022		0.040	-0.028	***	0.008	-0.118	***	0.041	-0.009		0.007
Respondent's mother has at least secondary education	-0.099	**	0.046	-0.025	***	0.009	-0.089	*	0.046	-0.006		0.008
Mother's employment at 14	-0.140	***	0.041	-0.004		0.007	-0.088	**	0.038	-0.011	*	0.007
Father's employment at 14	0.310	***	0.049	-0.040	***	0.009	0.221	***	0.056	0.005		0.011
Respondent lived with both parents at 14	-0.087		0.067	0.037	***	0.014	0.172	*	0.085	-0.009		0.014
Number of siblings at 14 if IT-SILC wave is 2005	–		–	0.004		0.003	–		–	0.004		0.003
Number of siblings at 14 if IT-SILC wave is 2011	–		–	0.011	***	0.004	–		–	0.016	***	0.003
Quarter of birth - Reference category: October, November, December												
January, February, March	-0.028		0.045	-0.008		0.009	0.119	**	0.047	0.014	*	0.008
April, May, June	-0.074		0.047	-0.013		0.009	0.168	***	0.048	0.017	**	0.008
July, August, September	-0.022		0.045	-0.006		0.009	0.079		0.048	0.013		0.009
Year of birth/10 (normalized to its minimum)	0.110		0.083	-0.030	*	0.016	-0.112		0.087	-0.025	*	0.014
Geographical area at birth - Reference category: North-West												
North-East	0.126	**	0.052	0.031	***	0.008	0.159	***	0.049	0.022	***	0.007
Center	-0.016		0.059	-0.015		0.011	-0.047		0.061	0.000		0.009
South	0.591	***	0.072	-0.012		0.015	0.717	***	0.076	-0.004		0.015
Islands	0.504	***	0.094	0.012		0.019	0.670	***	0.098	0.003		0.021
Regional unemployment rate at birth	-0.035	**	0.016	-0.010	***	0.003	-0.058	***	0.016	-0.009	***	0.003
Regional employment rate at birth	-0.027	***	0.007	-0.002		0.001	-0.027	***	0.006	-0.001		0.001
Regional growth rate at birth	-1.766	**	0.788	-0.277	*	0.012	1.005		0.833	-0.193		0.172
IT-SILC wave 2011	-0.498	***	0.037	-0.014		0.010	-0.533	***	0.037	-0.012		0.008
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	-0.174	***	0.062	0.042	***	0.012	-0.074		0.060	-0.011		0.011
Between 1986 and 1991	-0.405	***	0.098	0.050	***	0.020	-0.086		0.095	0.022		0.016
Between 1991 and 1996	-0.527	***	0.133	0.050	*	0.027	-0.195		0.138	0.029		0.024
After 1996	-0.567	***	0.182	0.119	***	0.035	-0.062		0.182	0.066	**	0.030
Constant	3.337	***	0.464	0.219	**	0.090	3.530	***	0.447	0.184	***	0.071
$\ln(\sigma^2)$	0.183	***	0.010	-3.068	***	0.022	0.138	***	0.014	-3.518	***	0.021

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table C2.3: Estimated coefficients of the treatment equation with time-constant unobserved heterogeneity

	Days (%) of nonemployment during the first 3 years after school completion (Males)			Days (%) of nonemployment during the first 3 years after school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	0.001		0.001	0.002	**	0.001
Mother's age at respondent's birth is missing	0.025		0.025	0.051	*	0.028
Respondent's father has at least secondary education	0.045	***	0.010	0.030	***	0.011
Respondent's mother has at least secondary education	0.040	***	0.010	0.055	***	0.011
Mother's employment at 14	0.019	**	0.009	0.008		0.010
Father's employment at 14	-0.012		0.015	-0.013		0.017
Respondent lived with both parents at 14	-0.013		0.021	-0.005		0.023
Number of siblings at 14 if IT-SILC wave is 2005	0.000		0.005	-0.005		0.005
Number of siblings at 14 if IT-SILC wave is 2011	-0.006		0.006	-0.017	**	0.006
Quarter of birth - Reference category: October, November, December						
January, February, March	-0.021	*	0.011	-0.012		0.013
April, May, June	-0.018		0.011	-0.005		0.013
July, August, September	-0.002		0.011	0.004		0.013
Year of birth/10 (normalized to its minimum)	0.004		0.025	0.047		0.028
Geographical area at birth - Reference category: North-West						
North-East	-0.030	***	0.011	-0.062	***	0.012
Center	0.036	***	0.012	0.074	***	0.013
South	0.020		0.017	0.072	***	0.019
Islands	-0.021		0.023	0.041		0.029
Average regional unemployment rate 3 years after diploma	0.009	***	0.002	0.013	***	0.002
Average regional employment rate 3 years after diploma	-0.008	***	0.001	-0.007	***	0.001
Average regional growth rate 3 years after diploma	0.132		0.311	0.624	*	0.341
IT-SILC wave 2011	0.023	*	0.013	0.031	**	0.014
Calendar year of t - Reference category: before 1981						
Between 1981 and 1986	-0.012		0.019	-0.050	**	0.021
Between 1986 and 1991	0.032		0.029	-0.047		0.031
Between 1991 and 1996	0.008		0.040	-0.066		0.044
After 1996	-0.026		0.054	-0.096		0.060
Average number of kids 3 years after diploma	-0.168	***	0.060	0.034		0.027
Age at school completion	-0.006		0.004	-0.021	***	0.005
Fraction of time spent at work 1 year before diploma	-0.345	***	0.021	-0.397	***	0.031
Constant	1.056	***	0.122	1.194	***	0.140
ln(σ^2)	-2.530	***	0.029	-2.453	***	0.032

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table C2.4: Estimated distribution of the discrete time-constant unobserved heterogeneity with $H = 5$ support points

	Location of the mass			Logistic weight of the probability masses (p^h)			Resulting probabilities (p^h)
	Coeff.		Std. Error	Coeff.		Std. Error	
a) Males							
θ^1	0.00		–	2.151	***	0.191	0.132
θ^2	-13268.00	***	854.94	1.774	***	0.253	0.091
θ^3	-4629.90	***	308.79	3.397	***	0.209	0.460
θ^4	-8334.58	***	542.71	2.978	***	0.216	0.302
θ^5	6033.00	***	432.65	–		–	0.015
b) Females							
θ^1	0.00		–	2.959	***	0.193	0.437
θ^2	-19597.73	***	500.62	1.476	***	0.227	0.099
θ^3	-9867.77	***	295.17	2.603	***	0.195	0.306
θ^4	10241.62	***	305.16	1.794	***	0.188	0.136
θ^5	24080.07	***	546.42	–		–	0.023

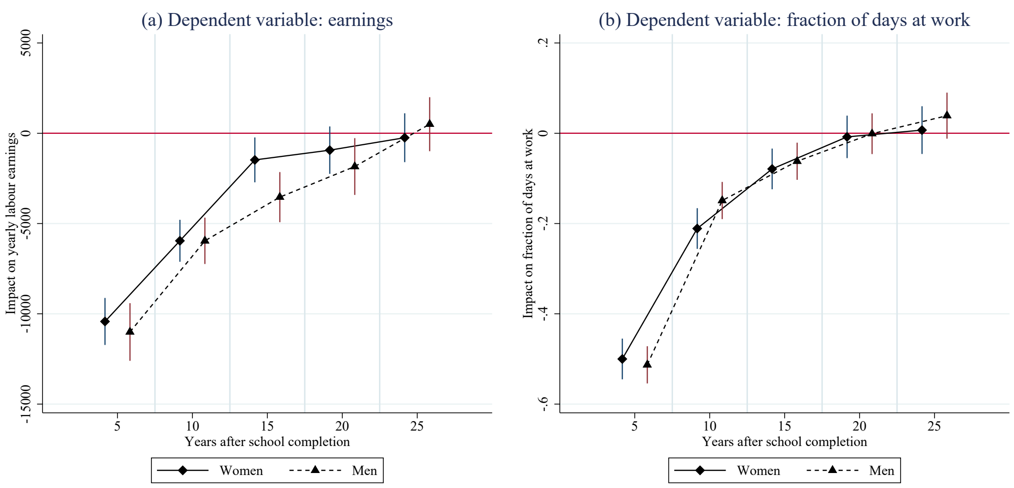
Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. Since the loading factor of one earnings equation is normalized to 1 for both genders, all the figures are in 2014 Euro.

Table C2.5: Estimated loading factors with time-constant unobserved heterogeneity (discrete distribution with $H = 5$ support points)

	Males			Females		
Equations	Loading factor		Std. Error	Loading factor		Std. Error
a) Measurement equations						
Number of siblings when 14 years old	-0.142	**	0.058	-0.074	***	0.026
Employment 1 year before school completion	-0.033	***	0.011	-0.006	*	0.005
b) Selection into treatment equation						
Days (%) of nonemployment 3 years after school completion	-0.069	***	0.017	-0.043	***	0.007
c) Labor market outcomes						
Yearly labor earnings 5 years after school completion	1.000		–	0.380	***	0.024
Yearly labor earnings 10 years after school completion	1.879	***	0.144	0.680	***	0.021
Yearly labor earnings 15 years after school completion	2.569	***	0.168	0.884	***	0.023
Yearly labor earnings 20 years after school completion	3.217	***	0.210	1.000		–
Yearly labor earnings 25 years after school completion	3.439	***	0.221	0.973	***	0.024
Yearly fraction of days spent at work 5 years after school completion	0.366	***	0.039	0.152	***	0.009
Yearly fraction of days spent at work 10 years after school completion	0.439	***	0.031	0.220	***	0.010
Yearly fraction of days spent at work 15 years after school completion	0.406	***	0.033	0.239	***	0.010
Yearly fraction of days spent at work 20 years after school completion	0.334	***	0.030	0.204	***	0.011
Yearly fraction of days spent at work 25 years after school completion	0.277	***	0.030	0.150	***	0.011

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Figure C2.1: Impact of nonemployment during the first 3 years after school completion on yearly labor earnings and fraction of days spent at work with time-constant unobserved heterogeneity



Notes: The vertical segments crossing the dots are 95% confidence intervals.

Table C2.6: Estimated coefficients of the covariates of the labor market outcome equations with time-varying unobserved heterogeneity

	Yearly labor earnings (Males)			Yearly fraction of days spent at work (Males)			Yearly labor earnings (Females)			Yearly fraction of days spent at work (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	20.806	*	11.245	0.000		0.000	9.113		8.950	0.000		0.000
Mother's age at respondent's birth is missing	329.345		358.722	-0.008		0.008	792.190	***	286.44	0.002		0.010
Respondent's father has at least secondary education	1445.685	***	132.032	0.000		0.003	902.926	***	106.394	0.001		0.004
Respondent's mother has at least secondary education	532.359	***	142.303	-0.002		0.003	446.787	***	113.122	0.003		0.004
Mother's employment at 14	-393.146	***	119.783	-0.003		0.003	364.577	***	98.355	-0.001		0.004
Father's employment at 14	494.609	***	191.530	0.004	**	0.004	123.384		179.471	0.005		0.006
Respondent lived with both parents at 14	-44.886		267.468	-0.004		0.006	-132.891		235.879	0.000		0.008
Number of siblings at 14 if IT-SILC wave is 2005	-256.002	***	66.715	0.001		0.001	161.295	***	52.917	-0.002		0.002
Number of siblings at 14 if IT-SILC wave is 2011	-445.398	***	79.238	-0.001		0.002	-4.512		66.883	0.001		0.002
Quarter of birth - Reference category: October, November, December												
January, February, March	7.548		149.086	-0.003		0.004	-391.449	***	127.477	0.000		0.004
April, May, June	-278.014	*	151.806	0.000		0.004	10.587		129.874	0.001		0.004
July, August, September	-474.224	***	148.496	-0.002		0.004	55.382		127.500	0.001		0.004
Year of birth/10 (normalized to its minimum)	-241.181		349.994	-0.013		0.008	-1525.899	***	295.127	-0.016		0.010
Geographical area at t - Reference category: North-West												
North-East	-619.835	***	137.398	-0.004		0.004	-665.015	***	117.967	-0.006		0.004
Center	-2546.118	***	151.540	0.002		0.004	-946.969	***	123.743	0.009	**	0.004
South	-1688.532	***	338.974	0.013	**	0.006	-362.773		274.421	0.022	***	0.008
Islands	-848.629	**	430.689	0.032	***	0.007	624.539		384.740	0.030	***	0.009
Regional unemployment rate at t	-113.750	**	45.115	-0.004	***	0.001	-134.470	***	40.576	-0.006	***	0.001
Regional employment rate at t	155.359	***	27.135	0.000		0.000	50.148	**	24.209	-0.007		0.001
Regional growth rate at t	-582.076		2423.789	-0.036		0.038	-1855.768		2252.019	-0.088	*	0.046
IT-SILC wave 2011	638.596	***	169.251	0.000		0.004	1.807		138.049	-0.003		0.005
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	561.075	**	223.859	0.004		0.006	834.730	***	192.126	0.008		0.006
Between 1986 and 1991	-252.934		375.067	0.004		0.010	233.212		314.834	0.004		0.011
Between 1991 and 1996	-1201.755	**	539.341	0.008		0.013	-217.472		461.812	0.009		0.016
After 1996	-2228.168	***	753.382	0.006		0.018	40.012		651.958	0.014		0.023
Age at school completion	941.353	***	59.829	-0.002		0.001	534.276	***	540.557	-0.002		0.002
Number of kids at t	950.919	***	71.227	0.004	**	0.002	-1385.903	***	57.741	-0.011	***	0.002
Fraction of time spent at work 1 year before diploma	-1868.949	***	234.217	-0.003		0.006	-4002.103	***	342.061	-0.021	**	0.009
Constant at $t = 5$	-19367.220	***	2499.585	0.188	***	0.043	-3153.973		2132.559	0.344	***	0.048
Constant at $t = 10$	-17571.980	***	2361.651	0.221	***	0.040	-4395.957	**	2161.705	0.222	***	0.050
Constant at $t = 15$	-19284.120	***	2404.844	0.124	***	0.044	-4897.454	**	2117.043	0.247	***	0.049
Constant at $t = 20$	-16297.360	***	2298.743	0.320	***	0.040	-4928.042	**	2127.937	0.342	***	0.048
Constant at $t = 25$	-16129.820	***	2330.063	0.296	***	0.040	-3663.224	*	2130.286	0.354	***	0.049
$\ln(\sigma^2)$	0.206	***	0.004	-3.786	***	0.008	-0.318	***	0.004	-3.508	***	0.009

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labour earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labour earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table C2.7: Estimated coefficients of the measurement equations with time-varying unobserved heterogeneity

	Number of siblings at 14 (Males)			Employment 1 year before school completion (Males)			Number of siblings at 14 (Females)			Employment 1 year before school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	-0.013	***	0.003	0.000		0.001	-0.016	***	0.003	-0.001		0.001
Mother's age at respondent's birth is missing	-0.271	**	0.108	0.042	**	0.019	-0.380	***	0.106	-0.006		0.018
Respondent's father has at least secondary education	-0.027		0.041	-0.026	***	0.008	-0.122	***	0.041	-0.009		0.007
Respondent's mother has at least secondary education	-0.108	**	0.046	-0.023	**	0.009	-0.089	*	0.046	-0.004		0.008
Mother's employment at 14	-0.138	***	0.041	-0.001		0.007	-0.088	**	0.038	-0.011	*	0.007
Father's employment at 14	0.306	***	0.050	-0.043	***	0.009	0.221	***	0.056	0.004		0.011
Respondent lived with both parents at 14	-0.082		0.067	0.039	***	0.014	0.169	**	0.086	-0.010		0.014
Number of siblings at 14 if IT-SILC wave is 2005	–		–	0.005		0.003	–		–	0.005	*	0.003
Number of siblings at 14 if IT-SILC wave is 2011	–		–	0.012	***	0.004	–		–	0.016	***	0.003
Quarter of birth - Reference category: October, November, December												
January, February, March	-0.026		0.045	-0.009		0.009	0.118	**	0.047	0.015	*	0.008
April, May, June	-0.074		0.047	-0.012		0.009	0.169	***	0.049	0.017	**	0.008
July, August, September	-0.020		0.045	-0.005		0.009	0.079		0.048	0.013		0.009
Year of birth/10 (normalized to its minimum)	0.119		0.084	-0.029	*	0.017	-0.105		0.087	-0.025	*	0.014
Geographical area at birth - Reference category: North-West												
North-East	0.128	**	0.053	0.028	***	0.008	0.163	***	0.049	0.022	***	0.007
Center	-0.010		0.059	-0.011		0.011	-0.048		0.062	0.002		0.009
South	0.592	***	0.073	-0.001		0.016	0.710	***	0.077	0.001		0.015
Islands	0.505	***	0.094	0.018		0.020	0.661	***	0.099	0.007		0.021
Regional unemployment rate at birth	-0.035	**	0.016	-0.010	***	0.003	-0.058	***	0.016	-0.008	***	0.003
Regional employment rate at birth	-0.026	***	0.007	-0.002	*	0.001	-0.028	***	0.006	-0.001		0.001
Regional growth rate at birth	-1.814	**	0.791	-0.261		0.164	-1.046		0.836	-0.206		0.169
IT-SILC wave 2011	-0.500	***	0.037	-0.012		0.010	-0.536	***	0.037	-0.012		0.008
Calendar year of t - Reference category: before 1981												
Between 1981 and 1986	-0.171	***	0.062	0.038	***	0.012	-0.073		0.060	-0.013		0.011
Between 1986 and 1991	-0.403	***	0.098	0.048	**	0.020	-0.084		0.095	0.020		0.016
Between 1991 and 1996	-0.522	***	0.133	0.043		0.027	-0.193		0.139	0.024		0.024
After 1996	-0.565	***	0.183	0.109	***	0.035	-0.068		0.183	0.059	**	0.030
Constant	3.433	***	0.465	0.235	***	0.089	3.556	***	0.451	0.184	***	0.071
$\ln(\sigma^2)$	0.184	***	0.010	-3.073	***	0.022	0.142	***	0.014	-3.520	***	0.021

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table C2.8: Estimated coefficients of the treatment equation with time-varying unobserved heterogeneity

	Days (%) of nonemployment during the first 3 years after school completion (Males)			Days (%) of nonemployment during the first 3 years after school completion (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Mother's age at respondent's birth	0.001		0.001	0.002	**	0.001
Mother's age at respondent's birth is missing	0.015		0.024	0.040		0.027
Respondent's father has at least secondary education	0.034	***	0.010	0.023	**	0.010
Respondent's mother has at least secondary education	0.022	**	0.010	0.045	***	0.011
Mother's employment at 14	0.015	*	0.009	0.006		0.009
Father's employment at 14	-0.005		0.014	-0.009		0.016
Respondent lived with both parents at 14	-0.013		0.020	-0.003		0.022
Number of siblings at 14 if IT-SILC wave is 2005	-0.001		0.005	-0.003		0.005
Number of siblings at 14 if IT-SILC wave is 2011	-0.005		0.005	-0.015	**	0.006
Quarter of birth - Reference category: October, November, December						
January, February, March	-0.016		0.011	-0.014		0.012
April, May, June	-0.018		0.011	-0.007		0.012
July, August, September	-0.004		0.011	0.004		0.012
Year of birth/10 (normalized to its minimum)	0.021		0.024	0.063	**	0.027
Geographical area at birth - Reference category: North-West						
North-East	-0.022	**	0.011	-0.060	***	0.011
Center	0.029	**	0.011	0.052	***	0.012
South	-0.006		0.016	0.035	*	0.018
Islands	-0.035		0.022	0.012		0.026
Average regional unemployment rate 3 years after diploma	0.007	***	0.002	0.010	***	0.002
Average regional employment rate 3 years after diploma	-0.006	***	0.001	-0.005	***	0.002
Average regional growth rate 3 years after diploma	-0.006		0.291	0.762	**	0.328
IT-SILC wave 2011	0.014		0.012	0.027	**	0.013
Calendar year of t - Reference category: before 1981						
Between 1981 and 1986	0.011		0.018	-0.033	*	0.020
Between 1986 and 1991	0.039		0.027	-0.034		0.030
Between 1991 and 1996	0.026		0.037	-0.037		0.042
After 1996	-0.018		0.050	-0.082		0.057
Average number of kids 3 years after diploma	-0.131	**	0.056	-0.002		0.024
Age at school completion	-0.003		0.004	-0.017	***	0.005
Fraction of time spent at work 1 year before diploma	-0.298	***	0.020	-0.350	***	0.028
Constant	1.083	***	0.115	1.153	***	0.132
ln(σ^2)	-2.643	***	0.026	-2.573	***	0.029

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table C2.9: Estimated distribution of the discrete time-varying unobserved heterogeneity with $H = 10$ support points

	Location of the mass				Logistic weight of the probability masses (p^h)			Resulting probabilities (p^h)
	$t = 5$	$t = 10$	$t = 15$	$t = 20/25$	Coeff.		Std. Error	
a) Males								
θ^1	0.00	0.00	0.00	0.00	0.890	***	0.133	0.045
θ^2	16123.09***	18555.35***	22615.87***	22180.61***	3.325	***	0.126	0.510
	(1026.68)	(828.50)	(909.56)	(510.55)				
θ^3	-5.114	-1583.01***	22103.38***	21996.60***	1.277	***	0.138	0.066
	(363.15)	(279.69)	(941.49)	(543.72)				
θ^4	387.92	18243.65***	22913.90***	21316.69***	2.307	***	0.126	0.184
	(302.67)	(829.60)	(912.61)	(507.13)				
θ^5	10745.48***	18276.38***	21379.39***	1856.64***	0.888	***	0.134	0.045
	(719.71)	(852.88)	(896.74)	(196.62)				
θ^6	248.43	6101.99***	445.38	19926.73***	0.957	***	0.131	0.048
	(349.94)	(307.07)	(510.41)	(512.03)				
θ^7	15517.97***	-1032.95***	18628.75***	19900.97***	0.653	***	0.141	0.035
	(1017.95)	(303.78)	(796.93)	(505.18)				
θ^8	332.38	-1279.99***	21592.88***	4927.54***	0.136		0.157	0.021
	(434.07)	(396.06)	(995.74)	(254.27)				
θ^9	15267.27***	18282.33***	1514.03***	18937.44***	0.453	**	0.145	0.008
	(993.91)	(903.88)	(501.95)	(496.73)				
θ^{10}	15839.88***	3500.28***	1411.94***	-955.01***	–		–	0.018
	(1052.84)	(244.84)	(498.58)	(244.73)				
b) Females								
θ^1	0.00	0.00	0.00	0.00	0.144		0.099	0.084
θ^2	13487.72***	16263.15***	19280.14***	12546.85***	1.694	***	0.069	0.396
	(755.95)	(740.81)	(443.27)	(748.39)				
θ^3	12546.85***	649.54**	15931.55***	16666.54***	-0.445	***	0.099	0.047
	(748.39)	(308.87)	(701.10)	(437.11)				
θ^4	-673.88***	15779.45***	16377.12***	18977.70***	0.629	***	0.077	0.136
	(185.44)	(738.52)	(660.36)	(458.99)				
θ^5	13129.48***	15789.42***	14937.07***	2700.26***	-0.370	***	0.110	0.050
	(806.24)	(767.28)	(632.06)	(202.16)				
θ^6	12943.14***	100.16	-574.51	13821.84***	-0.789	***	0.113	0.033
	(811.48)	(359.54)	(409.04)	(376.11)				
θ^7	-969.87**	14615.73***	14353.83***	3174.40***	-0.860	***	0.131	0.031
	(480.61)	(719.57)	(620.67)	(239.93)				
θ^8	-1180.99***	-156.62	16141.49***	17041.98***	0.194	**	0.084	0.088
	(300.18)	(283.82)	(668.55)	(426.95)				
θ^9	8739.69***	15626.11***	-268.93	12113.33***	-0.166		0.103	0.062
	(507.47)	(767.91)	(312.78)	(330.65)				
θ^{10}	-1145.74***	-439.08	-760.73**	17147.36***	–		–	0.073
	(327.83)	(341.63)	(350.53)	(425.63)				

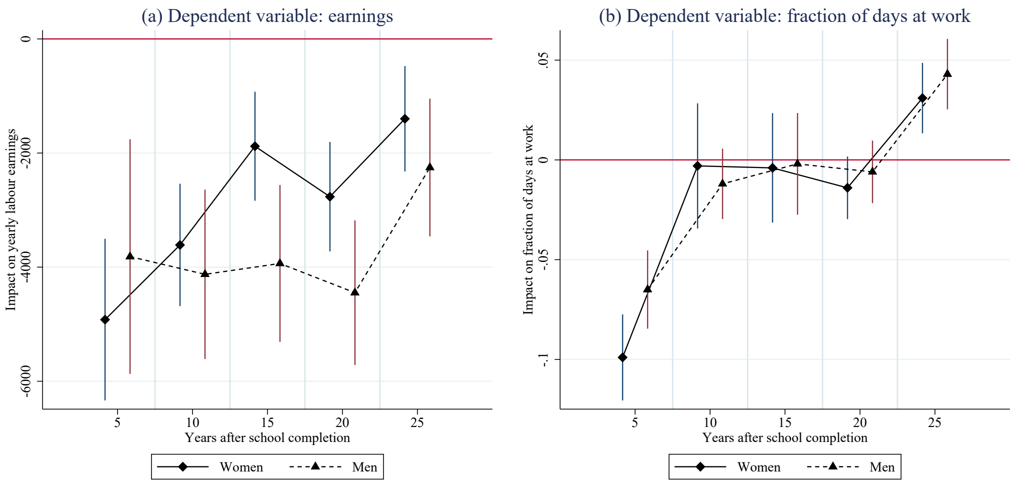
Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. Since the loading factors of the earnings equations are normalized to 1, all the figures are in 2014 Euro. The normalisation $\theta^1 = 0$ is innocuous: all the support points are indeed in deviation from the time-varying constant terms displayed in the last part of Table C2.6.

Table C2.10: Estimated loading factors with time-varying unobserved heterogeneity (discrete distribution with $H = 10$ support points)

	Males			Females		
Equations	Loading factor		Std. Error	Loading factor		Std. Error
a) Measurement equations						
Number of siblings when 14 years old	-0.029		0.023	-0.006		0.026
Employment 1 year before school completion	0.030	***	0.006	0.019	***	0.005
b) Selection into treatment equation						
Days (%) of nonemployment 3 years after school completion	-0.138	***	0.011	-0.170	***	0.011
c) Labor market outcomes						
Yearly labor earnings 5 years after school completion	1.000		–	1.000		–
Yearly labor earnings 10 years after school completion	1.000		–	1.000		–
Yearly labor earnings 15 years after school completion	1.000		–	1.000		–
Yearly labor earnings 20 years after school completion	1.000		–	1.000		–
Yearly labor earnings 25 years after school completion	1.000		–	1.000		–
Yearly fraction of days spent at work 5 years after school completion	0.558	***	0.036	0.610	***	0.038
Yearly fraction of days spent at work 10 years after school completion	0.462	***	0.022	0.558	***	0.027
Yearly fraction of days spent at work 15 years after school completion	0.420	***	0.017	0.527	***	0.023
Yearly fraction of days spent at work 20 or 25 years after school completion	0.333	***	0.009	0.401	***	0.011

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. The loading factors of the yearly labor earnings equations are normalized to 1.

Figure C2.2: Impact of nonemployment during the first 3 years after school completion on yearly labor earnings and fraction of days spent at work with time-varying unobserved heterogeneity



Notes: The vertical segments crossing the dots are 95% confidence intervals.

D. Full set of sensitivity analysis

Table D2.1: Impact of early nonemployment or unpaid internship/stage/training on labor market outcomes with time-varying unobserved heterogeneity

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment or unpaid internship/stage/training during the first 3 years after school completion (men)	-4526.72*** (1067.79)	-3914.81*** (772.27)	-4045.99*** (720.25)	-4084.16*** (675.11)	-2122.72*** (625.39)
Nonemployment or unpaid internship/stage/training during the first 3 years after school completion (women)	-5292.39*** (725.68)	-3596.06*** (550.48)	-1619.49*** (499.82)	-1997.51*** (494.46)	-412.97 (481.90)
b) Yearly fraction of days spent at work					
Nonemployment or unpaid internship/stage/training during the first 3 years after school completion (men)	-0.090*** (0.011)	-0.005 (0.010)	-0.003 (0.014)	-0.022** (0.009)	0.014 (0.010)
Nonemployment or unpaid internship/stage/training during the firs3 years after school completion (women)	-0.112*** (0.011)	-0.004 (0.017)	-0.001 (0.014)	-0.000 (0.008)	0.033*** (0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.2: Impact of early nonemployment during the first 2 years after school completion on labor market outcomes with time-varying unobserved heterogeneity

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first 2 years after school completion (men)	-3017.37*** (984.62)	-3338.88*** (717.11)	-2896.27*** (683.97)	-3090.98*** (613.93)	-940.90 (582.34)
Nonemployment during the first 2 years after school completion (women)	-4254.46*** (683.26)	-3082.40*** (524.82)	-1697.36*** (471.96)	-2242.27*** (463.27)	-1109.14*** (451.11)
b) Yearly fraction of days spent at work					
Nonemployment during the first 2 years after school completion (men)	-0.048*** (0.010)	-0.008 (0.009)	-0.001 (0.013)	0.000 (0.008)	0.039*** (0.008)
Nonemployment during the first during the firs3 years after school completion (women)	-0.075*** (0.011)	-0.003 (0.016)	-0.001 (0.013)	-0.000 (0.008)	0.033*** (0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.3: Impact of early nonemployment during the first 4 years after school completion on labor market outcomes with time-varying unobserved heterogeneity

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first	-5041.78***	-5066.98***	-4812.33***	-5569.02***	-3715.05***
4 years after school completion (men)	(1120.76)	(776.44)	(705.03)	(654.41)	(622.90)
Nonemployment during the first	-6116.17***	-4252.87***	-2492.40***	-3345.95***	-1867.65***
4 years after school completion (women)	(770.08)	(565.10)	(501.91)	(499.20)	(475.19)
b) Yearly fraction of days spent at work					
Nonemployment during the first	-0.098***	-0.018**	-0.002	-0.010	0.042***
4 years after school completion (men)	(0.010)	(0.009)	(0.014)	(0.008)	(0.009)
Nonemployment during the first	-0.134***	-0.011	-0.003	-0.018**	0.034***
4 years after school completion (women)	(0.016)	(0.017)	(0.015)	(0.009)	(0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.4: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity by including geographical area at birth and regional rates at birth in the outcome and treatment equations

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first	-3852.39***	-4255.17***	-4035.90***	-4502.19***	-2269.94***
3 years after school completion (males)	(1050.95)	(758.34)	(702.87)	(647.58)	(618.07)
Nonemployment during the first	-4965.48***	-3660.18***	-1901.50***	-2753.99***	-1361.83***
3 years after school completion (females)	(723.10)	(548.34)	(487.52)	(489.39)	(476.16)
(b) Yearly fraction of days spent at work					
Nonemployment during the first	-0.065***	-0.012	-0.003	-0.005	0.043***
3 years after school completion (males)	(0.010)	(0.009)	(0.013)	(0.008)	(0.009)
Nonemployment during the first	-0.101***	-0.006	-0.005	-0.014	0.030***
3 years after school completion (females)	(0.011)	(0.016)	(0.014)	(0.008)	(0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.5: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity by including geographical area at birth, regional rates at birth and average regional rates across 3 years after school completion in the outcome equations

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first	-4001.40***	-4394.44***	-4136.76***	-4551.30***	-2329.76***
3 years after school completion (males)	(1050.16)	(757.81)	(702.55)	(646.73)	(618.08)
Nonemployment during the first	-4984.55***	-3672.19***	-1890.95***	-2743.19***	-1353.11***
3 years after school completion (females)	(723.56)	(548.74)	(489.54)	(489.94)	(477.29)
(b) Yearly fraction of days spent at work					
Nonemployment during the first	-0.067***	-0.015	-0.005	-0.007	0.043***
3 years after school completion (males)	(0.010)	(0.009)	(0.013)	(0.008)	(0.009)
Nonemployment during the first	-0.101***	-0.006	-0.005	-0.014*	0.030***
3 years after school completion (females)	(0.011)	(0.016)	(0.014)	(0.008)	(0.009)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.6: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using only individuals born in 1960s

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first	-3612.84***	-3952.90***	-3429.31***	-4958.71***	-2000.35**
3 years after school completion (males)	(1950.23)	(1470.90)	(1193.26)	(919.97)	(791.84)
Nonemployment during the first	-4373.17***	-3431.93***	-1892.24***	-2834.94***	-1638.87***
3 years after school completion (females)	(1181.81)	(935.12)	(746.15)	(679.62)	(586.33)
(b) Yearly fraction of days spent at work					
Nonemployment during the first	-0.095***	-0.013	0.003	0.004	0.053***
3 years after school completion (males)	(0.012)	(0.017)	(0.031)	(0.011)	(0.011)
Nonemployment during the first	-0.099***	-0.002	-0.001	0.001	0.028***
3 years after school completion (females)	(0.018)	(0.025)	(0.021)	(0.011)	(0.012)
Observations (men)	2724	2717	2706	2674	2571
Observations (women)	2570	2491	2400	2310	2188

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.7: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using only individuals born in 1970s-1980s

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first	-4987.74***	-4672.43***	-4362.71***	-2239.73***	-2522.86
3 years after school completion (males)	(1122.23)	(792.34)	(810.93)	(936.45)	(2318.14)
Nonemployment during the first	-5891.23***	-3985.04***	-1700.73***	-2815.37***	-2432.92
3 years after school completion (females)	(833.28)	(612.94)	(612.38)	(703.74)	(1810.72)
(b) Yearly fraction of days spent at work					
Nonemployment during the first	-0.066***	-0.003	0.008	0.006	0.018
3 years after school completion (males)	(0.012)	(0.014)	(0.015)	(0.012)	(0.023)
Nonemployment during the first	-0.099***	-0.013	-0.004	-0.054***	0.056***
3 years after school completion (females)	(0.013)	(0.018)	(0.017)	(0.014)	(0.021)
Observations (men)	2672	2593	2158	1273	221
Observations (women)	2329	2231	1835	1073	235

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.8: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using only individuals we follow up to 25 years later

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first 3 years after school completion (males)	-3369.32 (2096.02)	-3487.70** (1497.57)	-3105.83*** (1194.46)	-4220.93*** (917.24)	-1822.05** (773.47)
Nonemployment during the first 3 years after school completion (females)	-3885.86*** (1250.19)	-3073.97*** (975.09)	-1455.63*** (778.74)	-2707.29*** (683.06)	-1502.73*** (595.14)
(b) Yearly fraction of days spent at work					
Nonemployment during the first 3 years after school completion (males)	-0.085*** (0.013)	-0.014 (0.017)	0.005 (0.032)	0.006 (0.010)	0.051*** (0.010)
Nonemployment during the first 3 years after school completion (females)	-0.082*** (0.020)	-0.002 (0.032)	-0.003 (0.019)	-0.010 (0.012)	0.035*** (0.011)
Observations (men)	2792	2792	2792	2792	2792
Observations (women)	2423	2423	2423	2423	2423

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.9: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using only individuals born and graduated in Central or Northern Italy

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first 3 years after school completion (males)	-5486.99*** (1334.70)	-4632.72*** (954.99)	-3703.93*** (906.23)	-3329.53*** (851.74)	-1665.45*** (772.62)
Nonemployment during the first 3 years after school completion (females)	-4982.19*** (906.35)	-3573.52*** (675.16)	-1476.99** (615.76)	-1518.21*** (600.17)	-607.70 (580.89)
(b) Yearly fraction of days spent at work					
Nonemployment during the first 3 years after school completion (males)	-0.085*** (0.011)	-0.013 (0.011)	-0.002 (0.015)	-0.014 (0.009)	0.009 (0.010)
Nonemployment during the first 3 years after school completion (females)	-0.079*** (0.015)	-0.010 (0.018)	0.005 (0.013)	0.008 (0.009)	0.024** (0.010)
Observations (men)	3739	3693	3429	2807	2000
Observations (women)	3552	3466	3176	2581	1854

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.10: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using only individuals born and graduated in Southern Italy or Islands

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
a) Yearly labor earnings (€)					
Nonemployment during the first 3 years after school completion (males)	-3740.46**	-2691.44*	-4783.19***	-3813.67***	-1243.61
	(1861.58)	(1509.31)	(1418.20)	(1297.11)	(1448.55)
Nonemployment during the first 3 years after school completion (females)	-3199.77**	-2811.95**	-681.14	-4019.22***	334.50
	(1415.13)	(1285.06)	(1224.82)	(1299.04)	(1404.07)
(b) Yearly fraction of days spent at work					
Nonemployment during the first 3 years after school completion (males)	-0.106***	0.018	-0.017	-0.016	0.003
	(0.035)	(0.031)	(0.038)	(0.030)	(0.034)
Nonemployment during the first 3 years after school completion (females)	-0.107***	0.009	0.002	0.005	0.017
	(0.036)	(0.052)	(0.038)	(0.030)	(0.029)
Observations (men)	1657	1617	1435	1140	792
Observations (women)	1347	1256	1059	802	569

Notes: Labor earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.11: Impact of early nonemployment on labor market outcomes with time-varying unobserved heterogeneity using daily earnings as outcome variable

	Years since school completion				
	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
Nonemployment during the first 3 years after school completion (males)	-8.75***	-10.55***	-11.02***	-13.44***	-8.14***
	(2.41)	(1.90)	(1.72)	(1.66)	(1.66)
Nonemployment during the first 3 years after school completion (females)	-11.70***	-9.40***	-6.50***	-8.58***	-5.84***
	(1.70)	(1.40)	(1.30)	(1.31)	(1.36)
Observations (men)	5396	5310	4864	3947	2792
Observations (women)	4899	4722	4235	3383	2423

Notes: Labor daily earnings are in 2014 prices and deflated by the ISTAT consumer price index.
* Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are reported in parentheses.

Table D2.12: Estimated impacts of early nonemployment during the first 3 years after diploma on future daily labor earnings, relative to the average in t for individuals who did not experienced early nonemployment

Years since school completion	$t = 5$	$t = 10$	$t = 15$	$t = 20$	$t = 25$
Men	-21.98%	-19.05%	-18.00%	-20.19%	-12.09%
Women	-25.29%	-18.04%	-12.25%	-14.55%	-8.87%

Notes: These figures are computed by evaluating the change in the daily labor earnings in a year implied by the estimated coefficients reported in Table D2.11 relative to the average daily labor earnings in t of individuals who did not experienced nonemployment after diploma.

Appendix (Ch. 3)

Table A3.1: Articles included in the meta-analysis ($N = 85$)

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Apouey et al. (2019)	7	SAH, PH, MH	Australia	HILDA	2001-2014	FE	+, 0	No
Ardito et al. (2020)	1	HC	Italy	WHIP	2001-2014	IV	+	O, Ph, I
Atalay and Barrett (2014)	36	SAH, PH, MH	Australia	NHSs	1995-2008	IV	0 (SAH), 0/+ (PH, MH)	G
Atalay et al. (2019)	12	MH	Australia	HILDA	2012-2016	FD-IV	0	No
Bamia et al. (2008)	92	M	Greece	EPIC	1994-2006	Other	–	No
Barrett and Kecmanovic (2013)	35	MH	Australia	HILDA	2007	Other	0	V
Bauer and Eichenberger (2018)	1	SAH, PH	Switzerland	Swiss LFS	2004-2015	DiD	0	No
Behncke (2012)	247	SAH, PH, MH	England	ELSA	2002-2007	PSM, IV	– (SAH, PH), 0 (MH)	No
Belloni et al. (2016)	31	MH	10 EU	SHARE	2004-2013	FE-IV	0	O
Bertoni and Brunello (2017)	23	MH	Japan	JPSS	2008-2013	IV	–	No
Bianchini and Borella (2016)	16	MH	10 EU	SHARE	2004-2012	FE-IV	0, + (RD)	RD
Binh Tran and Zikos (2019)	8	SAH, PH, MH	Australia	HILDA	2002-2015	FE-IV	+	No
Blake and Garrouste (2019)	9	SAH, PH, MH	France	Health Barometer	1994-2003	DiD	0	E
Bloemen et al. (2017)	55	M	Netherlands	Administrative Data	2000-2005	FE-IV	0	No
Bonsang et al. (2012)	444	MH	USA	HRS	1998-2008	FE-IV	–	No
Bonsang and Klein (2012)	134	SAH	Germany	GSOEP	1995-2010	FE-IV	+, 0 (MI)	V
Bozio et al. (2021)	2	M	France	Administrative Data	2004-2017	IV	0	No
Brockmann et al. (2009)	104	M	Germany	Gmunder Ersatzkasse	1990-2004	Other	+ (M), 0 (F)	G, Ph, I
Butterworth et al. (2006)	256	MH	Australia	NSMHWB	1997	Other	0	O, AC
Calvo et al. (2013)	152	PH, MH	USA	HRS	1992-2010	FE-IV	–	T
Carrino et al. (2020)	1	SAH, PH, MH	UK	Understanding Society	2009-2016	DiD	0	O
Celidoni et al. (2017)	55	MH	10 EU	SHARE	2004-2012	IV	+	T
Celidoni and Rebba (2017)	51	HC	10 EU	SHARE	2004-2012	FE-IV	0	MS
Che and Li (2018)	10	SAH	China	CHNS	1991-2006	IV	+	No
Chung et al. (2009)	64	PH	USA	HRS	1992-2002	FE-IV	–	AC, I, O
Coe et al. (2012)	153	MH	USA	HRS	1996-2008	IV	+, 0	O
Coe and Zamarro (2011)	506	SAH, MH	11 EU	SHARE	2004-2007	IV	+ (SAH), 0 (MH)	No
Dave et al. (2008)	485	SAH, PH, MH	USA	HRS	1992-2005	FE	–	MS, V
Dayaram and McGuire (2019)	1	PH, MH	Australia	HILDA	2003-2015	PSM	0	No
Eibich (2015)	268	SAH, PH, MH, HC	Germany	GSOEP	2002-2009	RDD	+ (SAH, MH, HC), 0 (PH)	E
Eyjólfsdóttir et al. (2019)	6	PH, M	Sweden	LNU, LISA, SWEOLD	2004-2014	PSM	0	No
Fé and Hollingsworth (2016)	14	SAH, MH	UK	BHPS	1991-2005	RDD	– (SAH), + (MH)	No
Feng et al. (2020)	12	PH	China	CHARLS	2001-2015	RDD	– (M), 0 (F)	E, G
Fitzpatrick and Moore (2018)	84	M	USA	MCOD, SSDMF	1979-2012	RDD	– (M), 0 (F)	G, E
Frimmel and Pruckner (2020)	7	HC	Austria	ASSD	1998-2012	FE-IV	+, 0 (F)	O, G
Gill et al. (2006)	95	MH	Australia	HILDA	2002-2003	Other	0	No
Godard (2016)	90	PH	8 EU	SHARE	2004-2011	FE-IV	0	O
Gorry et al. (2018)	84	SAH, MH, PH, HC	USA	HRS	1992-2014	IV	+, 0 (MH)	No
Grip et al. (2012)	121	SAH, MH, HC	Netherlands	Administrative Data	1997-2006	RDD	+ (MH), 0 (HC, SAH)	O, I
Grøtting and Lillebø (2020)	5	PH, HC, M	Norway	NORLAG	2002-2012	RDD	0, + (M PH)	No
Hagen (2018)	40	HC, M	Sweden	LOUISE	1987-2010	DiD	0	No
Hallberg et al. (2015)	64	M	Sweden	Administrative Data	1985-2010	DiD	+, 0	No
Heller-Sahlgren (2017)	67	MH	10 EU	SHARE	2004-2012	FE-IV	– (M), 0 (F)	E, G, O
Hernaes et al. (2013)	111	M	Norway	Administrative Data	1992-2010	IV	0	No
Hessel (2016)	54	SAH, PH	12 EU	EU-SILC	2009-2012	RE-IV	+, 0 (Chronic)	No

(continued on next page)

Table A3.1: *Continued from previous page*

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Horner and Cullen (2016)	23	PH, MH, HC	USA	Administrative Data	1997-2009	IV	0, + (Chronic)	No
Hult et al. (2010)	57	M	Sweden	Health Monitoring	1971-1993	Other	0	Ph
Insler (2014)	198	SAH	USA	HRS	1992-2010	FE-IV	+	No
Johnston and Lee (2009)	149	SAH, PH, MH	UK	HSE	1997-2005	RDD	+, 0 (PH)	No
Jokela et al. (2010)	161	PH, MH	UK	WhiteHall II Cohort Study	1991-2006	Other	+ (MH), 0 (PH)	T, RD
Kajitani et al. (2017)	35	MH	Japan	NSJE	1987-2002	IV	0	O
Kalwij et al. (2013)	11	M	Netherlands	IPO	1996-2010	Other	0	I
Kim and Koh (2020)	1	SAH	Singapore	SLP	2015-2019	RDD	+	No
Kim and Moen (2002)	364	MH	USA	Cornell Retirement Study	1994-1999	Other	0	G, MS, Ph
Kolodziej and García-Gómez (2019)	20	MH	11 EU	SHARE	2004-2013	IV	+ (F), 0 (M)	MS, Ph, G
Kuhn et al. (2020)	19	M	Austria	ASSD	1972-2017	IV	0 (F), – (M)	Ph, O, G
Kuusi et al. (2020)	4	MH, HC	Finland	Statistics Finland	2000-2012	FE-IV	+, 0 (HC)	G, O, I
Latif (2011)	59	MH	Canada	CNPHS	1994-2006	FE-IV	+	AC, MS
Lei and Liu (2018)	14	MH	China	CHARLS	2011-2015	FE-IV	+, 0 (M), 0, – (F)	G, O
Litwin (2007)	51	M	Israel	NHS	1997-2004	Other	0	No
Lucifora and Vigani (2018)	16	HC	10 EU	SHARE	2004-2006	FE-IV	0	G, O
Mandal and Roe (2007)	117	MH	USA	HRS	1992-2002	IV	+, –	No
Mazzonna and Peracchi (2017)	138	SAH, MH	10 EU	SHARE	2004-2006	FD-IV	–, 0 (F SAH)	O
Mein et al. (2003)	348	PH, MH	UK	WhiteHall II Cohort Study	1991-1995	Other	+, 0 (PH F)	No
Messe and Wolff (2019b)	9	SAH, PH	France	LFS	2013-2016	DiD	+ (M), 0 (F PH)	G, O
Messe and Wolff (2019a)	7	SAH, PH	France	INSEE	2012	IV	0	No
Mojon-Azzi et al. (2007)	80	SAH, PH, MH	Switzerland	SHP	1999-2003	Other	+	No
Mosca and Barrett (2016)	17	MH	Ireland	TILDA	2009-2013	FD	0	V
Müller and Shaikh (2018)	49	SAH	19 C.	SHARE	2004-2013	RDD	+	No
Neuman (2008)	235	SAH, PH, MH	USA	HRS	1992-2004	IV	0, + (SAH)	No
Nielsen (2019)	13	PH, HC, M	Denmark	Administrative Data	1980-2010	IV, RDD	0 (IV), 0, + (HC)	G, AC, T
Oksanen et al. (2011)	103	MH	Finland	National Records	1995-2004	Other	+	No
Oshio and Kan (2017)	28	SAH, MH	Japan	PSMOA	2005-2014	FE-IV	+, 0 (F SAH)	G
Picchio and van Ours (2020)	12	MH	Netherlands	LISS	2007-2018	RDD	+ (M), 0 (F)	G, MS
Quaade et al. (2002)	63	M	Denmark	Administrative Data	1986-1996	Other	–	No
Rijs et al. (2012)	41	SAH	Netherlands	LASA	1995-2009	Other	0	E, AC
Roberts et al. (2011)	51	MH	UK	WhiteHall II Cohort Study	1985-1988	Other	0	No
Rose (2020)	4	SAH, PH, MH, HC, M	England	ELSA, BHPS	1990-2011	RDD, FE-IV	+, 0, – (M M)	G, E, MS
Shai (2018)	44	SAH, HC	Israel	IHS, SHARE	1997-2013	DiD	0 (HC), +	E
Syse et al. (2017)	35	PH, MH	Norway	NORLAG	2002-2007	Other	+, 0	No
Tsai et al. (2005)	116	M	USA	Shell Oil	1973-2003	Other	0	No
Westerlund et al. (2009)	340	SAH	France	GAZEL	1990-2006	Other	+	O
Wu et al. (2016)	49	M	USA	HRS	1992-2010	Other	–	Ph
Zhang et al. (2018)	27	SAH, HC	China	CHARLS	2011-2013	RDD	0, – (F)	G
Zhu (2016)	51	SAH, PH, MH	Australia	HILDA	2001-2011	FE-IV	+	No

The sign of the effect is based on the value of t-stat: "–" means $t \leq -1.96$; "+" is for $t \geq 1.96$; "0" when $-1.96 < t < 1.96$. Note that the sign is positive even when articles estimate a negative effect of postponed retirement;

Identification Strategy: PSM = Propensity Score Matching; IV = Instrumental Variables; DiD = Difference-in-differences; FE = Fixed Effects; RDD = Regression Discontinuity Design;

Other methods = Ordinary Least Squares, Duration models, Logit, Multinomial logit, Ordered probit. .

Outcome: MH = Mental health; SAH = Self-assessed / General health; PH = Physical health; HC = Healthcare utilization; M = Mortality.

Effects: M = Males; F = Females; B = Both; MI = Mandatory/Involuntary; ER = Early retirement; PP = Postponed; SR = Short-run; LR = Long-run; RD = Retirement duration. .

Heterogeneity: G = Gender; O = Occupation; Ph = Previous health; MS = Marital status; E = Educational attainment; I = Income; V = Voluntariness; AC = Age cohort; T = Timing; C = Country; LS = Living standard. .

Table A3.2: Descriptive statistics of explanatory variables used in the meta-regressions by the sign of the partial correlation coefficient

	$r \leq 0$			$r > 0$		
	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.
Scimago subject areas						
Multi area	32	0.248	0.434	55	0.307	0.463
Economics/Business	36	0.279	0.450	52	0.291	0.455
Medicine/Psychology	61	0.473	0.501	72	0.402	0.492
Health outcomes						
Mortality (Reference category)	21	0.163	0.371	21	0.117	0.323
General and self-reported health	18	0.140	0.348	28	0.156	0.364
Physical health	33	0.256	0.438	38	0.212	0.410
Mental health	38	0.295	0.438	67	0.374	0.485
Healthcare utilization	19	0.147	0.356	25	0.140	0.348
Identification strategies						
Other methods (Reference category)	12	0.093	0.292	26	0.145	0.353
Regression discontinuity design (RDD)	22	0.171	0.378	37	0.207	0.406
Instrumental variables (IV)	60	0.465	0.501	92	0.514	0.501
Difference-in-differences (DiD)	14	0.109	0.312	9	0.050	0.219
Propensity score matching (PSM)	8	0.062	0.242	9	0.050	0.219
Fixed-effects/First-differences	13	0.101	0.302	6	0.034	0.180
Institutional contexts						
Statutory retirement (Reference category)	67	0.519	0.502	132	0.737	0.441
Mandatory or involuntary retirement	21	0.163	0.371	11	0.061	0.241
Early retirement	25	0.194	0.397	24	0.134	0.342
Postponed retirement	16	0.124	0.331	12	0.067	0.251
Geographical areas						
Multi-country analyses (Reference category)	19	0.147	0.356	22	0.123	0.329
Europe	57	0.442	0.499	86	0.480	0.501
Extra-European countries	53	0.411	0.494	71	0.397	0.491
Sex						
Males (Reference category)	38	0.295	0.458	64	0.358	0.481
Females	31	0.240	0.429	62	0.346	0.477
Males+Females	60	0.465	0.501	53	0.296	0.458
Calculation of t-statistic						
from 95% CI or from OR (Reference category)	13	0.101	0.302	27	0.151	0.359
t-statistic from $\widehat{\beta}_i/SE_i$	116	0.899	0.302	152	0.849	0.359
Birth cohorts						
Other (Reference category)	86	0.667	0.473	126	0.704	0.458
Only birth cohorts $\leq$ 1950	43	0.333	0.473	53	0.296	0.458
Type of previous occupation						
White collars (Reference category)	5	0.039	0.194	11	0.061	0.241
Blue collars	11	0.085	0.280	13	0.073	0.260
Not specified	113	0.876	0.331	155	0.866	0.342
Study-related characteristics						
Google scholar citations per year	129	12.826	10.943	179	10.046	10.418
Scimago Journal Ranking	129	1.895	1.449	179	1.763	1.069
Year of publication	129	2015.023	4.523	179	2015.637	4.728
Observations		129			179	

Notes: Females+Males = observations for which authors do not separate estimates for men and women. Other methods = OLS regressions and non-linear models (logit, multinomial logit, ordered probit and Cox proportional hazard models).
[a] At the time of publication, some journals did not have the SJR index yet, either because they were published in too recent years or because the journal was not indexed yet in Scimago. In these cases, we assigned to the journal the available value of the SJR index which was chronologically closer.

To check whether publication bias varies across different disciplines, we distinguish the results in three subject areas according to the Scimago classification: medicine/psychology, economics/business and multi-area journals. Table A3.4 shows the results.

Table A3.3: Heterogeneity in the estimated effects of retirement on health (FAT-PET specification)

	Bayesian Model Averaging[a]			Weighted-Average Least Square $(q=1)$[b]		Weighted-Average Least Square $(q=0.5)$[b]		OLS check after BMA[c]			OLS check after WALS[d]		
	PM	PSD	PIP	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Focus Regressors													
Publication bias	**0.308**	**0.174**	**1.000**	**0.245**	**0.175**	**0.217**	**0.178**	0.345		0.258	0.376		0.250
Precision effect	**0.003**	**0.003**	**1.000**	**0.012**	**0.004**	**0.012**	**0.004**	0.001	*	0.001	0.006	**	0.003
Auxiliary Regressors													
Google scholar citations per year	0.000	0.000	0.230	0.000	0.000	0.000	0.000	–		–	–		–
Scimago Journal Ranking	0.000	0.001	0.300	**-0.002**	**0.001**	**-0.002**	**0.001**	–		–	-0.002	***	0.001
Year of publication	**0.000**	**0.000**	**0.630**	**0.001**	**0.000**	**0.001**	**0.000**	0.000	***	0.000	0.001	***	0.000
Scimago subject areas (Reference category: Multi-area journals)													
Economics/Business	0.000	0.000	0.060	0.000	0.002	0.000	0.002	–		–	–		–
Medicine/Psychology	0.000	0.000	0.070	0.001	0.002	0.001	0.002	–		–	–		–
Health outcomes (Reference category: Mortality)													
General and self-reported health	**0.011**	**0.003**	**1.000**	**0.010**	**0.002**	**0.010**	**0.002**	0.010	***	0.002	0.011	***	0.002
Physical health	0.000	0.001	0.130	0.001	0.001	0.001	0.001	–		–	–		–
Mental health	**0.008**	**0.002**	**0.990**	**0.007**	**0.002**	**0.007**	**0.002**	0.008	***	0.002	0.007	***	0.002
Healthcare utilization	0.001	0.001	0.390	**0.001**	**0.001**	**0.001**	**0.001**	–		–	0.002	*	0.001
Identification strategies (Reference category: Other methods)													
Regression discontinuity design (RDD)	0.000	0.001	0.100	**-0.006**	**0.004**	**-0.005**	**0.004**	–		–	-0.002	*	0.001
Instrumental variables (IV)	0.000	0.001	0.070	-0.004	0.005	-0.004	0.005	–		–	–		–
Difference-in-differences (DiD)	0.000	0.001	0.050	-0.001	0.004	-0.001	0.004	–		–	–		–
Propensity score matching (PSM)	0.000	0.003	0.050	**-0.010**	**0.008**	**-0.010**	**0.009**	–		–	-0.009		0.009
Fixed-effects/First-differences	**-0.013**	**0.003**	**1.000**	**-0.014**	**0.005**	**-0.014**	**0.005**	-0.013	***	0.003	-0.012	***	0.002
Institutional contexts (Reference category: Statutory retirement)													
Mandatory or involuntary retirement	**-0.026**	**0.008**	**0.970**	**-0.021**	**0.006**	**-0.022**	**0.007**	-0.027	***	0.008	-0.027	***	0.008
Early retirement	0.000	0.001	0.100	-0.001	0.001	-0.001	0.001	–		–	–		–
Postponed retirement	-0.001	0.002	0.230	**-0.005**	**0.002**	**-0.006**	**0.002**	–		–	-0.006	***	0.002
Geographical areas (Reference category: Multi-country analyses)													
Europe	0.000	0.000	0.050	-0.002	0.002	-0.002	0.002	–		–	–		–
Extra-European countries	0.000	0.001	0.060	**-0.003**	**0.002**	**-0.003**	**0.002**	–		–	-0.001		0.001
Sex (Reference category: Males)													
Females	0.000	0.000	0.050	0.000	0.001	0.000	0.001	–		–	–		–
Males+Females	0.000	0.001	0.080	**-0.002**	**0.002**	**-0.003**	**0.002**	–		–	-0.003		0.002
Calculation of t-statistic (Reference category: from 95% CI)													
t-statistic from $\widehat{\beta}_i / SE_i$	-0.001	0.002	0.260	-0.002	0.003	-0.002	0.003	–		–	–		–
Type of previous occupation (Reference category: White collars)													
Blue collars	0.000	0.000	0.040	0.001	0.002	0.001	0.002	–		–	–		–
Not specified	0.000	0.000	0.050	0.001	0.001	0.001	0.002	–		–	–		–
Birth cohorts (Reference category: Others)													
Birth cohorts $\leq$ 1950	**-0.002**	**0.002**	**0.650**	-0.001	0.001	-0.001	0.001	-0.003	***	0.001	–		–

Notes: The results are from the FAT-PET specification by using the inverse of the SE_i^2 as weights. PM = Posterior Mean of the coefficient; PSD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability. The number of observations (studies) is 308 (85). Auxiliary variables for which the PIP is above 0.5 in BMA or the corresponding one-standard error band does not include zero in WALS are in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

[a] In the BMA, we use the uniform distribution for model priors, the Zellner's g prior for the distributions of the coefficients and a Markov Chain Monte Carlo algorithm to search over the model space, by distinguishing between focus and auxiliary regressors.

[b] $q = 1$ indicates the Laplace model prior distribution; $q = 0.5$ implies the Subbotin model prior distribution.

[c] The model specification under "OLS" includes those variables which have a PIP > 0.5 in BMA ($R^2 = 0.30$).

[d] The second model specification under "OLS" includes those variables which are relevant according to WALS ($R^2 = 0.34$).

Table A3.4: FAT-PET and PEESE tests and corrections for publication bias by subject area

	FAT-PET								PEESE	
	(1) OLS		(2) WLS-FE		(3) WLS-FE[(a)]		(4) FAIVE		(5) WLS-FE	
Publication bias in economics/business	0.256	(0.376)	0.384	(0.401)	0.385	(0.391)	0.523*	(0.057)	19.434*	(0.081)
Publication bias in medicine/psychology	0.612	(0.570)	0.350	(0.371)	0.357	(0.374)	0.212	(0.742)	9.485	(0.651)
Publication bias in multi-area	0.492	(0.557)	0.250	(0.635)	0.241	(0.659)	0.107	(0.845)	3.378	(0.762)
Precision effect in economics/business	0.003	(0.527)	0.000	(0.982)	0.000	(0.981)	-0.001	(0.821)	0.001	(0.871)
Precision effect in medicine/psychology	-0.005	(0.645)	0.000	(0.713)	0.000	(0.737)	0.001	(0.534)	0.001	(0.283)
Precision effect in multi-area	-0.002	(0.800)	0.004**	(0.040)	0.004*	(0.051)	0.005*	(0.078)	0.004**	(0.023)
R^2	0.058		0.082		0.082		0.039		0.078	

We report in parentheses wild cluster bootstrap p-values obtained from the wild cluster bootstrap-t procedure proposed by Cameron et al. (2008b), with clusters at study level (5,000 bootstraps using the Webb's (2014) six-point distribution as weights). We report wild cluster bootstrap p-values to take into account that, in each subject area, the number of clusters is small (from 16 to 36). The number of observations (studies) is 308 (85), 88 (16) in economics/business, 133 (36) in medicine/psychology, and 87 (33) in the residual multi-area category.

(a) The inverse of the square root of the sample size is used instead of $SE(r_i)$ as precision measure.

Table A3.5: FAT-PET and PEESE tests and corrections for publication bias without winsorization

	FAT-PET								PEESE[(b)]	
	(1) OLS		(2) WLS-FE		(3) WLS-FE[(a)]		(4) FAIVE		(5) WLS-FE	
Publication bias	0.044	(0.403)	0.409	(0.305)	0.416	(0.307)	0.269	(0.267)	7.730	(6.677)
Precision effect	0.005	(0.005)	0.000	(0.001)	0.000	(0.001)	0.001	(0.002)	0.001	(0.001)
R^2	0.000		0.011		0.011		0.008		0.003	
Publication bias in economics/business	0.222	(0.577)	-0.119	(0.912)	-0.119	(0.908)	0.512**	(0.036)	12.883	(0.245)
Publication bias in medicine/psychology	-0.271	(0.763)	0.583	(0.226)	0.593	(0.220)	0.187	(0.743)	6.371	(0.595)
Publication bias in multi-area	0.470	(0.591)	0.244	(0.705)	0.250	(0.700)	0.071	(0.901)	2.479	(0.844)
Precision effect in economics/business	0.002	(0.831)	0.001	(0.973)	0.001	(0.974)	-0.002	(0.714)	0.000	(0.870)
Precision effect in medicine/psychology	0.009	(0.445)	0.000	(0.785)	0.000	(0.775)	0.001	(0.438)	0.001	(0.682)
Precision effect in multi-area	-0.001	(0.884)	0.005*	(0.052)	0.004*	(0.064)	0.006	(0.139)	0.005**	(0.043)
R^2	0.035		0.050		0.050		0.007		0.042	

We report in parentheses wild cluster bootstrap p-values obtained from the wild cluster bootstrap-t procedure proposed by Cameron et al. (2008b), with clusters at study level (5,000 bootstraps using the Webb's (2014) six-point distribution as weights). We report wild cluster bootstrap p-values to take into account that, in each subject area, the number of clusters is small (from 16 to 36). The number of observations (studies) is 308 (85), 88 (16) in economics/business, 133 (36) in medicine/psychology, and 87 (33) in the residual multi-area category.

(a) The inverse of the square root of the sample size is used instead of $SE(r_i)$ as precision measure.

(b) PEESE is a meta-regression improved correction for publication bias.

Table A3.6: Heterogeneity in the estimated effects of retirement on health without winsorization

	Bayesian Model Averaging[a]			Weighted-Average Least Square $(q = 1)$[b]		Weighted-Average Least Square $(q = 0.5)$[b]		OLS check after BMA[c]			OLS check after WALS[d]		
	PM	PSD	PIP	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.		Std. Error	Coeff.		Std. Error
Focus Regressors													
Publication bias	**5.635**	**6.888**	**1.000**	**7.010**	**6.628**	**7.083**	**6.686**	7.790		7.566	6.501		7.609
Precision effect	**0.010**	**0.005**	**1.000**	**0.021**	**0.006**	**0.021**	**0.006**	0.009	**	0.004	0.019	***	0.007
Auxiliary Regressors													
Google scholar citations per year	0.000	0.000	0.050	**0.000**	**0.000**	**0.000**	**0.000**	–		–	0.000		0.000
Scimago Journal Ranking	0.000	0.000	0.100	-0.001	0.001	-0.001	0.001	–		–	–		–
Year of publication	**0.001**	**0.000**	**1.000**	**0.001**	**0.000**	**0.001**	**0.000**	0.001	***	0.000	0.001	***	0.000
Scimago subject areas (Reference category: Multi-area journals)													
Economics/Business	0.000	0.001	0.110	**-0.003**	**0.003**	**-0.003**	**0.003**	–		–	-0.004		0.003
Medicine/Psychology	0.000	0.001	0.140	-0.002	0.002	-0.002	0.002	–		–	–		–
Health outcomes (Reference category: Mortality)													
General and self-reported health	**0.014**	**0.003**	**1.000**	**0.011**	**0.003**	**0.011**	**0.003**	0.015	***	0.005	0.013	***	0.004
Physical health	0.000	0.001	0.100	**0.002**	**0.002**	**0.002**	**0.002**	–		–	0.001		0.002
Mental health	**0.014**	**0.003**	**1.000**	**0.010**	**0.003**	**0.010**	**0.003**	0.015	***	0.004	0.012	***	0.005
Healthcare utilization	0.000	0.001	0.080	0.001	0.001	0.001	0.001	–		–	–		–
Identification strategies (Reference category: Other methods)													
Regression discontinuity design (RDD)	0.000	0.001	0.060	**-0.009**	**0.006**	**-0.009**	**0.006**	–		–	-0.004		0.002
Instrumental variables (IV)	0.000	0.001	0.060	-0.007	0.007	-0.006	0.007	–		–	–		–
Difference-in-differences (DiD)	0.000	0.001	0.050	-0.004	0.006	-0.003	0.006	–		–	–		–
Propensity score matching (PSM)	0.000	0.003	0.050	**-0.013**	**0.013**	-0.013	0.014	–		–	–		–
Fixed-effects/First-differences	**-0.023**	**0.004**	**1.000**	**-0.025**	**0.007**	**-0.027**	**0.007**	-0.023	***	0.007	-0.021	***	0.006
Institutional contexts (Reference category: Statutory retirement)													
Mandatory or involuntary retirement	**-0.016**	**0.015**	**0.590**	**-0.021**	**0.010**	**-0.023**	**0.010**	-0.027	***	0.008	-0.025	***	0.009
Early retirement	0.000	0.001	0.100	-0.001	0.002	-0.001	0.002	–		–	–		–
Postponed retirement	-0.001	0.001	0.150	**-0.005**	**0.003**	**-0.006**	**0.003**	–		–	-0.005		0.003
Geographical areas (Reference category: Multi-country analyses)													
Europe	-0.001	0.003	0.210	**-0.005**	**0.003**	**-0.005**	**0.004**	–		–	-0.008		0.007
Extra-European countries	-0.001	0.003	0.140	**-0.006**	**0.004**	**-0.006**	**0.004**	–		–	-0.006		0.007
Sex (Reference category: Males)													
Females	0.000	0.000	0.040	-0.001	0.001	-0.001	0.001	–		–	–		–
Males+Females	0.000	0.000	0.050	**-0.004**	**0.003**	**-0.004**	**0.003**	–		–	-0.003		0.004
Calculation of t-statistic (Reference category: from 95% CI)													
t-statistic from $\widehat{\beta}_i / SE_i$	**-0.008**	**0.004**	**0.850**	**-0.006**	**0.004**	**-0.006**	**0.005**	-0.009	*	0.005	-0.009	*	0.005
Type of previous occupation (Reference category: White collars)													
Blue collars	0.000	0.001	0.040	0.001	0.003	0.001	0.003	–		–	–		–
Not specified	0.000	0.001	0.060	0.001	0.002	0.001	0.002	–		–	–		–
Birth cohorts (Reference category: Others)													
Birth cohorts $\leq$ 1950	**-0.004**	**0.002**	**0.870**	-0.001	0.002	-0.001	0.002	-0.004	***	0.001	-0.003	*	0.002

Notes: The results are from the PEESE specification by using the inverse of the SE_i^2 as weights. PM = Posterior Mean of the coefficient; PSD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability. The number of observations (studies) is 308 (85). Auxiliary variables for which the PIP is above 0.5 in BMA or the corresponding one-standard error band does not include zero in WALS are in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

[a] In the BMA, we use the uniform distribution for model priors, the Zellner's g prior for the distributions of the coefficients and a Markov Chain Monte Carlo algorithm to search over the model space, by distinguishing between focus and auxiliary regressors.

[b] $q = 1$ indicates the Laplace model prior distribution; $q = 0.5$ implies the Subbotin model prior distribution.

[c] The model specification under "OLS" includes those variables which have a PIP > 0.5 in BMA ($R^2 = 0.38$).

[d] The second model specification under "OLS" includes those variables which are relevant according to WALS ($R^2 = 0.41$).

Appendix (Ch. 4)

A. Overview of previous studies

Table A4.1 presents an overview of studies focused on the effect of retirement on mortality by distinguishing between study-related characteristics. In particular, we report the main features of each article, i.e. composition of the sample, country, selected birth cohorts, age at which mortality is evaluated, econometric strategy, and if the study focuses or not on any kind of policy intervention. The last two columns reports the effect of retirement on mortality and analysis of heterogeneity of the estimated effects.

Table A4.1: Overview of empirical evidence on the effect of retirement on mortality

Study	Sample	Country	Birth cohorts	Death age	Method	Policy change	Effects	Heterogeneity
Bloemen et al. (2017)	M civil servants	Netherlands	1940-1952	Within 5 years	FE-IV	ERA decrease	+	–
Bozio et al. (2021)	Private sector	France	1933-1943	65-74	IV	SRA increase	0	–
Brockmann et al. (2009)	M, F	Germany		Within 15 years	Other		+ (M), 0 (F)	G, I
Coe and Lindeboom (2008)	M, F	USA	1931-1941	Within 4/6 years	RDD	ERA offer	0	–
Eyjólfsdóttir et al. (2019)	M, F	Sweden	1920-1944	70-85	PSM		0	–
Fitzpatrick and Moore (2018)	M, F	USA	1921-1948	62	RDD		- (M), 0 (F)	G, E
Grøtting and Lillebø (2020)	M, F	Norway	1929-1952	Within 2014	RDD		0	–
Hagen (2018)	Civil servants	Sweden	1938-1942	65-69	DiD	NRA increase	0	–
Hallberg et al. (2015)	Military officers	Sweden	1931-1939	<=71	DiD	ERA offer	+	–
Hernaes et al. (2013)	M, F	Norway	1928-1938	67, 70, 74, 77	DiD	ERA offer	0	–
Hult et al. (2010)	M blue-collars	Sweden	1920-1932	65-72	Other		0	H
Kalwij et al. (2013)	M, F	Netherlands	1931-1945	58-65	Other		0	I
Kuhn et al. (2020)	M, F	Austria	1927-1944	<=73	IV	NRA decrease	- (M), 0 (F)	G, O
Lalive and Staubli (2015)	F	Switzerland	1938-1942	Within 2012	RDD	NRA increase	+	–
Litwin (2007)	M, F	Israel	<1937	7 years later	Other		0	E, H
Nielsen (2019)	M, F	Denmark	1939	Before 2012	RDD	NRA decrease	0	–
Rose (2020)	M, F	UK		65 (M), 60 (F)	RDD		0	–
Zulkarnain and Rutledge (2018)	M, F	Netherlands	1943-1954	62-65	IV	LR incentive	- (M), 0 (F)	G

Notes: Not all papers are clear about selected birth cohorts and death age.
Sample: M = Males; F = Females.
Methods: PSM = Propensity score matching; IV = Instrumental Variables; DiD = Difference-in-differences; RDD = Regression discontinuity design; Other = Estimation methods not controlling for endogeneity.
Policy Evaluation: ERA = Early retirement age; NRA = Normal retirement age; LR = Late retirement.
Heterogeneity: E = Education; G = Gender; H = Previous health conditions; I = Income; O = Previous occupation.

B. Institutional context

In this section we discuss in detail the pension rules in Italy before and after the 1992 pension reform. According to Brugiavini and Peracchi (2012), about two-thirds of the labor force is insured with the INPS, where the FPLD (fondo pensioni lavoratori dipendenti) is the most important in covering the private sector employees, with the exception of agricultural sector. Since 1969, Italy adopted a mandatory pay-as-you-go pension system, and the way benefit were computed changed over time, but was essentially of a defined benefit scheme, with a financial benefit equal to the average of the last 5 years of gross earnings (pension base). In 1976, pension benefits were automatically indexed to the contractual wages in the industrial sector. Retirement is not mandatory, but individuals who intend to work beyond the normal retirement age (NRA) are not protected by law and could be fired (Brugiavini, 1999). Before 1990s, the Italian pension system included two channels of eligibility to a full pension benefit: the "old age" and the "seniority" pensions. The former depended on the workers age, and the NRA was 60 years old for men and 55 years old for women with at least 15 years of contributions. The latter was based on the number of years of contributions and, in this case, 35 years of contributions for both males and females in the private sector were required to be entitled to pension, regardless of age. Otherwise, benefit calculations were more generous in the public sector, as the last earnings prior to retirement worked as the basis for the benefit and also an early retirement options were available for men and women with just 20 and 15 years of contribution, respectively. Empirical evidence suggests that the prevailing exit route was the seniority pension (Brugiavini and Peracchi, 2012; Brunello and Comi, 2015). Thus, male employees in the private sector with a continuous working and insurance career from age 15 could retire as early at age 50.

The most relevant pension reforms took place in 1990s, providing different treatments of different cohorts of workers. The 1992 pension reform (d. lgs. 503/1992, also called *Riforma Amato*) gradually increased both the NRA and the years of contributions needed for claiming the pension, so postponed retirement of both males and females workers by one every two years, starting from January 1994. The eligibility requirement for old-age pension varies across individuals on the basis of their birth date. This new system applies to workers with less than 15 years of paid contributions in December 1992, and the reference period for computing pensionable earnings was increased gradually to include the whole working life. These eligibility criteria aimed at tightening conditions to claim an old-age pension increasing it by 5 years and reaching a minimum pension age of 65 and 60 years old for men and women, respectively. Moreover, the 1992 reform also increased the years of contributions needed to access old age pension from 15 to 20 years, and the number of years of last salaries used to compute the pension base from 5 to 10, maintaining the same defined benefit formula. Otherwise, the reform left the rules governing the early retirement provision almost untouched. For this reason, individuals with many years of paid contributions were unaffected by the reform, and tended to anticipate their retirement decision opting for seniority pension claims (Brugiavini and Peracchi, 2012). In Table B4.1, such eligibility conditions are exploited for private sector employees following Brugiavini and Peracchi (2012).

Table B4.1: Eligibility conditions required for old-age pensions as established by the 1992 pension reform (male and female private sector employees)

	Old age pension		
Year	NRA (Males)	NRA (Females)	Contributions
1992	60	55	15
1993	60	55	16
1994	61	56	16
1995	61	56	17
1996	62	57	17
1997	63	58	18
1998	64	59	18
1999	65	60	19
2000	65	60	19
2001	65	60	20
2002	65	60	20

Notes: Data retrieved from Table 4.1 in Brugiavini and Peracchi (2012). The increase of the required contributions applies only to individuals with < 15 years of contributions in December 1992. Note that rules in 1992 where the same as in the previous years (ante-1992 reform). NRA = normal retirement age.

The disparities in treatment of older and younger cohorts were maintained, and the eligibility requirements were further tightened, in the subsequent 1995 and 1997 pension reforms (Law n. 335/1995 or *Riforma Dini*, and Law n. 449/1997 or *Riforma Prodi*, respectively). As suggested by Brugiavini and Peracchi (2012), the 1995 reform appears to be the most radical because it completely modifyied the eligibility rules and changed the benefit formula. However, because the changes are only introduced gradually and through a very long transitional period, the direct effects of the reform were lower if compared to the 1992 reform. In particular, the 1995 reform changed the pension formula only for those with less than 18 years of contributions in 1995 (tipically workers born between 1955 and 1965), and in practice it left the cohorts born before 1945 unaffected (Ardito et al., 2020). Indeed, the Amato reform had a major effect on retirement behavior as it was the first signal of a coherent redesigning of the social security system (Brugiavini, 1999). This is one of the main reasons why we focus only on it.

C. Descriptive statistics

In this section we provide further descriptive statistics. Figures C4.1 displays the distribution of individuals across the time elapsed since ages 50 and retirement by distinguishing between men and women. We use this measure as treatment variable in our empirical analysis in order to estimate if the retirement and its timing matter in explaining health outcomes at different years later in life. In our sample, 2,510 individuals switched into retirement at the normal retirement age before the reform (60 for men and 55 for women). Finally, Figure C4.2 shows the distribution of individuals at the age they died, Table C4.1 shows the distributions of the age at death and at retirement by distinguishing between men and women.

Table C4.1: Distributions of the age at death and at retirement

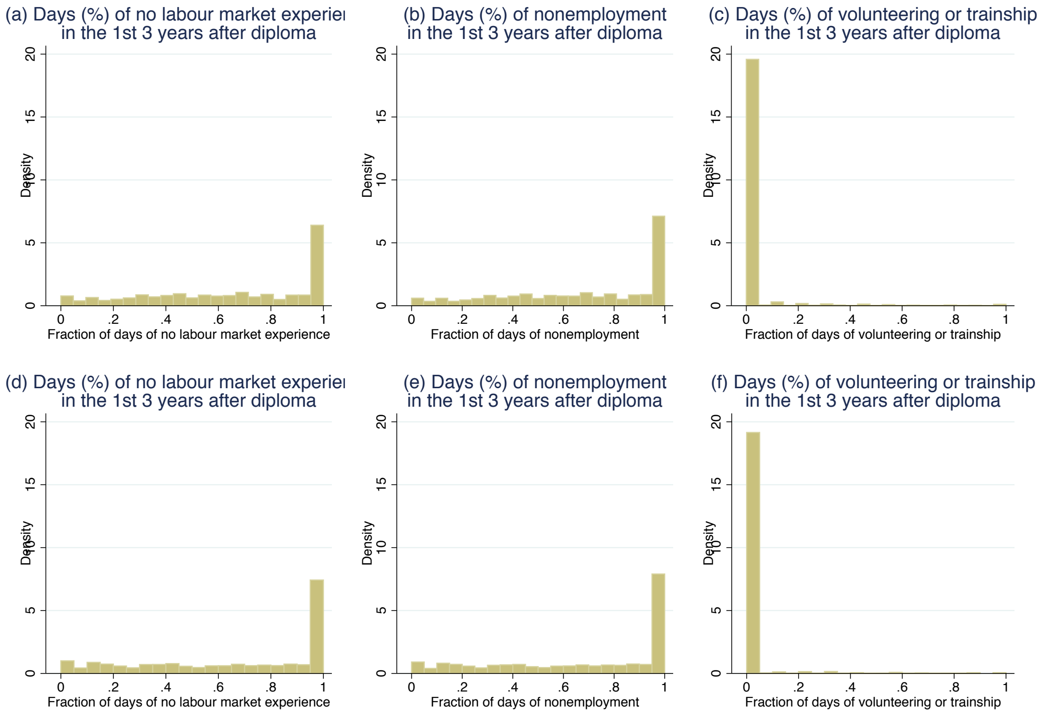
	Age at death		Age at retirement	
	Males	Females	Males	Females
Mean	74.6	75.1	59.3	57.8
Standard Deviation	4.7	4.7	4.4	3.3
10th percentile	68	68	53	55
25th percentile	71	72	56	55
50th percentile	75	76	60	58
75th percentile	78	78	63	60
90th percentile	81	81	65	61
Observations	1,396	553	7,120	5,414

First marginal correlation on the relationship between retirement (and its timing) and the health outcome considered in our analysis are provided running a series of separate linear probability models for each $t \in \{72, 75, 78\}$ separated by gender.

$$Y_{it} = \sum_{r=0}^{2} \beta_{tr} D_{ir} + x'_{it}\pi_t + \epsilon_{it} \tag{C4.1}$$

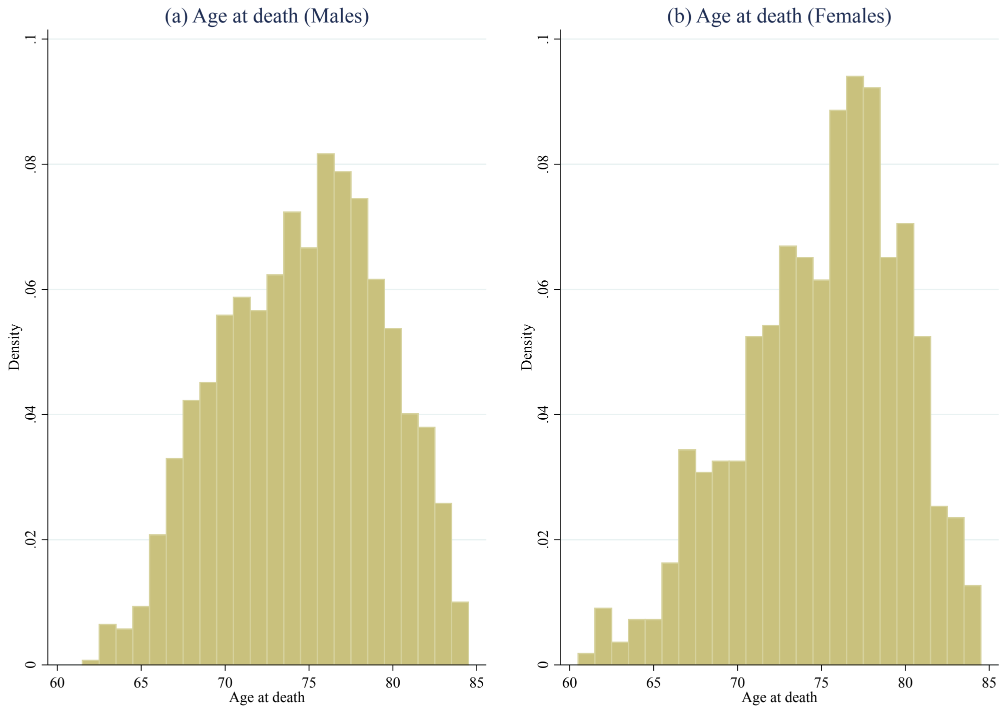
where Y_{it} is the probability of survival at ages t; x_{it} is a vector of covariates which includes the constant, educational attainment, geographical dummies, year of birth, age at the interview, average regional unemployment rate, employment rate, GDP growth rate and number of hospital beds per 1,000 inhabitants between the year of reaching 50 years old and t. D_{ir} is the treatment variable corresponding to the switch into retirement and its timing r, i.e. a dummy variable that takes value 1 if the individual is retired at time r, with $r = 0$ for early retirement, $r = 1$ for retirement at the NRA, and $r = 2$ for postponed

Figure C4.1: Spacing between ages 50 and retirement



Notes: The histograms display the distribution of individuals across the treatment variable, that is the time elapsed between the year of reaching 50 years old and the year of retirement. Graph (a) and (b) are drawn on male and female samples, respectively.

Figure C4.2: The age at death



Notes: The histograms display the distribution of employees across the age at death. Graph (a) and (b) are drawn using 1,396 men and 553 women who died in the time window under analysis.

retirement after the NRA;[8] β_{tr} is the associated coefficient which measures the impact of retirement and its timing on the probability of survival; ϵ_{it} is the error term. Tables C4.2 shows the point estimates of all the β_{tr}s for both men and women. The impact of 1 year delay of retirement on the probability of survival is fairly nil for both men and women. Furthermore, if we divide the timing of retirement in three intervals we note that the dummy for early retirement is positively correlated with the probability of survival at 72 for women (3.3 p.p.). The dummy for postponed retirement over the NRA reveals an initial positive effect of about 2 p.p., but at 75 it shows an increased risk of mortality of about 5 p.p. In contrast, postponing retirement seems to be positively correlated with the probability of survival for men in the long term (3.6 p.p.).

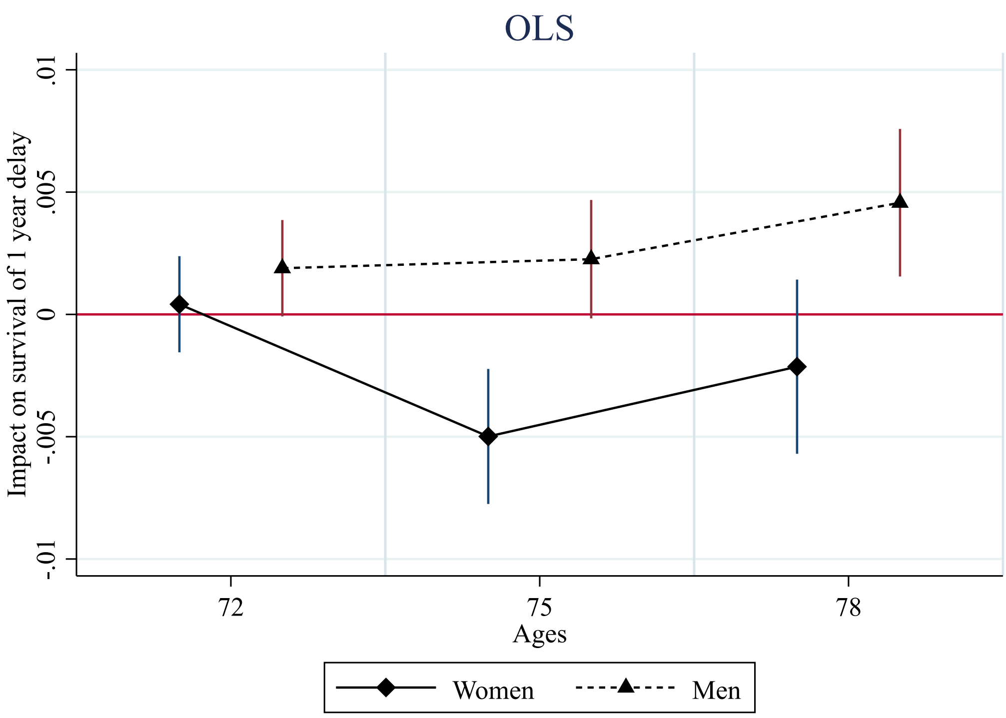
Table C4.2: LPM estimates of the timing of retirement on the probability of survival

	Males			Females		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Impact of retirement						
Impact of 1 year delay	0.002*	0.002*	0.005***	0.000	-0.005***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
b) Reference Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.016	0.016	-0.004			
	(0.013)	(0.014)	(0.017)			
Postponed retirement $\in 61, R$	-0.003	0.026*	0.036**			
	(0.013)	(0.014)	(0.016)			
c) Reference Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.033***	-0.025	0.000
				(0.012)	(0.019)	(0.026)
Postponed retirement $\in 56, R$				0.022***	-0.049***	-0.013
				(0.007)	(0.009)	(0.014)
Observations	4,652	4,288	3,658	3,394	3,176	2,594

Notes: The equations for the survival outcomes also include educational attainment, regional dummies, regional unemployment, employment and GDP growth rates, regional number of hospital beds per 1,000 inhabitants, predetermined information, year of birth, and age at the interview. Their OLS estimated parameters are not reported for the sake of brevity. *** Significant at 1%, ** significant at 5%, * significant at 10%. Standard errors robust to heteroskedasticity are reported in parentheses.

[8] We also used a second specification for the timing of retirement, i.e. the distance between age 50 and the age at retirement. The related coefficient has interpreted as the effect of 1 year delayed retirement. The estimated β_{tr} for the t and r of interest are graphically displayed in Figure C4.3 along with 95% confidence intervals, and in panel (a) of Table C4.2.

Figure C4.3: LPM estimates of 1 year delayed retirement on the probability of survival at different ages



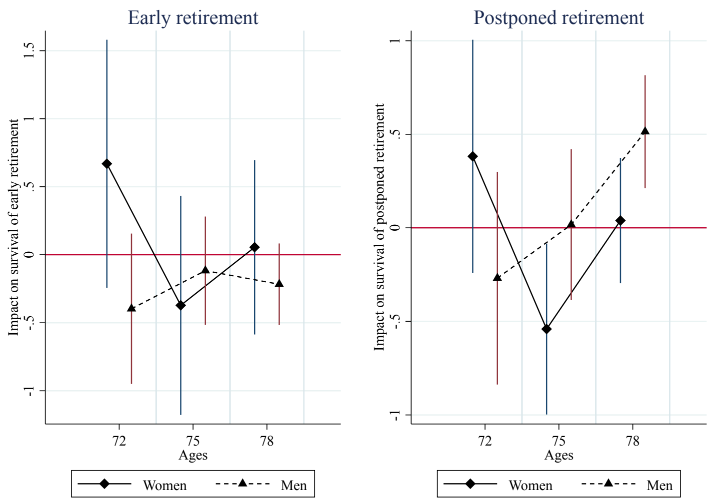
Notes: The graph is obtained by estimating linear probability models for survival at different ages and by plotting the LPM estimates of the coefficients of 1 year delayed retirement, that is panel (a) in Table C4.2. The vertical segments are 95% confidence intervals.

Table C4.3: Observed covariates and their exclusion restrictions across equations

	Selection-free measurements		Treatment equation	Outcomes
	5-year average fraction of days at work before 50	5-year average yearly labor earnings before 50	Timing of retirement	Survival at different ages
Average yearly labor earnings 5 years before reaching 50 years old	–	–	Yes	Yes
Average fraction of days at work 5 years before reaching 50 years old	–	–	Yes	Yes
Year of birth	Yes	Yes	Yes	Yes
Age at the interview	–	–	–	Yes
Education (primary, secondary, tertiary)	Yes	Yes	Yes	Yes
Geographical area of work (5 areas)	Yes	Yes	Yes	–
Geographical area at ages t (5 areas)	–	–	–	Yes
Average regional unemployment rate 5 years before reaching 50	Yes	Yes	Yes	–
Average regional employment rate 5 years before reaching 50	Yes	Yes	Yes	–
Average regional GDP growth rate 5 years before reaching 50	Yes	Yes	Yes	–
Average regional hospital beds x1,000 inhabitants 5 years before reaching 50	Yes	Yes	Yes	–
Average regional unemployment rate between 50 and t	–	–	–	Yes
Average regional employment rate between 50 and t	–	–	–	Yes
Average regional GDP growth rate between 50 and t	–	–	–	Yes
Average egional hospital beds x1,000 inhabitants between 50 and t	–	–	–	Yes
1992 pension reform	–	–	Yes	–
Timing of retirement	–	–	–	Yes

D. Full set of estimation results without unobserved heterogeneity

Figure D4.1: Impact of the timing of retirement on the probability of survival at different ages without unobserved heterogeneity



Notes: The graph is obtained by estimating the model for survival at different ages assuming no unobserved heterogeneity and by plotting the logit estimates of the coefficients of the timing of retirement displayed in Table 4.5. They have to be read in deviation from the reference category (NRA). The vertical segments are 95% confidence intervals.

Table D4.1: Estimated (logit) coefficients of the covariates of the health outcome equations without unobserved heterogeneity

	Probability of survival (Males)			Probability of survival (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.041		0.066	0.182		0.119
Year of birth/10 (normalized to its minimum)	-7.924	***	0.276	-8.537	***	0.422
Age at the interview/10	1.456	***	0.153	1.629	***	0.246
Geographical area at work - Reference category: North-West						
North-East	0.925	***	0.085	0.843	***	0.135
Center	1.026	***	0.081	0.999	***	0.132
South	1.038	***	0.147	1.158	***	0.269
Islands	1.137	***	0.202	1.451	***	0.367
Average regional unemployment rate between 50 and t	-0.431	***	0.388	-0.391	***	0.059
Average regional employment rate between 50 and t	-0.194	***	0.220	-0.147	***	0.035
Average regional growth rate between 50 and t	-2.358	***	0.104	-2.386	***	0.157
Average regional number of beds x 1,000 inhabitants between 50 and t	-0.092		0.056	0.059		0.099
Average yearly labour earnings 5 years before reaching 50/10,000	0.058	***	0.021	-0.153	**	0.060
Average yearly fraction of days at work 5 years before reaching 50	-0.026		0.104	0.206		0.174
Constant at $t = 72$	-0.092		0.056	0.059		0.099
Constant at $t = 75$	0.058	***	0.021	-0.153	**	0.060
Constant at $t = 78$	-0.026		0.104	0.206		0.174

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table D4.2: Estimated coefficients of the measurement equations without unobserved heterogeneity (Males)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.055	***	0.010	-5216.46	***	239.43
Year of birth/10 (normalized to its minimum)	0.024	*	0.020	3042.45	***	535.45
Geographical area at work - Reference category: North-West						
North-East	-0.040	***	0.015	-2426.14	***	331.33
Center	-0.053	***	0.014	-3121.10	***	604.38
South	-0.182	***	0.021	-8978.82	***	815.82
Islands	-0.190	***	0.027	-8825.32	***	815.82
Average regional unemployment rate before 50	-0.002		0.003	-181.59	**	88.99
Average regional employment rate before 50	0.004	**	0.002	29.19		39.82
Average regional growth rate before 50	0.027		0.375	19393.04	*	10410.09
Average regional number of beds x 1,000 inhabitants before 50	0.002		0.004	-48.75		95.71
Constant	0.379	***	0.110	21850.66	***	2956.88
ln(σ^2)	–		–	-0.767	***	0.006

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table D4.3: Estimated coefficients of the measurement equations without unobserved heterogeneity (Females)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.140		0.099	1996.43	***	621.25
Year of birth/10 (normalized to its minimum)	-0.008		0.114	2066.47	***	733.65
Geographical area at work - Reference category: North-West						
North-East	0.002		0.076	-167.59		461.68
Center	0.041		0.072	193.13		440.97
South	-0.079		0.119	-1198.91		766.92
Islands	-0.031		0.160	-532.58		1069.61
Average regional unemployment rate before 50	-0.006		0.019	-149.12		124.33
Average regional employment rate before 50	0.005		0.009	117.12	**	54.11
Average regional growth rate before 50	-0.043		0.023	-1607.02		141.09
Average regional number of beds x 1,000 inhabitants before 50	-0.009		0.019	-49.51		116.12
Constant	0.176		0.625	526.23		3969.58
ln(σ^2)	–		–	-0.980	***	0.010

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

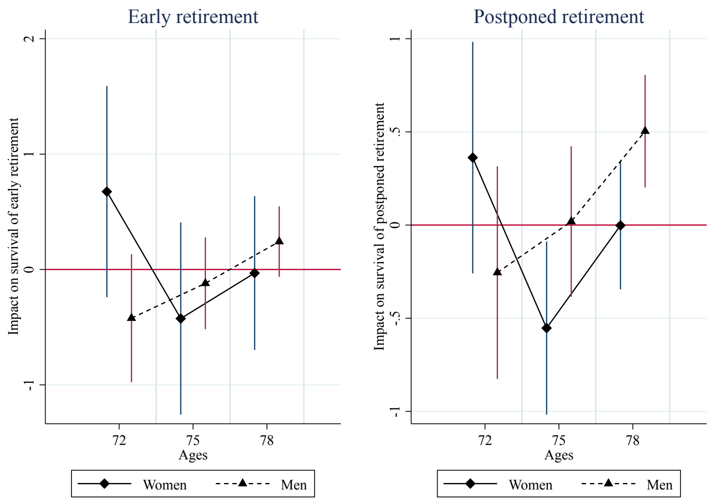
Table D4.4: Estimated coefficients of the (ordered logit) equation for the timing of retirement without unobserved heterogeneity

	Selection into retirement (Males)			Selection into retirement (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.652	***	0.065	-0.232	***	0.092
Year of birth/10 (normalized to its minimum)	-1.037	***	0.127	-0.004		0.159
Geographical area at work - Reference category: North-West						
North-East	0.142	*	0.082	0.264	***	0.095
Center	0.394	***	0.076	0.561	***	0.091
South	1.075	***	0.121	0.952	***	0.152
Islands	0.704	***	0.164	1.617	***	0.228
Average regional unemployment rate before 50	0.015		0.018	-0.070	***	0.023
Average regional employment rate before 50	-0.034	***	0.009	-0.010		0.011
Average regional growth rate before 50	-0.370	**	0.247	0.848	***	0.029
Average regional number of beds x 1,000 inhabitants before 50	0.024		0.022	-0.014		0.027
Average yearly labour earnings 5 years before reaching 50/10,000	-0.157	***	0.018	-0.052		0.046
Average yearly fraction of days at work 5 years before reaching 50	0.718	***	0.107	1.069	***	0.137
1992 pension reform	2.904	***	0.138	1.794	***	0.116
Ordered logit thresholds of the timing of retirement						
δ_0 (retired at $r = 0$)	-2.468	***	0.657	-2.780	***	0.844
ln($\delta_1 - \delta_0$) [retired at $r = 1$]	-0.356	***	0.031	0.692	***	0.028

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

E. Full set of estimation results with time-constant unobserved heterogeneity

Figure E4.1: Impact of the timing of retirement on the probability of survival at different ages with time-constant unobserved heterogeneity



Notes: The graph is obtained by estimating the model for survival at different ages assuming time-constant unobserved heterogeneity and by plotting the logit estimates of the coefficients of the timing of retirement displayed in Table 4.5. They have to be read in deviation from the reference category (NRA). The vertical segments are 95% confidence intervals.

Table E4.1: Estimated (logit) coefficients of the covariates of the health outcome equations with time-constant unobserved heterogeneity

	Probability of survival (Males)			Probability of survival (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.039		0.066	0.157		0.120
Year of birth/10 (normalized to its minimum)	-7.937	***	0.279	-8.570	***	0.429
Age at the interview/10	1.457	***	0.153	1.623	***	0.247
Geographical area at work - Reference category: North-West						
North-East	0.927	***	0.087	0.852	***	0.136
Center	1.023	***	0.083	1.021	***	0.133
South	1.030	***	0.150	1.148	***	0.269
Islands	1.131	***	0.205	1.455	***	0.367
Average regional unemployment rate between 50 and t	-0.434	***	0.039	-0.392	***	0.059
Average regional employment rate between 50 and t	-0.195	***	0.022	-0.148	***	0.035
Average regional growth rate between 50 and t	-2.358	***	0.105	-2.397	***	0.157
Average regional number of beds x 1,000 inhabitants between 50 and t	-0.097		0.056	0.055	***	0.100
Average yearly labour earnings 5 years before reaching 50/10,000	0.051	**	0.039	-0.356	***	0.113
Average yearly fraction of days at work 5 years before reaching 50	-0.042		0.117	-0.077		0.203
Constant at $t = 72$	18.885	***	2.183	14.854	***	3.468
Constant at $t = 75$	16.137	***	2.167	12.738	***	3.457
Constant at $t = 78$	13.150	***	2.136	9.425	***	3.440

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table E4.2: Estimated coefficients of the measurement equations with time-constant unobserved heterogeneity (Males)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.123	***	0.021	-2633.20	***	150.24
Year of birth/10 (normalized to its minimum)	0.001		0.040	-122.99		291.76
Geographical area at work - Reference category: North-West						
North-East	-0.040		0.029	-2149.52	***	190.92
Center	-0.058	**	0.027	-3407.07	***	179.42
South	-0.139	***	0.040	-5543.39	***	296.97
Islands	-0.160	***	0.052	-5921.50	***	422.88
Average regional unemployment rate before 50	-0.005		0.006	-302.01	***	43.60
Average regional employment rate before 50	0.002		0.003	-96.29	***	21.66
Average regional growth rate before 50	0.012		0.078	-74.79		581.38
Average regional number of beds x 1,000 inhabitants before 50	0.002		0.007	-324.87	***	53.94
Constant	0.540	***	0.219	35657.09	***	1610.61
ln(σ^2)	–		–	-1.355	***	0.006

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table E4.3: Estimated coefficients of the measurement equations with time-constant unobserved heterogeneity (Females)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.087	***	0.016	-324.55	**	136.82
Year of birth/10 (normalized to its minimum)	-0.023		0.022	-1339.68	***	215.81
Geographical area at work - Reference category: North-West						
North-East	0.013		0.014	126.08		130.27
Center	0.051	***	0.013	252.26	**	127.61
South	-0.026		0.021	-152.62		206.84
Islands	0.018		0.028	177.19		292.12
Average regional unemployment rate before 50	0.002		0.003	39.59		33.68
Average regional employment rate before 50	0.003	**	0.002	39.56	***	15.25
Average regional growth rate before 50	-0.017		0.041	731.40	*	38.87
Average regional number of beds x 1,000 inhabitants before 50	-0.006		0.004	18.94		33.91
Constant	-0.080		0.111	-1136.68		1109.58
ln(σ^2)	–		–	-1.914	***	0.008

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table E4.4: Estimated coefficients of the (ordered logit) equation for the timing of retirement with time-constant unobserved heterogeneity

	Selection into retirement (Males)			Selection into retirement (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.624	***	0.066	-0.225	**	0.092
Year of birth/10 (normalized to its minimum)	-1.020	***	0.127	0.049		0.161
Geographical area at work - Reference category: North-West						
North-East	0.197	**	0.083	0.257	***	0.094
Center	0.478	***	0.078	0.548	***	0.091
South	1.202	***	0.123	0.962	***	0.150
Islands	0.859	***	0.166	1.615	***	0.228
Average regional unemployment rate before 50	-0.005		0.018	-0.071	***	0.023
Average regional employment rate before 50	-0.031	***	0.009	-0.010		0.011
Average regional growth rate before 50	-0.350		0.245	0.832	***	0.292
Average regional number of beds x 1,000 inhabitants before 50	0.035		0.022	-0.013		0.027
Average yearly labour earnings 5 years before reaching 50/10,000	0.047		0.032	0.126		0.086
Average yearly fraction of days at work 5 years before reaching 50	0.981	***	0.119	1.245	***	0.158
1992 pension reform	2.896	***	0.139	1.783	***	0.116
Ordered logit thresholds of the timing of retirement						
δ_0 (retired at $r = 0$)	-1.368	**	0.672	-2.755	***	0.844
ln($\delta_1 - \delta_0$) [retired at $r = 1$]	-0.347	***	0.031	0.693	***	0.028

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table E4.5: Estimated distribution of the discrete time-constant unobserved heterogeneity with $H = 3$ support points

	Location of the mass		Logistic weight of the probability masses (p^h)			Resulting probabilities (p^h)
	Coeff.	Std. Error	Coeff.		Std. Error	
a) Men						
θ^1	0.000	–	1.636	***	0.068	0.556
θ^2	-0.141	0.136	1.132	***	0.070	0.336
θ^3	0.183	0.176	–		–	0.108
b) Women						
θ^1	0.000	–	1.178	***	0.070	0.529
θ^2	0.315	0.251	0.635	***	0.084	0.308
θ^3	0.573	0.456	–		–	0.163

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. The normalisation $\theta^1 = 0$ is innocuous: all the constant terms displayed in the last part of Table E4.1.

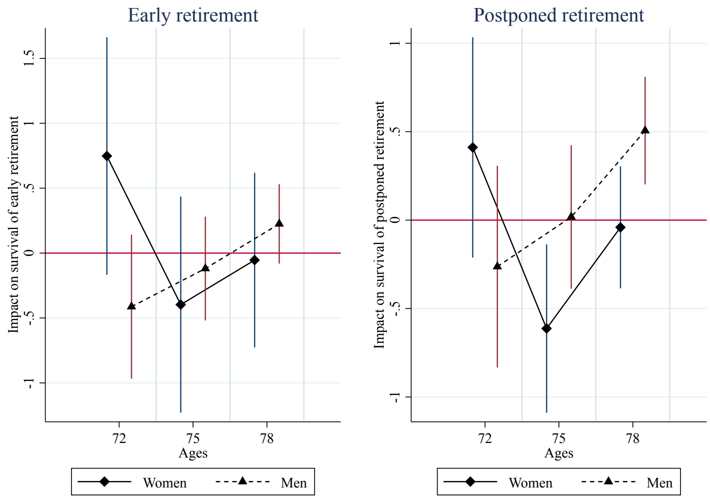
Table E4.6: Estimated loading factors with time-constant unobserved heterogeneity (discrete distribution with $H = 3$ support points)

	Men		Women		
Equations	Loading factor	Std. Error	Loading factor		Std. Error
a) Measurement equations					
Average fraction of days at work during 5 years before 50	11.155	10.686	3.856		3.068
Average yearly labor earnings during 5 years before 51	1.788	1.714	1.350		1.073
b) Selection into treatment equation					
Timing of retirement	4.209	4.044	1.110		1.012
c) Health outcomes					
Probability of survival at $t = 72$	1.000	–	1.000		–
Probability of survival at $t = 75$	0.201	0.946	1.500		1.129
Probability of survival at $t = 78$	-0.695	1.307	1.686	*	0.994

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. The loading factors of the health outcomes are normalized to 1.

F. Full set of estimation results with time-varying unobserved heterogeneity

Figure F4.1: Impact of the timing of retirement on the probability of survival at different ages with time-varying unobserved heterogeneity



Notes: The graph is obtained by estimating the model for survival at different ages assuming time-varying unobserved heterogeneity and by plotting the logit estimates of the coefficients of the timing of retirement displayed in Table 4.5. They have to be read in deviation from the reference category (NRA). The vertical segments are 95% confidence intervals.

Table F4.1: Estimated (logit) coefficients of the covariates of the health outcome equations with time-varying unobserved heterogeneity

	Probability of survival (Males)			Probability of survival (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.052		0.067	0.152		0.120
Year of birth/10 (normalized to its minimum)	-7.955	***	0.281	-8.615	***	0.434
Age at the interview/10	1.453	***	0.153	1.649	***	0.247
Geographical area at work - Reference category: North-West						
North-East	0.929	***	0.087	0.846	***	0.136
Center	1.028	***	0.083	1.001	***	0.133
South	1.045	***	0.150	1.133	***	0.271
Islands	1.150	***	0.205	1.428	***	0.369
Average regional unemployment rate between 50 and t	-0.433	***	0.040	-0.393	***	0.059
Average regional employment rate between 50 and t	-0.195	***	0.022	-0.150	***	0.035
Average regional growth rate between 50 and t	-2.362	***	0.105	-2.402	***	0.159
Average regional number of beds x 1,000 inhabitants between 50 and t	-0.094		0.057	0.067	***	0.100
Average yearly labour earnings 5 years before reaching 50/10,000	0.080	**	0.039	-0.175	***	0.121
Average yearly fraction of days at work 5 years before reaching 50	-0.095		0.134	-0.248		0.221
Constant at $t = 72$	18.970	***	2.208	14.936	***	3.541
Constant at $t = 75$	16.054	***	2.184	12.978	***	3.501
Constant at $t = 78$	12.764	***	2.149	9.745	***	3.493

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table F4.2: Estimated coefficients of the measurement equations with time-varying unobserved heterogeneity (Males)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.123	***	0.021	-2653.59	***	150.46
Year of birth/10 (normalized to its minimum)	0.001		0.040	-70.19		291.86
Geographical area at work - Reference category: North-West						
North-East	-0.040		0.029	-2146.56	***	191.16
Center	-0.058	**	0.027	-3412.97	***	179.56
South	-0.139	***	0.040	-5539.36	***	296.90
Islands	-0.160	***	0.052	-5935.17	***	422.70
Average regional unemployment rate before 50	-0.005		0.006	-308.16	***	43.58
Average regional employment rate before 50	0.002		0.003	-100.42	***	21.66
Average regional growth rate before 50	0.013		0.078	-169.20		581.69
Average regional number of beds x 1,000 inhabitants before 50	0.002		0.007	-325.01	***	53.96
Constant	0.872	***	0.222	56450.84	***	1611.26
ln(σ^2)	–		–	-1.355	***	0.006

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table F4.3: Estimated coefficients of the measurement equations with time-varying unobserved heterogeneity (Females)

	Average fraction of days at work during 5 years before 50			Average yearly labor earnings during 5 years before 50		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	0.087	***	0.016	-323.24	**	136.79
Year of birth/10 (normalized to its minimum)	-0.024		0.022	-1371.31	***	216.46
Geographical area at work - Reference category: North-West						
North-East	0.014		0.014	131.92		130.18
Center	0.052	***	0.013	257.72	**	127.59
South	-0.026		0.021	-150.87		206.82
Islands	0.018		0.028	184.51		291.86
Average regional unemployment rate before 50	0.002		0.003	39.84		33.67
Average regional employment rate before 50	0.003	**	0.002	40.40	***	15.27
Average regional growth rate before 50	-0.016		0.040	769.84	**	38.89
Average regional number of beds x 1,000 inhabitants before 50	-0.006	*	0.004	16.30		33.90
Constant	0.694	***	0.112	20961.55	***	1101.95
ln(σ^2)	–		–	-1.913	***	0.008

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.
We estimated the model using labor earnings divided by 10,000 to reduce numerical problems. Then, we multiplied all the estimated coefficients by 10,000 before reporting results, apart from the natural logarithms of the variances of the underlying normal distributions. Hence, the latter must be interpreted as the log of the variance of the normal distribution of labor earnings divided by 10,000, i.e. $\ln(\sigma_t^2 \cdot 10,000)$.

Table F4.4: Estimated coefficients of the (ordered logit) equation for the timing of retirement with time-varying unobserved heterogeneity

	Selection into retirement (Males)			Selection into retirement (Females)		
	Coeff.		Std. Error	Coeff.		Std. Error
Education level - Reference: Secondary or tertiary						
Lower secondary or less	-0.624	***	0.066	-0.225	**	0.092
Year of birth/10 (normalized to its minimum)	-1.021	***	0.127	0.517		0.161
Geographical area at work - Reference category: North-West						
North-East	0.196	**	0.083	0.257	***	0.094
Center	0.477	***	0.078	0.548	***	0.091
South	1.199	***	0.123	0.962	***	0.150
Islands	0.857	***	0.166	1.615	***	0.228
Average regional unemployment rate before 50	-0.005		0.018	-0.071	***	0.023
Average regional employment rate before 50	-0.031	***	0.009	-0.010		0.011
Average regional growth rate before 50	-0.350		0.245	0.832	***	0.292
Average regional number of beds x 1,000 inhabitants before 50	0.035		0.022	-0.013		0.027
Average yearly labour earnings 5 years before reaching 50/10,000	0.046		0.032	0.132		0.086
Average yearly fraction of days at work 5 years before reaching 50	0.977	***	0.120	1.250	***	0.159
1992 pension reform	2.895	***	0.139	1.783	***	0.116
Ordered logit thresholds of the timing of retirement						
δ_0 (retired at $r = 0$)	-0.611		0.704	-2.098	**	0.884
ln(δ_1-δ_0) [retired at $r = 1$]	-0.347	***	0.031	0.693	***	0.028

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table F4.5: Estimated distribution of the discrete time-varying unobserved heterogeneity with $H = 3$ support points

	Location of the mass			Logistic weight of the probability masses (p^h)			Resulting probabilities (p^h)
	$t = 72$	$t = 75$	$t = 78$	Coeff.		Std. Error	
a) Men							
θ^1	0.000	0.000	0.000	-1.642	***	0.068	0.068
θ^2	-0.156	0.077	0.464*	-0.505	***	0.037	0.581
	(0.311)	(0.319)	(0.275)				
θ^3	-0.088	0.108	0.465**	–		–	0.351
	(0.175)	(0.260)	(0.226)				
b) Women							
θ^1	0.000	0.000	0.000	-0.635	***	0.084	0.162
θ^2	-0.112	-0.324	-0.444	0.544	***	0.045	0.530
	(0.455)	(0.460)	(0.423)				
θ^3	-0.050	0.288	0.268	–		–	0.308
	(0.205)	(0.387)	(0.340)				

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. The normalisation $\theta^1 = 0$ is innocuous: all the constant terms displayed in the last part of Table F4.1.

Table F4.6: Estimated loading factors with time-varying unobserved heterogeneity (discrete distribution with $H = 3$ support points)

	Men		Women	
Equations	Loading factor	Std. Error	Loading factor	Std. Error
a) Measurement equations				
Average fraction of days at work during 5 years before 50	23.254	46.284	19.675	79.767
Average yearly labor earnings during 5 years before 50	3.725	7.411	6.886	27.909
b) Selection into treatment equation				
Timing of retirement	8.693	17.288	5.850	23.897
c) Health outcomes				
Probability of survival at $t = 72$	1.000	–	1.000	–
Probability of survival at $t = 75$	1.000	–	1.000	–
Probability of survival at $t = 78$	1.000	–	1.000	–

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%. The loading factors of the health outcomes are normalized to 1.

G. Sensitivity analysis

Table G4.1: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
Spacing between 50 and year of retirement	0.023 (0.014)	0.023 (0.015)	0.036*** (0.014)	-0.007 (0.034)	-0.047 (0.031)	-0.008 (0.024)
Observations	4,652	4,288	3,658	3,394	3,176	2,594

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.2: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity using a different set of exclusion restrictions

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.469 (0.287)	-0.104 (0.206)	0.323** (0.156)			
Postponed retirement $\in 61, R$	-0.241 (0.294)	0.021 (0.207)	0.549*** (0.156)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.778 (0.469)	-0.372 (0.423)	-0.017 (0.343)
Postponed retirement $\in 56, R$				0.405 (0.322)	-0.625*** (0.242)	-0.068 (0.176)
Observations	4,652	4,288	3,658	3,394	3,176	2,594

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.3: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (blue collar workers)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.226	0.037	0.195			
	(0.333)	(0.239)	(0.189)			
Postponed retirement $\in 61, R$	-0.085	0.189	0.488***			
	(0.346)	(0.237)	(0.184)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.563	-0.945*	-0.620
				(0.664)	(0.520)	(0.412)
Postponed retirement $\in 56, R$				0.361	-0.943***	-0.151
				(0.410)	(0.308)	(0.222)
Observations	3,215	2,942	2,508	2,260	1,954	1,450

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.4: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (white collar workers)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.680	-0.793	-0.121			
	(0.819)	(0.576)	(0.426)			
Postponed retirement $\in 61, R$	-0.557	-0.419	0.334			
	(0.825)	(0.596)	(0.426)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				1.199	0.761	1.840*
				(1.087)	(0.998)	(0.995)
Postponed retirement $\in 56, R$				0.759	0.115	1.434
				(0.840)	(0.916)	(0.975)
Observations	978	858	705	394	294	212

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.5: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (low educated workers)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.364	-0.168	0.231			
	(0.336)	(0.242)	(0.169)			
Postponed retirement $\in 61, R$	-0.294	-0.155	0.445***			
	(0.346)	(0.241)	(0.155)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.969*	-0.383	-0.372
				(0.573)	(0.477)	(0.372)
Postponed retirement $\in 56, R$				0.439	-0.674***	-0.195
				(0.361)	(0.259)	(0.191)
Observations	3,600	3,468	3,024	2,907	2,812	2,320

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.6: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (high educated workers)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.493	-0.119	0.042			
	(0.544)	(0.419)	(0.456)			
Postponed retirement $\in 61, R$	-0.288	0.605	0.775*			
	(0.551)	(0.438)	(0.428)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				-0.111	-0.421	1.267
				(1.020)	(1.485)	(1.149)
Postponed retirement $\in 56, R$				0.241	-0.401	0.611
				(0.827)	(0.865)	(0.609)
Observations	1,052	820	634	487	364	274

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.7: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (married at the interview)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	-0.501	-0.080	0.206			
	(0.322)	(0.228)	(0.173)			
Postponed retirement $\in 61, R$	-0.325	0.065	0.455***			
	(0.330)	(0.231)	(0.174)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.817	-0.807	-0.072
				(0.615)	(0.661)	(0.727)
Postponed retirement $\in 56, R$				0.581	-0.893**	-0.430
				(0.446)	(0.357)	(0.283)
Observations	3,879	3,542	2,969	2,210	1,878	1,293

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table G4.8: Estimated (logit) coefficients of the timing of retirement on the probability of survival with time-varying unobserved heterogeneity (single/divorced/widowed at the interview)

	Probability of survival (Males)			Probability of survival (Females)		
	$t = 72$	$t = 75$	$t = 78$	$t = 72$	$t = 75$	$t = 78$
a) Ref. Category: Retirement at NRA (60)						
Early retirement $\in 50, 59$	0.031	-0.263	0.339			
	(0.638)	(0.500)	(0.401)			
Postponed retirement $\in 61, R$	0.144	-0.131	0.818**			
	(0.669)	(0.506)	(0.369)			
b) Ref. Category: Retirement at NRA (55)						
Early retirement $\in 50, 54$				0.929	0.191	0.021
				(0.810)	(0.663)	(0.403)
Postponed retirement $\in 56, R$				0.333	-0.252	0.281
				(0.511)	(0.354)	(0.237)
Observations	773	746	689	1,184	1,298	1,301

Notes: * Significant at 10%, ** significant at 5%, *** significant at 1%.