Ca' Foscari University of Venice

Department of Economics

Working Paper

Petru Crudu

Long-term effects of early adverse labour market conditions: A Causal Machine Learning approach

ISSN: 1827-3580
No. 21/WP/2023

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 21/WP/2023
ISSN 1827-3580

Long-term effects of early adverse labour market conditions: A Causal Machine Learning approach

Petru Crudu
Ca' Foscari University of Venice

Abstract
This study estimates the long-term causal effects of completing education during adverse labour market conditions, measuring outcomes 35 years post-education. To achieve this, the study combines historical regional unemployment rates with detailed SHARE microdata for European cohorts completing education between 1960 and 1990 in a novel database. A systematic heterogeneity analysis is conducted by leveraging the Causal Forest, a causal machine learning estimator that allows estimates at various aggregation levels. Furthermore, the causal link is validated using an instrumental variable approach. The main findings reveal that a one-percentage-point increase in the unemployment rate at the time of completing education leads to a significant decline in earnings (-5.2%) and self-perceived health (-2.23%) after 35 years. The heterogeneity analysis uncovers that the results are primarily driven by less educated individuals and highlights a permanent disadvantage for women in labour market participation. This study also provides evidence that systematic divergence in life trajectories can be explained by search theory and human capital models. Overall, the research suggests that the consequences of limited post-education opportunities can be permanent, underscoring the importance of identifying vulnerable groups for effective policy interventions.

Keywords
Long-term Effects, Unemployment, Heterogeneous Effects, GRF

JEL Codes
J31, I1, J24, I24, E24

Address for correspondence:
Petru Crudu
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: petru.crudu@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/pag/16882/)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Long-term effects of early adverse labour market conditions: A Causal Machine Learning approach

Petru Crudu*

Abstract

This study estimates the long-term causal effects of completing education during adverse labour market conditions, measuring outcomes 35 years post-education. To achieve this, the study combines historical regional unemployment rates with detailed SHARE microdata for European cohorts completing education between 1960 and 1990 in a novel database. A systematic heterogeneity analysis is conducted by leveraging the Causal Forest, a causal machine learning estimator that allows estimates at various aggregation levels. Furthermore, the causal link is validated using an instrumental variable approach. The main findings reveal that a one-percentage-point increase in the unemployment rate at the time of completing education leads to a significant decline in earnings (-5.2%) and self-perceived health (-2.23%) after 35 years. The heterogeneity analysis uncovers that the results are primarily driven by less educated individuals and highlights a permanent disadvantage for women in labour market participation. This study also provides evidence that systematic divergence in life trajectories can be explained by search theory and human capital models. Overall, the research suggests that the consequences of limited post-education opportunities can be permanent, underscoring the importance of identifying vulnerable groups for effective policy interventions.

Keywords: Long-term Effects, Unemployment, Heterogeneous Effects, GRF
JEL Codes: J31, I1, J24, I24, E24

*Ca' Foscari University of Venice. Email: petru.crudu@unive.it

1 Introduction

Economic crises are a cause for concern for young people because the negative impact of poor labour market conditions does not manifest only in the short term but may influence also adulthood outcomes (Kahn (2010); Oreopoulos et al. (2012); Maclean (2013); Schwandt and Von Wachter (2019)). The first years spent in the labour market are crucial for determining potential career paths. During these early ages, workers acquire general and/or industry-specific knowledge and know-how, switch jobs in search of better opportunities, and eventually stick to a given job. As a consequence, even temporary recessions might affect long-term outcomes by shrinking the available opportunities to secure the optimal job position.

It is unclear, however, whether these effects persist and weaken over time or if they can become permanent, shaping outcomes well into later stages of life. Moreover, little is known about the long-term effects on European cohorts who experienced the stagflation periods of the 1970s. Yet, there is very little dedicated research aimed at identifying heterogeneity and elucidating the mechanisms that underlie the long-term effects throughout the entirety of the life cycle.

This paper exploits a unique dataset of historical unemployment rates at the regional level gathered from different administrative sources to provide new empirical evidence on the effect of adverse labour market conditions at the time of completing education on old-age outcomes. The microdata is retrieved from SHARE (Börsch-Supan et al. (2013); Brugiavini et al. (2019)) and consists of individuals completing education between 1960 and 1990 in Austria, Belgium, Denmark, West Germany, and Italy. To the best of my knowledge, this is the first paper investigating the causal effects after more than 35 years of completing education. Moreover, I offer a comprehensive assessment by combining administrative data, survey data, standard identification and estimation strategies combined with cutting-edge Causal Machine Learning methods. This enables me to document the sizeable, long-lasting impacts on individuals who encountered unfavourable initial labour market conditions. Beyond the causal effect on labour market outcomes, my analysis provides a better understanding of (i) additional life outcomes linked to the labour market, encompassing health and family dynamics, (ii) elucidation of underlying mechanisms considering the entirety of lifetime, and (iii) exploration of the potential usefulness of cutting-edge Causal Machine Learning methods in unravelling and clarifying heterogeneity.

In detail, I use the Causal Forests to estimate long-term effects by relying on the Generalized Random Forest (GRF) (Athey et al. (2019)). Causal Machine Learning is a cutting-edge and fast-growing literature that combines the microeconometric literature on identifying causal effects with the prediction power of the statistical learning literature (Hastie et al. (2009)). The most significant advantages stem from their explicit design to effectively capture heterogeneity in a highly flexible manner. The Causal Forest estimates the treatment effects at the individual level, allowing a more detailed analysis of heterogeneity. The granularity in treatment effects allows me to recognise existing theories to explain heterogeneity, while also opening pathways for the discovery of new explanations.

The main empirical challenge lies in the potential linkage between the characteristics of graduation[1] cohorts and prevailing labour market conditions. Local unemployment rates may influence migration or graduation timing, which may depend on an individual's socioeconomic background or ability. Specifically, my analysis addresses potential self-selection issues following two strategies. First, I rely on a broad set of childhood conditions including socioeconomic status, ability, and health to estimate the long-term effects with the Causal Forest estimator (Athey et al. (2019)). Second, I use a standard instrumental variable approach to validate the previous results. Particularly, I instrument the graduation unemployment rate using the exogenous unemployment rate determined by birth year and the compulsory minimum school-leaving age laws. In a comprehensive assessment, the results do not appear to be substantially driven by endogenous mobility across regions or graduation decisions. To begin with, the instrumental variable approach yields highly comparable results with the first baseline approach. Furthermore, I provide supporting evidence regarding the capacity of childhood circumstances to predict and control for potential omitted variables.

I find that adverse labour market conditions at the time of concluding education permanently downgrade the long-term labour market and health outcomes. Overall, after more than 35 years, individuals have 5.2% fewer earnings from work and significantly worse self-perceived health (-2.23%) and less grip strength (-1%) for a percentage point increase in the unemployment rate at the time of completing education. The impact on earnings is primarily attributable to the hourly remuneration rather than the number of hours worked. This suggests that individuals in question experience suboptimal job placement, and labour market rigidities fail to facilitate adjustments at the intensive margin. Furthermore, the magnitude of these effects is economically significant, revealing the potential pervasiveness of the effects of early conditions during older ages. Results remain robust to a vast set of robustness checks.

Importantly, the average effects hide strong heterogeneity. A clear education gradient is present for all the outcomes under analysis, i.e., university-educated individuals are able to hedge from adverse initial conditions. For instance, the effect on hourly earnings

[1]For the sake of time I will improperly refer to graduation as the time of ending full-time education, which may consist of having no diploma, a high-school diploma or a university degree.

is -6.3% for individuals with less than a high school diploma, -4% for those with a high school diploma, and -1.6% (but not statistically significant) for university degree holders. The second form of heterogeneity identified concerns how individuals are impacted in their labour market outcomes based on gender. Men are more affected in earnings (-5%), whereas women are more affected in labour market participation later in life (-2.8%). This outcome is far from trivial, and the accessibility of lifetime information enables me to delineate a clear pathway for understanding it.

The explaining mechanism relies on cumulative disadvantage (DiPrete and Eirich (2006)) spawned from poor labour market opportunities at the time of completing education, which results in systematic divergence in life course trajectories. The empirical evidence in my dataset shows that initial unfavourable conditions reduce the opportunity to find jobs and acquire human capital. Besides, the mechanisms differ significantly based on gender. In instances of unfavourable labour market conditions at the time of completing their education, women experience an immediate labour market exclusion, with a -4.1% likelihood of securing employment in the first year. Furthermore, they attain significantly fewer job opportunities (-0.053). This effect exhibits a strong path dependency, as evidenced by a 2.1% decrease in the probability of ever entering the workforce and a cumulative impact resulting in a -1.36 reduction in total years worked. Conversely, men experience only a slight impact on their likelihood of entering the labour force (-1.9% but not statistically significant) but are notably affected in terms of their initial remuneration. This phenomenon is especially pronounced among uneducated men (worse placement of -3.66 percentiles) who attempt to ameliorate their disadvantaged starting position by transitioning between more jobs (0.052). The permanent effects of early labour market conditions are further revealed in the characteristics of the main job that are of significantly inferior quality. In particular, individuals who encountered elevated unemployment rates upon completing their education found themselves in workplaces characterised by increased physical and emotional demands, along with frequent conflicts. Additionally, those commencing their careers under unfavourable initial labour market conditions also reported an inadequate working environment for the acquisition of human capital. This inferior placement is further substantiated by measures of perceived recognition, support, and fairness within the work environment. The effects are more pronounced among men, although women are not exempt from experiencing them. Notably, the heterogeneity analysis states that this effect is more pronounced among cohorts that completed their education in the early 1970s, underlining the consequences of stagflation periods.

Apart from estimating the causal effects, I conduct numerous simulations to illustrate the capabilities of Causal Machine Learning techniques in addressing model misspecification and capturing underlying heterogeneity. Additionally, I employ simulations to assess the robustness of the results to hyperparameter tuning and their reliability in finite samples.

This work leads to straightforward policy implications. Early labour market conditions affect permanently long-term outcomes, yet the extent and nature of this effect vary among individuals and are not uniform. Hypothetically, the economic cost of supporting individuals during an economic crisis may be outweighed by potential long-term gains. For instance, implementing tax relief schemes or reducing the barriers to accessing university education could potentially enhance both human capital development and job availability. Since resources are always limited, policy interventions should prioritise individuals who are more severely affected. The in-depth analysis of heterogeneity proposed in this paper suggests that directing funds toward less educated individuals is optimal. Moreover, directing funds towards women could mitigate their exclusion from the labour market, while directing funds towards men could improve the quality of their employment.

This paper is organised as follows. Section 2 briefly discusses the relevant literature. Section 3 presents the data. Next, section 4 discusses the identification and the estimation strategies. Then, section 5 presents the baseline results and heterogeneity analysis. Afterwards, section 7 discusses the explaining mechanisms of the long-term effects. Section 8 presents validity and robustness checks. Finally, section 9 concludes.

2 Related literature

The long-term impact of initial adverse labour market conditions is a major concern in economic literature. Kahn (2010) points out that US college graduates who graduated during the deep recession in the early 1980s struggled with a persistent (up to 15 years) reduction in wages. Similarly, Maclean (2013) provides evidence that the above-mentioned sample is negatively and persistently hit also in health outcomes. Then, additional evidence is provided by Oreopoulos et al. (2012) who show that Canadian college graduates who entered into labour markets characterized by unfavourable conditions during the 1980s and 1990s have experienced earning declines lasting up to 10 years. More recently, Schwandt and Von Wachter (2019) aggregated several large US datasets to expand the analysis on subgroups usually not considered such as non-college graduates, women and minorities. Results show that the long-term decrease in wages lasts up to 10 years and the most disadvantaged categories are the most affected. Furthermore, Schwandt and Von Wachter (2020) show that worse economic opportunities at graduation do not impact only the wage trajectory but also might impact the outcomes over the life cycle by worsening the family outcomes and increasing mortality. Again, Gregg and Tominey

(2005) use the National Child Development Survey to estimate the long-term impact of unemployment experienced at the ages between 16 and 23 on UK young males born in 1958. Their main results show that one additional year of unemployment during young adulthood creates scarring effects on wages at age 42 of 13–21%.

Since the above-mentioned literature considers only Anglo-American countries, it cannot be broadly generalized. For instance, the European institutional frameworks consist of more rigid labour markets and provide more generous social support with respect to the US case. As a consequence, in European countries, the immediate negative effects of economic downturns (e.g. on health) might be better mitigated. Again, the longer-lasting unemployment benefits might be helpful to reduce job mismatching. At the same time, more generous social support might unnecessarily delay the entrance into the labour market. Moreover, the labour market rigidities may shrink the employment opportunities for certain subpopulations. A priori it is not clear which effect dominates in the long run. For instance, Cutler et al. (2015) confirm that European youngsters graduating during bad times have significantly worse long-term outcomes (e.g., health and life satisfaction). In particular, the effects are especially strong for less-educated workers. Next, Liu et al. (2016) uses data from Norway to provide evidence about the importance of the match between skills and occupations in determining the permanent negative effects on earnings. Then, Van den Berge (2018) exploits field-specific differences in the economic conditions of highly educated vocational and academic graduates in the Netherlands. Results suggest that both groups face lower wages with the academic group suffering more at the early stages, whereas, the vocational group experiences much more persistent dynamics. Again, the explaining channel through which the detrimental effects arise is the job mismatch and the observed catch-up mechanism relies on job mobility. Fernández-Kranz and Rodríguez-Planas (2018) show that in Spain, a country characterized by high wage rigidity and segmented labour markets, both college and non-college graduates entering into the labour market during a recession have fewer opportunities to become employed which lasts for at least 7 years. However, for college graduates the effect is smaller and less persistent. Cockx and Ghirelli (2016) show that the unlucky graduates in the Flanders region of Belgium experience a persistent negative effect on earnings. In detail, high educated workers face a reduction in the hourly wage and not in the annual hours worked, on the contrary, low-educated employees encounter a reduction in the working hours but not in the hourly wage. Again, Brunner and Kuhn (2014) point out that in Austria unfavourable entry conditions generate a sizable negative long-run effect, especially for blue-collar workers which remain permanently locked into low-paying jobs/tasks. A similar study conducted by Päällysaho (2017) shows that Finnish unlucky graduates experience sizable and persistent negative effects on labour market outcomes, such as earnings and employment status, for no less than 10 years.

3 Data

The analysis is based on two datasets. First, I am proposing a novel dataset containing regional unemployment rates for five Western countries (Austria, (West-)Germany, Italy, Denmark, and Belgium) during the period 1960-1990. Then, the microdata (outcomes, childhood conditions, other confounding variables, and schooling and migration histories) is collected from SHARE.

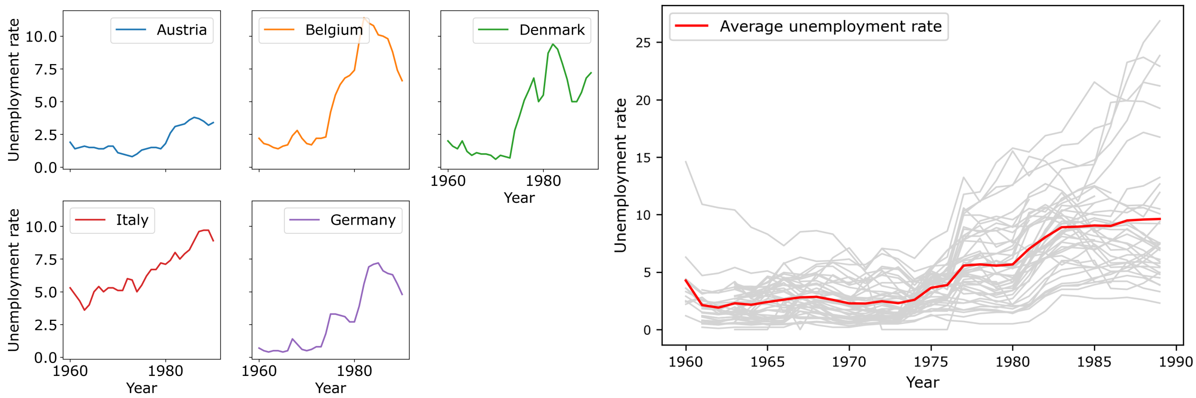
3.1 Unemployment rates

Following the vast majority of the literature, I capture labour market conditions at the time of completing education by making use of unemployment rates. Figure 1a displays the unemployment rate trajectories of the countries under analysis. It is interesting to notice that unemployment rates, owing to the economic booms following WWII, were relatively low for all countries until the beginning of the 1970s. Then, unemployment suddenly increased during the 1970s stagflation period, remaining anchored at larger values than before.

Next, since country-level data hides huge within-country variability (see section A in the appendix), employing the country level to measure the labour market conditions at graduation might raise remarkable measurement errors. To overcome this issue, I propose a novel dataset of regional unemployment rates collected from different historical sources (more details are available in section A). Figure 2 enforces the claim (from a geographical perspective) that it is important to consider also the local labour market variability.

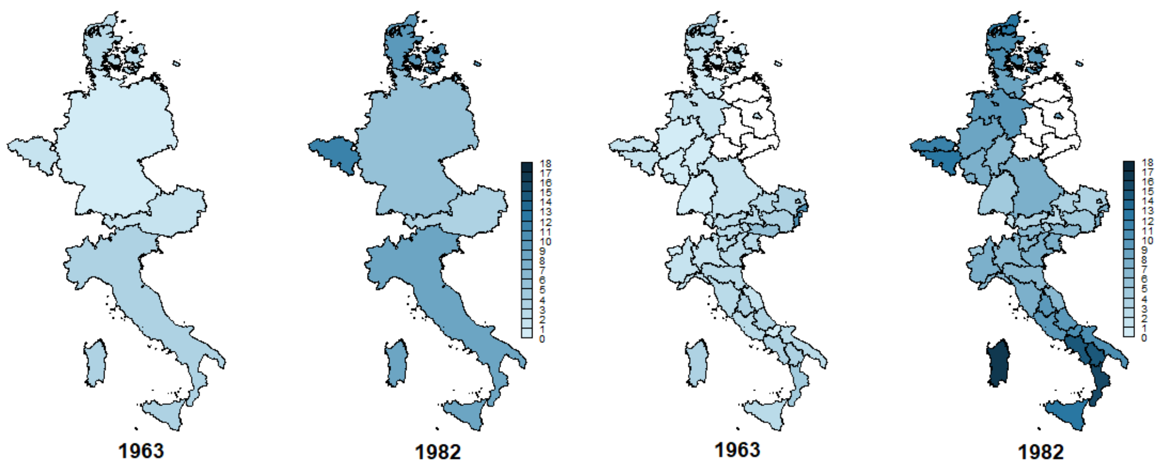
Country	Austria	Germany	Italy	Denmark	Belgium
Start	1960	1961	1963	1961	1960
End	1990	1986	1990	1990	1982
NUTS	2	1	2	2	1

Table 1: Time and geographical availability of regional unemployment rates. NUTS stays for the Nomenclature of Territorial Units for Statistics (NUTS), a hierarchical system for dividing up the economic territory of Europe.



(a) (b)

Figure 1: Figure 1a shows the unemployment rate at the country level, 1960 - 1990. Source: AMECO. Figure 1b presents the different trajectories of the regional unemployment rates along with the average unemployment rate (own data).



(a) (b)

Figure 2: Figure 2a displays country-level unemployment rate and figure 2b regional unemployment rates respectively in 1963 and 1982. Source: AMECO and own data.

Table 1 provides a concise overview of the regional unemployment rate data based on different time periods and geographical levels. With the exception of Belgium, all countries have nearly three decades' worth of data spanning from 1960 to 1990. Table 4 highlights that the majority of observations are concentrated around the year 1972, and hence containing also the stagflation periods. Austria, Italy, and Denmark offer data at the NUTS 2 level, while Germany and Belgium provide data at the NUTS 1 level.

3.2 SHARE microdata

Microdata is retrieved from the Survey of Health, Ageing and Retirement in Europe (SHARE)[2], a multidisciplinary, cross-national household panel survey which collected from 2004 until today more than 530,000 in-depth interviews with 140,000 people aged 50 or older from 29 European countries. The samples contain data on health, labour market outcomes, socio-economic status, family situation, and childhood circumstances at a nationally representative level for elderly people. Specifically, so far, 9 waves of interviews occurred including the last two which focused also on COVID-19 topics. It is important to point out that the information that I used is gathered both from the "regular" waves containing information at the time of the interview (up to wave 7) (Börsch-Supan et al. (2013)) and the "retrospective" waves (waves 3 and 7) containing retrospective information (Brugiavini et al. (2019)).

The first concern which could rise in front of survey data regards the validity of the data due to imprecision in responding or recall bias in case of retrospective answers. However, although SHARE may collect less precise data with respect to administrative sources, the wide availability of information far from graduation, detailed childhood conditions, and lifetime migration and schooling pathways provide an ideal setting to be employed for the current task. In addition, the literature provides several validation studies which support the quality and reliability of SHARE data. The first notable work is from Bingley and Martinello (2014) who link the same individuals participating to SHARE interviews to Danish Administrative Registers to compare their educational attainment, labour market status and household income. The authors claim that the measurement error of survey information is small. In fact, they claim that household income is on average not statistically different in SHARE Denmark and register data and only the level of schooling is slightly overstated in the survey data. Next, Garrouste and Paccagnella (2011) employ an "internal" validation test by comparing "retrospective" and "regular" information regarding family and labour market outcomes. Their main conclusion is that the two sources are highly consistent. Then, Mazzonna and Havari (2011) employs external country-level data (e.g., GDP per capita, average years of schooling and war episodes) to confirm the validity of the childhood histories from retrospective interviews.

The sample used in the main analysis contains individuals concluding full-time education in the 1960-1990 time period in Austria, Germany, Italy, Denmark, and Belgium[3] who are interviewed between ages 50 and 70. The baseline sample contains 10893 unique individuals.

[2]http://www.share-project.org/

[3]Only these five countries have been selected owing to the availability of historical regional unemployment rates.

Country	Austria	Germany	Italy	Denmark	Belgium	Total
Observations	1,564	1,782	2,168	2,194	3,185	10,893

Table 2: Sample frequency of individuals by country.

3.2.1 Outcome data

The first relevant aspect of SHARE data is that it allows for analysing the long-term outcomes very far away from the year of completing education. In detail, the average time span elapsed from finishing studies to the measure of the outcomes is 38 years. Additionally, a broad range of outcomes can be considered: labour market, health, and family outcomes.

It is worth mentioning that, when dealing with labour market outcomes, the main sample is restricted only to not retired individuals less than 65 years of age[4]. The first outcome of interest is labour market participation during late adulthood, i.e., an individual is considered "Active" if receives a positive income from employment or self-employment. Next, "Earnings from work" contains the earnings from employment and self-employment. Additionally, since the effects on earnings could be driven by the number of working hours or by the hourly remuneration, I consider the weekly "Working hours" and "Hourly earnings"[5]. It is important to stress that all original income data presents missing data due to non-responses and consequently SHARE provides carefully designed imputations (De Luca et al. (2015)).

Health, which represents the second main area of interest, can be observed from different dimensions. To cover the most important ones, I restrict the attention to three different categories. The first one, which is the baseline in the literature, is the general subjective health measure, "Self-perceived health"[6], containing 5 levels going from 1 = "Excellent" to 5 = "Poor". However, since subjective measures might not be correctly reported (Bound (1989)), a more objective health measure is considered: "Grip strength". It consists in physically testing the grip strength of interviewed individuals. The measure goes from 1 to 100, where 100 is the maximum achievable. Finally, since also mental health is an important health dimension which sometimes is overlooked, is checked the "EURO-D depression scale" which measures the number of depression symptoms (out of 12)[7].

[4]This restriction is done to avoid considering the outcomes of particular individuals such as entrepreneurs who still work after traditional retirement ages. The results remain robust if the threshold is reduced to 60 or increased to 70.

[5]"Earnings from work" are divided by 52, then divided again by the number of weekly working hours.

[6]Additional health measures such as "Number of chronic diseases", "Number of mobility limitations", "ADL", and "IADL" provide the same results as the baseline self-perceived measure.

[7]The EURO-D measure consists of the following questions: depression, pessimism, suicidality, guilt,

Finally, the well-being of individuals depends also on family outcomes. Hence, the following measures are analysed: "Number of children"[8] and "Never married or currently divorced".

The above-mentioned outcomes provide a good representation of socio-economic and health conditions later in life. Nevertheless, this approach does not allow for unravelling the complex inter-dependencies among the different outcomes.

	Outcome	Observations	Mean	Std dev	Min	Max	Years after completing education
Labour market	Active into labour market	8076	0.73	0.44	0	1	36
	Earnings from work	5899	9.86	0.99	2.59	12.95	35
	Working hours	6272	36.37	12.12	1	70	35
	Hourly earnings	5840	2.50	0.75	0.44	4.27	35
Health	Self-perceived health	10893	2.79	1.05	1	5	38
	Objective grip strength	10364	37.63	11.65	1	90	38
	EURO-D depression scale	8581	2.10	2.11	0	12	38
Family	Number of children	10770	1.98	1.21	0	10	38
	Never married or divorced	10893	0.18	0.38	0	1	38

Table 3: Descriptive statistics of the long-term outcomes variables. Labour market outcomes contain only individuals younger than 65 who do not perceive a retirement pension. Years after completing education is the mean of the years that have passed from the time of completing education to the time of the interview.

3.2.2 Control variables and childhood conditions

Along with the traditional controls such as the year of graduation, age of interview, year of birth, year of interview and region of graduation[9], and gender which are usually employed in the literature, I propose additional pre-treatment controls at the individual level. Literature shows that socio-economic status (SES) during childhood is an important indicator which influences almost all adulthood outcomes (Currie (2009), Case et al. (2002), Case et al. (2005)). One of the most important measures of SES during childhood which I can control for is the occupation of the breadwinner at age 10 according to the skill level and job characteristics (ISCO code at 1 digit level)[10]. The other SES measures consist of the accommodation conditions during childhood ages: the number of rooms

sleep, interest, irritability, appetite, fatigue, concentration (on reading or entertainment), enjoyment, and tearfulness

[8]The alternative measure of "At least one child" provides robust results.

[9]Since the estimation strategy relies on a random forest algorithm, regions of graduation were aggregated to reduce the number of categorical variables but still maintaining the coherency of "regional fixed effects". For more details see A.6 in the appendix.

[10]The aggregation of the main breadwinner occupation are defined by the following rules based on ISCO 1-digit codes. Low skill, Blue collar: 0) Armed forces, 6) Skilled agricultural or fishery worker, 8) Plant/machine operator or assembler, 9) Elementary occupation. High skill, Blue collar: 7) Craft or related trades worker. Low skill, White collar: 4) Clerical Support Workers. High skill, White collar: 1) Legislator, senior official or manager; 2) Professional; 3) Technician or associate professional.

per capita[11] and the number of features (fixed bath, cold running water supply, hot running water supply, inside toilet and central heating). Next, I employ different proxies of childhood ability. The first available ability proxy (but also SES measure) is the number of books at home at age 10. It assumes value 1 if "None or very few (0-10 books)" and 5 if the answer was "Enough to fill two or more bookcases (more than 200 books)". Next, I can use information about the relative position in math and language abilities at age 10. These proxies assume value 5 if "Much better" compared to others and value 1 if "Much worse". It is clear that the self-declared ability position is only a broad proxy of the real ability. Nevertheless, these measures implicitly capture another intrinsic characteristic which impacts both the educational pathway and long-term outcomes: self-confidence. Further, also health measured before graduation can be an important factor in the education pathway. To control for it, a principal component analysis is performed on a large set of childhood illnesses and then categorized according to the quintile (where 1 corresponds to the most healthy quintile). Finally, are considered dummies capturing whether the individuals were born in a different country with respect to the graduation one and if they were born in a rural area.

3.2.3 Migrations

Another relevant aspect of the SHARE microdata is that it allows to geographically locate each year of respondents' life at the relevant local labour market level. This provides three main advantages. First, I am able to set the local labour market condition at a more appropriate level with respect to the country. Consequently, the attenuation bias due to measurement errors in labour market conditions is drastically reduced. Then, pre-graduation migration histories allow me to understand how relevant endogenous migration decisions are. Finally, post-graduation migration decisions can be employed to test whether post-graduation migrations could hedge in the long run from early labour market shocks.

In literature, it is usually claimed that migration flows in European countries are a relatively small phenomenon, especially if compared with the US. However, by taking into account also movements between regions, migration flows are quite comparable with the US ones. As a matter of fact, 29.41% of individuals in my sample claimed to have moved their region of residence at least once in their life (in the US Molloy et al. (2011) claims that the lifetime cross-state migration rate is 32 %). As far as regards between-country migration, overall 259 individuals (2.38% of the sample) were born in a different country with respect to the interview location. On the contrary, when observing the

[11] The number of rooms per capita assumes value 1 if less than 1 room per capita, 3 if more than 3 rooms and 2 otherwise.

	Control variables	Mean	Std dev	Min	Max
Baseline controls	Year graduation	1972.05	6.33	1960	1989
	Age at interview	60.11	5.01	50	70
	Year of birth	1953.36	5.60	1938	1967
	Year of interview	2009.88	3.82	2004	2017
	Male	0.46	0.50	0	1
	Educational attainment	2.02	0.75	1	3
	Region graduation				
Childhood SES	Breadwinner occupation				
	• *Low skill, Blue collar*	0.35	0.47	0	1
	• *Low skill, White collar*	0.19	0.35	0	1
	• *High skill, Blue collar*	0.26	0.42	0	1
	• *High skill, White collar*	0.19	0.39	0	1
	Home, rooms per capita	1.23	0.45	0	3
	Home, features	3.17	1.67	0	5
Childhood Ability	Number of books	2.46	1.26	1	5
	Math ability	2.69	0.90	1	5
	Language ability	3.38	0.87	1	5
Childhood Health	Childhood Health	2.80	1.54	1	5
Other aspects	Born abroad	0.02	0.15	0	1
	Born in rural area	0.37	0.48	0	1

Table 4: Descriptive statistics of the confounding variables.

within-country movements, the picture radically changes. For instance, the year after graduation no less than 623 individuals (5.72% of the sample) moved into another region within the country of graduation and the number increases cumulatively to 1052 (slightly less than 10% of the sample) after three years. Again, those moving towards a different region in the five years prior to ending full-time education are 669 (6% of the sample).

4 Identification and estimation

4.1 Identification

To identify the causal impact of recessions on long-term outcomes, the characteristics of different graduation cohorts should not be related to local labour market conditions. The first point to notice is that the structure of educational systems is fixed and the year of completing education, and consequently the early opportunities, are to some degree exogenous. However, there is still the possibility for individuals to endogenously self-select the timing and/or the location of graduation according to the labour market conditions.

The first source of endogeneity derives from the fact that people might prolong or shorten their educational pathway in order to avoid bad conditions or to exploit favourable labour market circumstances. For instance, if the individuals with better potential outcomes are better at self-selecting the timing of graduation, then the effects of bad early labour market conditions on long-term earnings would be upward biased. To check this type of selection, I regress the unemployment rate at the time of completing education on childhood conditions controlling also for region of birth and year of birth fixed effects (equation 1).

$$UR_i^{rg} = \beta_0 + \beta_1 child\, condition_i + \theta_{region\, birth} + \phi_{year\, birth} + \epsilon_i, \tag{1}$$

where UR_i^{rg} is the unemployment rate experienced in region r in the year g of completing education. $child\, condition_i$ is a variable measuring the childhood conditions between 10 and 15. $\theta_{region\, birth}$ and $\phi_{year\, birth}$ are respectively the region and year of birth fixed effects.

Results in figure 3 show that graduation timing is not random. Surprisingly, there is a positive relationship between the unemployment rate at graduation and all the socioeconomic-ability proxies measured during childhood even after controlling for year and region of birth FEs. Hence, in my sample, the individuals with higher potential outcomes are worse at selecting the graduation timing. Nevertheless, childhood conditions are excellent candidates to control this possible bias.

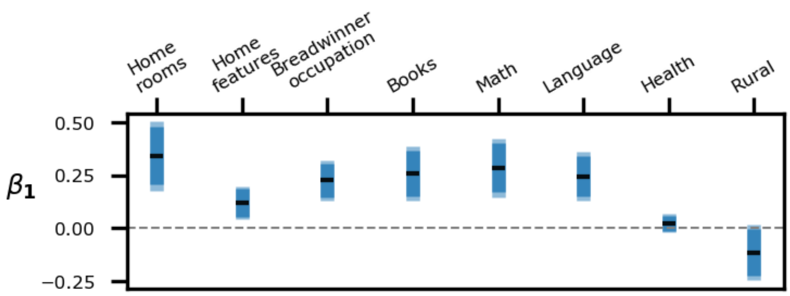


Figure 3: OLS regressions of the unemployment rate at graduation on childhood conditions. The additional control variables include birth cohort and region of birth fixed effects. Standard errors clustered at the level of year-region of graduation. The darker and lighter areas denote significance at 10% and at 5% respectively.

The second source of endogeneity regards migration decisions. As a matter of fact, since recessions can hit local labour markets in a very heterogeneous manner, people can react to bad economic conditions by migrating to labour markets that are better off. Consequently, endogenous migration can be a problematic source of bias. In this regard, SHARE data provides detailed information about migration pathways, allowing me to control for endogenous migration before graduation. To understand the role of endogenous migration in my context, I control for the relationship between the unemployment rate experienced prior to graduation and subsequent migration decisions. The first point to raise is that people might migrate independently of local labour market circumstances, and since in my sample only 6% of the individuals move to a different region in the five years before graduation, migrations are a relatively small phenomenon. To further corroborate this claim, I regress a migration dummy[12] on the unemployment rates the individual faced five years before graduation (equation 2).

$$migration_{i,(g-5,g)} = \beta_0 + \beta_1 UR_{i,g-5} + \theta_{region\,birth} + \phi_{year\,birth} + \epsilon_i, \tag{2}$$

where g is the year of completing education. $migration_{i,(g-5,g)}$ is a dummy equal to 1 if the individual migrated at least once in the 5 years before graduation. $UR_{i,g-5}$ is the unemployment rate experienced 5 years before completing education. $\theta_{region\,birth}$ and $\phi_{year\,birth}$ are respectively the region and year of birth fixed effects.

Figure 4 shows that unemployment rates experienced five years before graduation are mildly associated with migration choices. Individuals/families deciding to migrate do not have the same socio-economic background. As a matter of fact, once childhood controls are included in the regression, the relationship between unemployment rates experienced prior to graduation and migration decision is no longer statistically significant. It is noteworthy to mention that the bias spawned from migration decisions goes in the opposite direction with respect to timing decisions. Specifically, on average, migration is not random as the destination region has on average a lower unemployment rate with respect to the departure one (3.54% vs 3.91%[13]). Once more, childhood conditions stand as strong candidates for mitigating this potential bias.

To sum up, the main source of endogeneity in my setting is related to the possibility of having graduating cohorts with different characteristics due to self-selection in timing and location. Hence, although the educational system fixes to some degree schooling decisions, by ignoring childhood microdata conditions, the estimated effects could be biased owing to an omitted variable issue. Nevertheless, childhood health, ability, and socioeconomic status are able to control for timing and migration decisions, and so, are likely to represent a valid set of controls to radically mitigate the omitted variable bias. To identify the causal impact of early local labour market conditions on the long-run

[12] The migration dummy assumes value 1 if the individual migrated at least once in the 5 years before graduation. If the individual aims to select the location in order to maximize the labour market opportunities at graduation, it is reasonable to believe that most happen in the last cycle of education. Nevertheless, results remain robust if the 5-year span is changed.

[13] The average unemployment rate for individuals who migrated at least once in the 5 years before completing their education was 3.54%. The unemployment rate would have been 3.91% if the same individuals had not migrated.

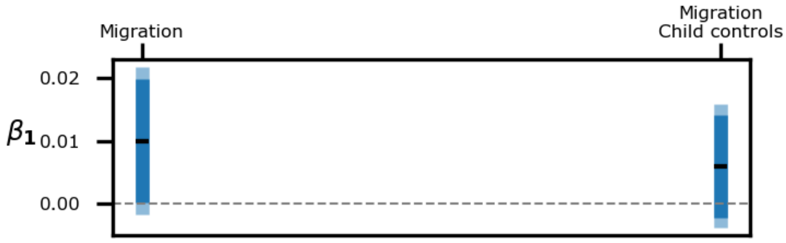


Figure 4: OLS regressions of migration dummy (migrated at least once in the 5 years before the end of full-time education) on unemployment rate five years before graduation. The additional control variables include birth cohort and region of birth fixed effects. Standard errors clustered at the level of year-region five years before graduation. The darker and lighter areas denote respectively significance at 10% and at 5%.

outcomes, I rely on two identification strategies: unconfoundedness after controlling for the rich set of childhood conditions and an instrumental variable approach.

4.1.1 Unconfoundedness

The first strategy to identify the causal effects relies on the unconfoundedness assumption:

$$Y_i(t) \perp\!\!\!\perp T_i | X_i = x_i, \quad t \in \mathcal{T}, \quad \mathcal{T} = [t_0, t_1],$$

where $Y_i(t)$ is the potential outcome of individual i in case of treatment t, T_i is the level of treatment received, and $\mathcal{T}$ is the set of continuous treatments in an interval $[t_0, t_1]$ (Hirano and Imbens (2004)).

Again, figure 5 provides the graphical perspective of unconfoundedness through Directed Acyclic Graphs (DAG).

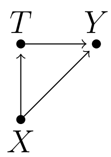


Figure 5: Directed Acyclic Graphs (DAG) for unconfoundedness.

Unconfoundedness implies that there are no features other than X that jointly influence treatment (T) and potential outcomes (Y). Or simply, treatment is as good as random conditionally on confounding variables. Hence, the validity of this identification strategy relies on the availability of very detailed information regarding X which assures the absence of omitted variable bias, i.e., X must contain all the variables associated with T and affect Y conditional on T. In my specific analysis, the credibility of the identifying assumption relies on the availability of a rich set of confounding variables able to control for the timing and migration selection. Along with the traditional controls such as the year of graduation, age of interview, year of birth, region of graduation, and gender which are usually employed in the literature, I rely on the bunch of childhood conditions presented in the previous section 3.2.2. In this regard, literature extensively shows that socio-economic status during childhood is an important indicator which influences almost all adulthood outcomes (among others, see Currie (2009), Case et al. (2002), Case et al. (2005)). Hence, the endowment of childhood health, ability, and socio-economic background are likely to impact both the potential outcomes in the long run and the graduation timing and migration decisions. To sum up, early childhood conditions are necessary information to control for the omitted variable bias.

4.1.2 Instrumental variable

The credibility of the unconfoundedness assumption is supported by the way the treatment is defined and the inclusion of a rich set of childhood conditions that adjust for endogenous timing and migrations. However, since it is not possible to directly test the unconfoundedness assumption, concerns may arise regarding the potential presence of unobserved confounding variables that could affect the validity of the causal inference. Figure 6 provides a graphical representation of the DAG, which illustrates that if any relevant unobserved confounding variables (u) are not captured, it is impossible to infer the causal effect of T on Y from the joint distribution of T and Y.

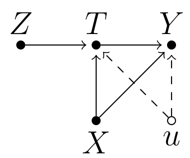


Figure 6: Directed Acyclic Graphs (DAG) for the instrumental variable.

To solve this issue, a specific new variable (the instrument Z) is needed. In detail, Imbens and Angrist (1994) and Angrist et al. (1996) provide the instrumental variable setting for identifying Local Average Treatment Effects. To perform a consistent estimate of the local average treatment effect for the compliers, the instrument Z must be a good predictor for the endogenous regressor T (Instrument Relevance). Furthermore, the instrument should be exogenous, i.e., there is no unmeasured confounder that affects the instrument Z and the outcome Y. In detail, the key identification assumption of the IV model is that, after controlling for the confounding variables X, the instrument Z is correctly excludable from the second stage. Figure 7 displays a DAG in which the exclusion restriction is violated. The last condition to satisfy states that the change in the instrument should have a monotone effect on the treatment, i.e., monotonicity should be satisfied.

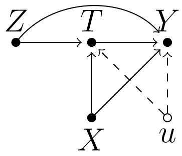


Figure 7: Directed Acyclic Graphs (DAG) for the instrumental variable. Violation of exclusion restriction.

To recover an appropriate instrumental variable, I rely on previous literature (Gregg and Tominey (2005), Kahn (2010), Oreopoulos et al. (2012), Maclean (2013), Brunner and Kuhn (2014)) that instruments the unemployment rate with rates of exogenous timing and location. In detail, my instrument consists of the unemployment rate the individual would have faced in the first year in which he/she would have been allowed to drop school by the compulsory minimum school-leaving age laws in the location he/she effectively had resided at that age. The validity of this approach relies on the fact that both the timing and location defining the assigned instrument are exogenously determined by the year of birth and the compulsory minimum school-leaving age laws.

First, I provide support for the relevance of the instrument by showing the first-stage F-statistics in appendix B.1. There is evidence of a strong relationship between the instrument and the unemployment rate at graduation[14]. However, to better define the estimand it is necessary to better investigate who are the compliers. In section B.1.1, I show that the compliers are especially low-educated individuals.

Then, I show that exogeneity is likely to be achieved since the effect of the instrument is likely to impact the long-term outcomes only through the path-dependency of the labour market conditions at graduation. To support the excludability of the instrument, I follow the same procedure as in section 4.1 to check whether the childhood conditions are related to the instrument. Figure 8 shows that the proposed instrument is not related to childhood conditions. Since the instrument is not related to the observables[15], it is likely not predicted also by other unobservables and is correctly excludable from the second stage. Additionally, the reverse effect of the long-term effect on the instrument is not possible in the current setting.

Finally, subsection B.3 provides evidence about the monotonicity condition in my setting.

[14] F-statistics are always between 42 and 123, according to the sample size of the outcome considered.

[15] Note that this evidence is particularly strong since the observables in my setting are childhood SES, ability, and health which are considered unobservable factors in most of the datasets.

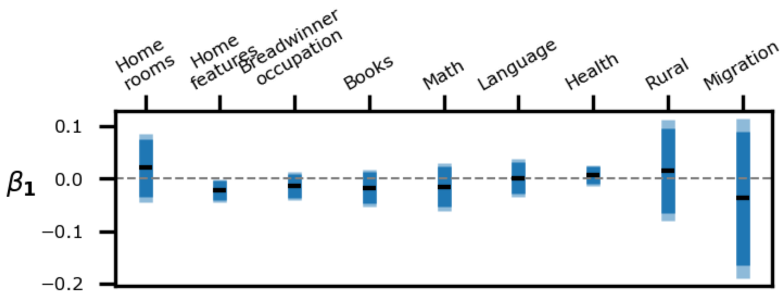


Figure 8: OLS regressions of childhood conditions on the unemployment rate determined by exogenous year of birth and compulsory schooling laws (instrument). The additional control variables include birth cohort and region of birth fixed effects. Standard errors clustered at the level of year-region of compulsory age. The darker and lighter areas denote respectively significance at 10% and at 5%.

4.2 Estimation

I estimate the causal effect of early labour market conditions on long-term outcomes following three different procedures: OLS, 2SLS and Generalized Random Forest (GRF). OLS and GRF exploit the rich set of childhood conditions under the unconfoundedness assumption to recover the causal effects. In the 2SLS procedure, in addition to the childhood conditions, the causal impact is identified by employing the exogenous unemployment rate defined by the compulsory minimum school-leaving age. Standard errors are clustered at the level of graduation year by region to account for region-cohort-specific serial correlation (Moulton (1990)). Further, I employ SHARE sample weights that account for the sampling survey design.

The first estimand of interest is the Average Partial Effect (APE), i.e., the change in the expected long-term outcome due to a small change in the unemployment rate faced at graduation.

$$APE(X,t,\delta) = \frac{E[Y(X,T=t+\delta) - Y(X,T=t)]}{\delta}, \tag{3}$$

where Y is the potential outcome, t is the treatment value, and δ is the infinitesimally small change in the treatment value.

However, the average effect could hide deep heterogeneity. In this regard, causal machine learning techniques are better suited to estimate heterogeneous treatment effects by being able to estimate the Conditional Average Partial Effects (CAPE), and so, estimate the effects of relevant groups of interest (GAPE) or even individualized treatment effects.

$$GAPE(X,t,\delta|G=g) = \frac{E[Y(X,T=t+\delta|G=g) - Y(X,T=t|G=g)]}{\delta}, \tag{4}$$

where g is the group of interest.

The three different estimation procedures could shed light on the robustness of the results to the choice of the estimators and their identification strategies. Additionally, since causal machine learning (CML) estimators are still in the early stages of adoption in empirical economics, it is interesting to understand how reliable are with respect to other widely accepted methods in the literature.

4.2.1 Standard Estimators: OLS and 2SLS

The first estimation procedure follows the linear regression model using OLS:

$$y_i = \beta_0 + \beta_1 UR_{rg} + \beta_2 \mathbf{X}_i + \gamma_g + \theta_r + \phi_b + \eta_t + \mu_a + u_i \tag{5}$$

where y_i is the long-term outcome of individual i, UR_{rg} is the unemployment rate faced at graduation in the region r by graduation-cohort g, $\mathbf{X}_i$ contains fixed individual characteristics such as gender and the rich set of childhood conditions. γ_g, θ_r, ϕ_b, η_t, and μ_a are respectively the year of completing education, region of graduation, year of birth, year of interview, and age at interview fixed effects. The year of completing education FE controls for differences between graduating cohorts. Regional fixed effects control for permanent local labour market characteristics, both in reporting styles and institutions affecting lifetime income. Year of birth and year of interview check respectively for specific characteristics of any given birth-cohort and the "current" conditions experienced in each year of interview. Age at interview controls for natural differences in outcomes according to age.

Next, I follow a standard two-stage least squares (2SLS) estimation procedure:

$$UR_i = \alpha + \Pi_1 Z_{lc} + \mathbf{\Pi}_2 \mathbf{X}_i + \gamma_g + \theta_r + \phi_b + \eta_t + \mu_a + v_i \tag{6}$$

$$y_i = \beta_0 + \beta_1 \widehat{UR}_i + \beta_2 \mathbf{X}_i + \gamma_g + \theta_r + \phi_b + \eta_t + \mu_a + u_i \tag{7}$$

In the first stage (equation 6) Z_{lc} is the instrument representing the unemployment rate the individual would have faced during the first year (c) in which he/she would have been allowed to drop school by the compulsory minimum school-leaving age laws in the location (l) he/she effectively had resided at that age. The remaining baseline control variables are the same as in the OLS case.

4.2.2 Generalized Random Forest (GRF)

The last estimation procedure which I am proposing is built on recent developments in the literature on machine learning methods applied to causal inference frameworks (for a brief review of the state of the art Athey and Imbens (2017) and Athey and Imbens (2019)). Causal Machine Learning (CML) combines the prediction power of the statistical learning literature (for an overview see Hastie et al. (2009)) with the microeconometric literature on defining and identifying causal effects. The combination of the two kinds of literature allows me to estimate the usual average partial effects in a more flexible way, and so, the risk of model misspecification is reduced[16]. In addition, CML naturally enables me to uncover partial effect heterogeneity by providing Group Average Partial Effects (GAPE) for the relevant subgroups of interest, or even partial effects at the individual level. This allows me to further test the reliability of the uncovered heterogeneity. Detecting heterogeneity is crucial since allows a better understanding of who are the winners and the losers of any given "policy treatment".

In detail, I use Causal Forests to estimate conditional average partial effects by relying on the Generalized Random Forest (GRF) framework (Athey et al. (2019)), a method for nonparametric statistical estimation based on random forests (Breiman (2001)). This estimation framework is advantageous from several perspectives. From a theoretical point of view, its flexibility allows it to adapt to different functional forms and, in addition, guarantees formal asymptotic results for statistical inference (asymptotic consistency and Gaussianity of the estimates). Next, from an operational point of view, a freely available and detailed R package is provided[17]. Again, as far as regards the current causal task, GRF allows for continuous treatment assignment.

Conversely to the traditional random forests which are understood as ensemble methods where predictions are an average of predictions made by individual trees[18], GRF reinterprets the forests as a type of adaptive nearest neighbour estimator, which is more suitable for statistical extensions.

Formally, by exploiting the identification via local moment conditions, GRF can be used to fit any quantity $\tau(x)$ of interest:

$$E[\psi_{\tau(x),\nu(x)}(Y_i, T_i)|X_i = x] = 0, \text{ for all } x \in \mathcal{X}, \tag{8}$$

where $\psi(\cdot)$ is some scoring function, $\tau(x)$ is the parameter of interest, $\nu(x)$ is a nuisance

[16] In section E.2 I perform simulations to show the robustness of GRF in contrast to the standard linear OLS model.

[17] https://grf-labs.github.io/grf/index.html

[18] However, the main elements of the conventional random forest such as recursive partitioning, subsampling and random split selection are still maintained in the GRF structure. For a broader view of random forest and the difference of GRF see section E.

parameter, Y_i is the outcome, and T_i is the treatment assignment.

One way to perform the estimation of $\tau(x)$ consists in defining some kind of similarity weights $\alpha_i(x)$ that capture the relevance of the i^{th} training observation to fitting $\tau(\cdot)$ at x. Then, in the second step, the target of interest is fitted via an empirical version (see equation 9) of the estimating equation 8.

$$(\hat{\tau}(x), \hat{\nu}(x)) \in \underset{\tau,\nu}{argmin}\left\{\left\|\sum_{i=1}^{n} \alpha_i(x)\psi_{\tau,\nu}(Y_i, T_i)\right\|_2\right\}, \tag{9}$$

where Y_i and T_i are respectively the outcome and the treatment assignment of observation i.

In the local maximum likelihood estimation literature (e.g., Fan et al. (1998)), heterogeneous estimating equations are obtained via a deterministic kernel weighting function. However, this approach works only in low dimensions owing to the curse of dimensionality. To overcome this issue, GRF employs forest-based algorithms to adaptively learn the weights $\alpha_i(x)$ to be used in the empirical version of the estimating equation.

In detail, the similarity weights are picked up by the frequency with which the training observation i falls into the same leaf as x

$$\alpha_{bi}(x) = \frac{1(\{X_i \in L_b(x)\}}{|L_b(x)|}, \quad \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x) \tag{10}$$

where, $b = 1, \ldots, B$ are the set of trees, $L_b(x)$ is the set of training units falling in the same "leaf" as x. The weights sum to 1 and define the forest-based adaptive neighbourhood of x. For a more intuitive graphical perspective, figure 9 illustrates how forests define similarity weights.

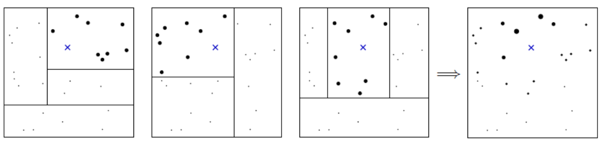


Figure 9: Illustration of the random forest weighting function. The rectangles depicted above correspond to terminal nodes of the random forest algorithm. Each tree starts by giving equal (positive) weight to the training examples in the same leaf as our test point x of interest, and zero weight to all the other training examples. Then, the forest averages all these tree-based weightings and effectively measures how often each training example falls into the same leaf as x. Source: Athey et al. (2019).

The empirical implementation of GRF to get the heterogeneous treatment effect estimation employs a forest-based method based on the R-learner objective function (Nie and Wager (2021)):

$$\hat{\tau}(\cdot) = \text{argmin}_{\tau} \left\{ \sum_{i=1}^{n} \left(\left(Y_i - \hat{m}^{(-i)}\left(X_i\right) \right) - \tau\left(X_i\right) \left(T_i - \hat{e}^{(-i)}\left(X_i\right) \right) \right)^2 + \Lambda_n(\tau(\cdot)) \right\}, \quad (11)$$

where $\Lambda_n(\tau(\cdot))$ is a regularization term which controls for the complexity of the $\hat{\tau}(\cdot)$ function. Y_i is the observed outcome and T_i is the treatment assignment. $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$ are the nuisance parameters $\nu(X_i)$ and are respectively the expected outcome and the generalized propensity score that allows for continuous treatments as in Hirano and Imbens (2004). The $^{(-i)}$-superscripts characterize the "out-of-bag" predictions, i.e., Y_i and T_i were not used to compute $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$.

In brief, by first computing the similarity weights $\alpha_i(x)$ and then combining them with the above-mentioned R-learner objective function, the heterogeneous treatment effects are obtained by

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)(Y_i - \hat{m}^{(-i)}(X_i))(T_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^{n} \alpha_i(x)(T_i - \hat{e}^{(-i)}(X_i))^2} \quad (12)$$

Hence, the Individualized Partial Effects $\hat{\tau}(X_i)$ can be estimated by evaluating $\hat{\tau}(x)$ at the covariate combination of each individual. Next, Average Partial Effects can be estimated by plugging in the Individualized Partial Effects into a variant of augmented inverse-propensity weighting (see for instance Chernozhukov et al. (2018a)):

$$\widehat{APE} = \sum_{i=1}^{N} (\hat{\tau}^{(-i)}(X_i) + \frac{T_i - \hat{e}^{(-i)}(X_i)}{\widehat{Var}(T_i|X_i)} [Y_i - \hat{\mu}^{(-i)}(X_i, T_i)]), \quad (13)$$

where $\widehat{Var}(T_i|X_i)$ is get from an auxiliary forest. Similarly, Group Average Partial Effects can be estimated by this doubly robust average treatment estimator by restricting the summation for individuals in the subgroups of interest (e.g. low-educated individuals). This procedure yields semiparametrically efficient average treatment effect estimates and accurate standard error estimates under considerable generality.

To sum up, GRF allows for continuous treatment assignment by effectively estimating an average partial effect $\frac{Cov[Y,T|X=x]}{Var[T|X=x]}$, which is interpreted as a treatment effect given unconfoundedness. Then, GRF can be naturally extended to account for sampling variability of potentially unexplained cluster-level effects (Athey and Wager (2019)). In the case in which each cluster is assumed to have some effect on an individual's outcome, the random forest algorithm is modified in the sub-sampling stage by drawing a subsample of clusters instead of observations. Then, in the out-of-bag predictions step, in order to account for potential correlations within each cluster, are considered out-of-bag observations only those who are part of a cluster not drawn in the sub-sampling stage.

5 Results

First, I display the CAPEs of completing education in worse economic conditions on the long-term outcomes at the individual level along with the 95% confidence intervals (figure 10). These plots provide the first insights into the results, including the heterogeneity of long-term effects. For instance, it can be seen that being "Active in the labour market" is more evenly distributed around the 0, contrarily to other outcomes such as "Hourly earnings" or "Number of children".

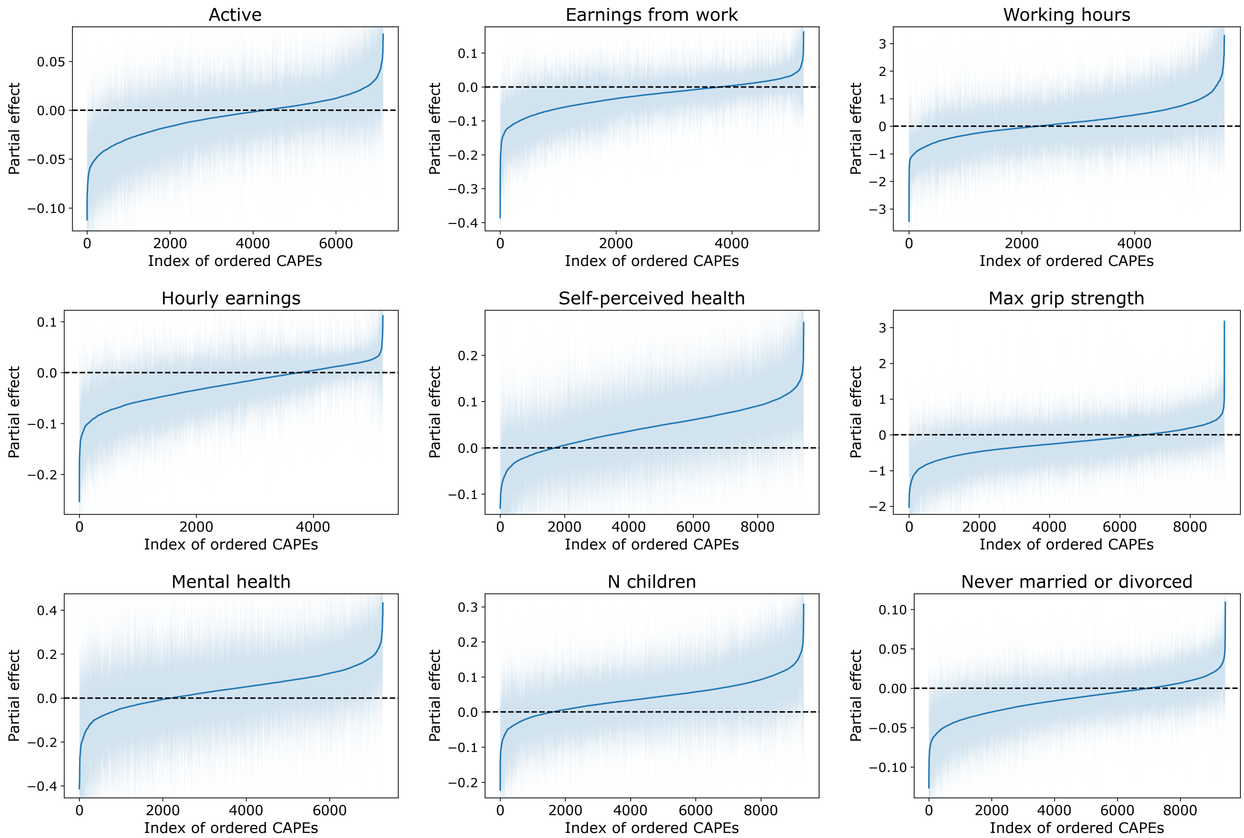


Figure 10: Ordered CATEs for each individual and 95% confidence interval.

Table 5 shows the long-term effects from a tabular perspective. The effects should be interpreted as the average effect of increasing the unemployment rate at the end of the educational pathway by one percentage point.

As a baseline, I employ the Causal Forest's results estimated by GRF. One percentage point increase in the unemployment rate at the time of completing education causes a reduction of 5.2% of earnings from work after more than 35 years after ending the educational pathway. This effect is entirely driven by hour earnings (-4%) whereas working hours are not significantly affected. Furthermore, worse labour market conditions at the time of completing education significantly deteriorate both self-perceived health (-2.23%)

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
OLS	0.002	-0.016	0.355*	-0.022*	0.061***	-0.050	0.030	0.010	0.004
	(0.006)	(0.016)	(0.210)	(0.012)	(0.012)	(0.106)	(0.031)	(0.015)	(0.005)
GRF	-0.008	-0.052***	0.346	-0.040***	0.053**	-0.342**	-0.001	0.059**	-0.019**
	(0.010)	(0.020)	(0.311)	(0.014)	(0.021)	(0.170)	(0.051)	(0.024)	(0.009)
2SLS	-0.059***	-0.151**	-0.182	-0.100**	0.089**	0.206	0.034	0.042	-0.010
	(0.022)	(0.068)	(0.923)	(0.048)	(0.037)	(0.308)	(0.096)	(0.040)	(0.015)
R-squared OLS	0.154	0.214	0.137	0.170	0.119	0.633	0.126	0.046	0.049
R-squared IV	0.153	0.190	0.146	0.153	0.112	0.603	0.131	0.092	0.072
First-Stage F-Statistic	87.47	42.44	48.02	43.67	123.1	121.45	99.18	123.15	123.1
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 5: Average partial effects estimated by OLS, 2SLS, and GRF. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate experienced at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender, born in a rural area, and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%. For 2SLS, the instrument is the unemployment rate at the age defined by the compulsory minimum school-leaving age in the region of effective residence at that given age. Observations refer to the ones used for GRF estimation.

and objective grip strength (almost -1%)[19]. No significant average effects can be reported for depressive symptoms. Furthermore, worse initial conditions seem to reduce the probability of divorcing or never marrying (-1.9%). Again, bad worse initial conditions increase also the average number of children (0.059). As a possible explanation, the worse initial conditions may provide additional time for individuals to focus on social and family dimensions. Or again, women may lose empowerment in case of worse initial conditions. This last hypothesis is investigated in section 7.2.

Notably, GRF and OLS do not always provide accordant results. In detail, OLS estimates provide similar insights about "Hourly earnings" and "Self-perceived health". On the contrary, OLS does not detect any effect for "Earnings from work", "Objective grip strength", "Number of children", and "Never married or divorced". Both OLS and GRF identify the causal effects under the unconfoundedness assumption. However, the OLS model has to assume that the model is correctly specified. In section E.2, I perform a simulation to show that the OLS functional form misspecification may generate severe biases. On the contrary, GRF, owing to the flexible forest-based setting, is not affected when more complex data-generating functions are proposed. To observe the treatment effects from a different perspective, figure 11 depicts the density distribution of the CAPEs at individual level. In addition, the vertical dashed blue and black lines represent respectively the APEs of the Causal Forest and the standard OLS point estimates. These plots show that the OLS estimates tend to ignore the heterogeneity of the effects far from the

[19]-2.23% and -1% are computed by referring to the mean of "Self-perceived health" and "Grip strength" respectively.

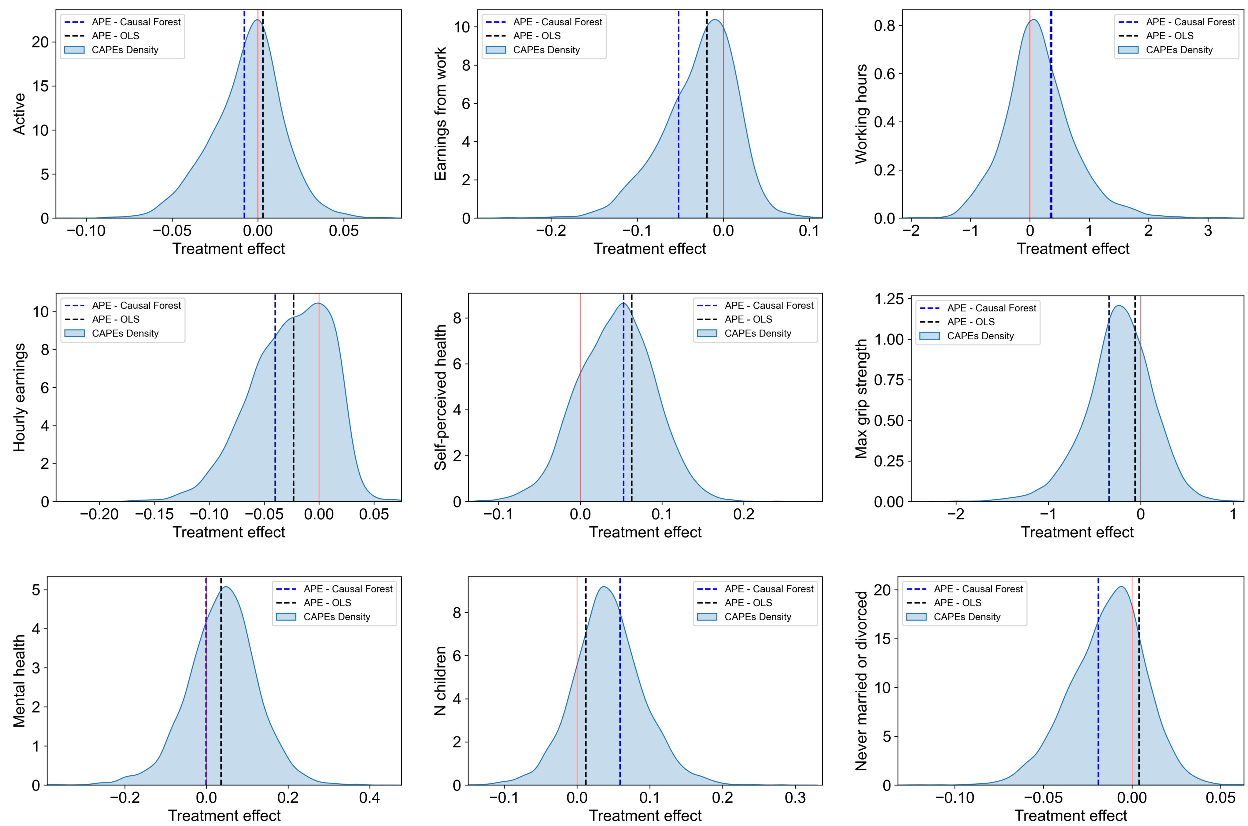


Figure 11: Distribution of CAPEs. The dashed blue vertical line represents the Average partial effect estimated by GRF, whereas the dashed black vertical line is the Average partial effect in the standard OLS.

mode of the distribution. On the contrary, the GRF uses the CAPEs to compute average partial effects which are more informed about the tails.

To provide further evidence about the robustness of the baseline results, I report the 2SLS estimations of the IV approach (table 5). 2SLS estimates show similar results with respect to the baseline GRF results. However, contrary to GRF, IV detects a negative effect on labour market participation (-5.9%) and no effect on grip strength. Again, although the point estimates of family outcomes are close to the GRF ones, the effects are not statistically significant owing also to larger standard errors. Moreover, the magnitude of the "Earnings from work", "Hourly earnings", and "Self-perceived health" coefficients tends to be larger. This is reassuring since if any omitted variable bias is present, the baseline effect would represent a more conservative bound. However, it is important to recall that, as discussed in section 4.1.2, the estimand of the IV approach refers to a particular subgroup of the sample and so, it is different with respect to GRF's estimand that targets the entire sample. Since the IV is estimating the local average treatment effect for the compliers[20] (individuals with low education), the IV estimand captures the

[20] F-statistics provide evidence that the instrument is much more relevant for those individuals with

effects on a subpopulation who might be more at risk of detrimental effects in the long-run. Further, IV estimates provide huge confidence intervals that contain almost always the entire GRF's confidence intervals. Finally, the simple linear specification might be an inappropriate functional form since it does not consider possible interaction terms and non-linearities which are automatically detected by GRF. In table 18, I provide evidence about different magnitudes of the effects according to alternative model specifications.

To sum up, table 5 reports that worse labour market conditions at the time of completing education caused a systematic divergence in life course trajectories. Europeans who graduated in the 1960-1990 time period earn less (roughly -5%) and suffer worse health. Results are robust across the different estimators.

6 Heterogeneity and non-linearity

One possible concern is that average effects could hide strong heterogeneity. Understanding which subgroups most suffered could help to better understand the underlying mechanisms and consequently devise optimal policy measures. The treatment effects at the individual level may be helpful to test and investigate in depth the heterogeneity and non-linearity of the effects.

6.1 Assessing heterogeneity

Following Athey and Wager (2019), I investigate the ability of the Causal Forests to uncover the heterogeneity in the treatment effect.

First, I use a heuristic approach to provide a qualitative understanding of the heterogeneity. The following steps are performed. First, the out-of-bag Conditional Average Partial Effect is estimated for each individual. Further, individuals are sorted above and below the 80th percentile of CAPEs. Next, for each subgroup is estimated the APE employing the double robust procedure. Finally, is performed a test of the difference in the APE of the two subgroups. The estimated 90% confidence intervals are displayed in table 6. This approach reveals significant heterogeneity only for "Hourly earnings" and "Never married or divorced". However, this procedure seems to be too simplistic to provide deep insights into heterogeneity. Hence, below are provided more refined procedures.

less than high school and high-school education. By removing from the sample college-educated people and running the same IV, the results remain robust with respect to the main IV analysis (see table 15).

	Active into labour market	Earnings from work	Working hours
90% CI	0.022 +/- 0.041	0.017 +/- 0.083	0.528 +/- 1.352
	Hourly earnings	Subjective health	Max grip strength
90% CI	0.041* +/- 0.040	-0.041 +/- 0.085	0.155 +/- 0.640
	Mental health	Number of children	Never married or divorced
90% CI	-0.075 +/- 0.196	0.075 +/- 0.110	0.048* +/- 0.031

Table 6: Heuristic approach to investigate heterogeneity.

The second method is motivated by the "best linear predictor" (Chernozhukov et al. (2018b)). The idea is to fit the CAPE as a linear function of the out-of-bag estimates $\hat{\tau}^{-i}(X_i)$. In detail, two predictors are constructed

$$C_i = \overline{\tau}(T_i - \hat{e}^{-i}(X_i))$$
$$D_i = (\hat{\tau}^{-i}(X_i) - \overline{\tau})(T_i - e^{-i}(X_i)),$$

where $\overline{\tau} = n^{-1} \sum_{i=1}^{n} \hat{\tau}^{-i}(X_i)$.

Then, $Y_i - \hat{m}^{-i}(X_i)$ is regressed against C_i and D_i[21] (equation 14).

$$Y_i - \hat{m}^{-i}(X_i) = \beta_C C_i + \beta_D D_i \tag{14}$$

A value of β_C close to 1 signals that the average prediction is correctly estimated. Similarly, D_i measures the quality of the calibration of the heterogeneity in the effects, i.e., $D_i = 1$ signals that the heterogeneity is well calibrated. Additionally, the p-value serves as an omnibus test for the existence of heterogeneity: if the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity[22]. To sum up, the best linear predictor models the outcome of interest as depending only on the average of the treatment effects and the deviation from this average. This test is possible only in the case CAPEs at individual levels are available. Although the test of this regression is not conclusive, it can provide insights into the ability of the estimator to provide reliable average and heterogeneity effects.

In table 7 are presented the point estimates and p-values. The first aspect to notice is the fact that β_C tends always to be close to 1, and so, the mean forest predictions seem to be correct. The evidence is particularly strong for "Earnings from work", "Hourly earnings", "Self-perceived health", "Number of children", and "Never married or divorced" which reject the null hypothesis that $\beta_C = 0$ at less than 10% level of significance and

[21]This procedure is a built-in function ("*test_calibration*") provided by the GRF package in R.

[22]It is important to notice that the asymptotic results justifying such inference are not presently available in the literature Athey and Wager (2019).

the point estimates are close to 1. Next, the calibration test provides interesting insights into the quality of the estimated heterogeneity. First of all, "Hourly earnings" assumes positive values and is statistically significant at the 5% level. Next, the "Self-perceived health" β_D coefficient assumes a positive value and is at the boundary of statistical significance. Finally, the Causal Forest can capture relevant heterogeneity for family outcomes, especially for "Never married or divorced". To sum up, the omnibus test confirms that the estimated CAPEs can adequately capture the average treatment effects for the majority of the outcomes of interest. Furthermore, this agnostic test can detect for which outcomes the underlying heterogeneity is most prominent.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
β_C	0.886	0.734*	-0.348	0.871**	0.964***	0.499	2.547*	0.880**	0.995**
p-value	0.235	0.068	0.617	0.022	0.001	0.177	0.061	0.023	0.024
β_D	0.226	0.337	-0.127	0.684**	0.418	−0.231	-0.325	0.110	0.988***
p-value	0.252	0.221	0.623	0.025	0.122	0.732	0.750	0.393	0.001

Table 7: Calibration test to assess the quality of the estimates of the treatment heterogeneity.

However, the above-mentioned heterogeneity tests are not conclusive in determining the presence of heterogeneity. As a matter of fact, the naive heuristic approach and the above omnibus tests are agnostic regarding the heterogeneity to spot. If any pre-determined theory regarding treatment heterogeneity is available, it can be used to increase the power of the tests focusing on specific variables of interest.

6.1.1 Detect heterogeneity

To investigate the heterogeneity in treatment effects of "Hourly earnings", individuals were ordered based on their partial effects, and subsequently, they were grouped into five quintiles. The resulting heatmap, depicted in figure 12, provides a visual representation of the average covariate values within each quintile. This visualisation serves the purpose of offering a clearer perspective on the heterogeneity observed in the treatment effect. Specifically, the visualisation helps to understand which theories are associated with significant variations in treatment effects across different groups. This analysis not only enhances the understanding of the factors that influence treatment outcomes but also opens routes for the discovery of new explaining channels. By identifying patterns and trends within the quintiles, I can uncover novel insights and potentially develop innovative theories or interventions that can further improve the efficacy of treatments.

The initial two rows of the heatmap clearly exhibit a discernible educational gradient. The average values of the covariates related to education level and age at completion of

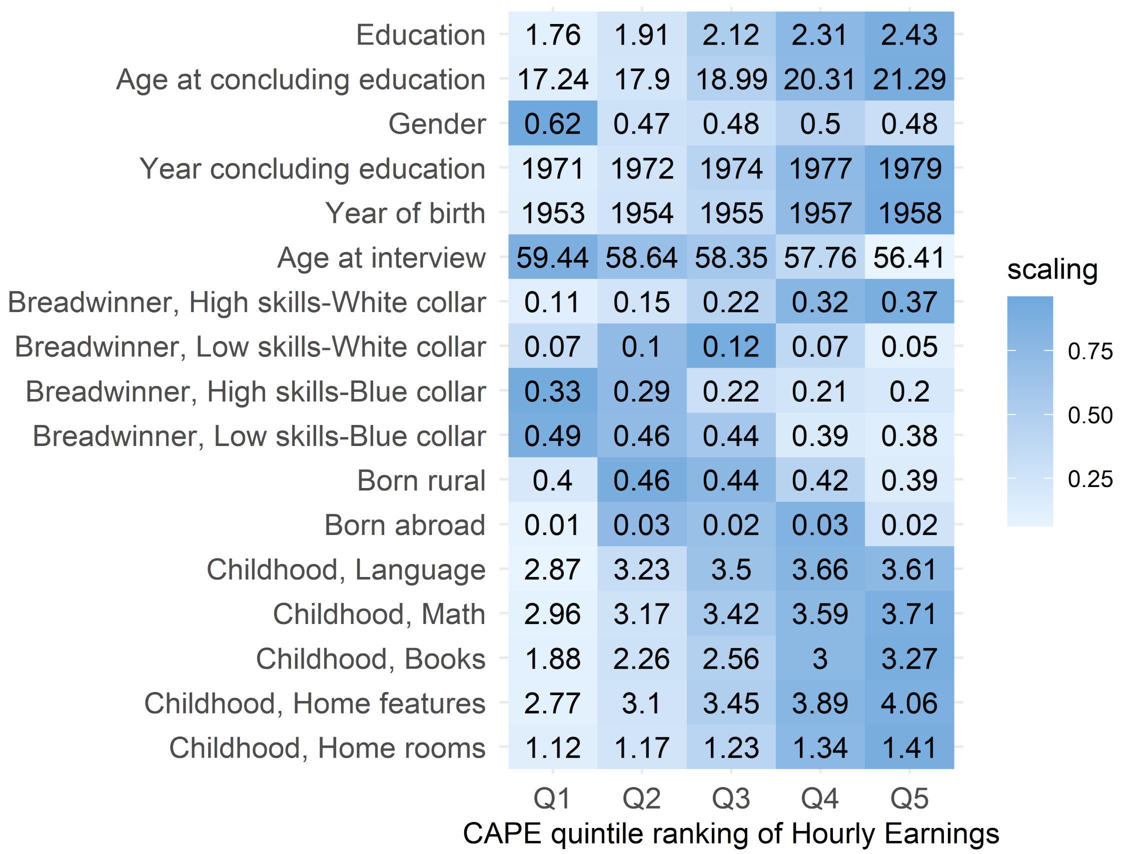


Figure 12: Average covariate values within quintile groups (based on CAPE estimate ranking) of "Hourly earnings". The scaling is achieved through a standardisation process where the proportion of a normal population is less than or equal to any given quantile.

education notably increase as we progress through the quintiles. On average, individuals in the first two (worse) quintiles have less than a high school diploma[23] and conclude education before age 18. Next, it is noteworthy that men exhibit a higher susceptibility to experiencing the most extreme long-term effects, as more than 60% of individuals in the worst quintile belong to the male gender. This finding suggests a gender disparity in the long-term "Hourly earnings" outcome, indicating that men are disproportionately affected by early adverse circumstances. Additionally, the cohorts who completed their education in the early 1970s are particularly vulnerable to experiencing adverse effects on earnings. Furthermore, it is evident that individuals in the worst quintiles also had blue-collar breadwinners during their childhood. However, being born abroad or in a rural area does not show a clear effect on long-term earnings outcomes. Conversely, indicators of worse childhood socioeconomic status are strongly associated with more adverse long-term effects on earnings. This highlights the significant impact of early-life socioeconomic conditions on the economic trajectories of individuals, underscoring the

[23]Education assumes values 1, 2, and 3 respectively for less than high school diploma, high school diploma, and university degree.

importance of controlling for childhood conditions.

In summary, the analysis indicates that the individuals most affected in terms of hourly earnings are typically less educated men who concluded their education in the early 1970s. Moreover, this group tends to have had childhood breadwinners engaged in blue-collar occupations and lower childhood socioeconomic status, indicating a potential intergenerational occupational influence.

To further gain insights about the underlying heterogeneity, I run a t-test[24] to determine whether there is a significant difference between the means of doubly robust treatment effects of subgroups of interest (educational attainment, gender,..). The null hypothesis is that there is no significant difference between the means of the two groups, against the true difference in means is not equal to 0. In table 8 are reported the p-values of the tests.

		Low vs Mid educ	Low vs High educ	Mid vs high educ	Men vs Women
Active	p-value	0.208	0.009***	0.111	0.006***
	t-test	(-1.258)	(-2.627)	(-1.595)	(2.73)
Earnings from work	p-value	0.808	0.104	0.03**	0.904
	t-test	(0.243)	(-1.624)	(-2.167)	(0.121)
Working hours	p-value	0.067*	0.001***	0.043**	0.843
	t-test	(1.831)	(3.48)	(2.027)	(-0.198)
Hourly earnings	p-value	0.906	0.01**	0.004***	0.682
	t-test	(-0.118)	(-2.577)	(-2.89)	(-0.41)
Self-perceived health	p-value	0.868	0.426	0.476	0.188
	t-test	(0.166)	(0.797)	(0.712)	(-1.315)
Objective grip strength	p-value	0.01**	0.177	0.166	0.424
	t-test	(-2.591)	(-1.351)	(1.387)	(-0.8)
Mental health	p-value	0.99	0.589	0.526	0.115
	t-test	(0.013)	(0.54)	(0.635)	(-1.576)
N children	p-value	0.141	0.707	0.042**	0.624
	t-test	(1.474)	(-0.376)	(-2.033)	(-0.491)
Never married or divorced	p-value	0.624	0.198	0.364	0.329
	t-test	(0.491)	(1.288)	(0.908)	(-0.975)

Table 8: T-tests to determine whether there is a significant difference between the means of doubly robust treatment effects of subgroups of interest. The null hypothesis is that there is no significant difference between the means of the two groups, against the true difference in means is not equal to 0. P-values from Welch Two Sample t-test.

The tests provide evidence that there is (almost) no statistical difference between the treatment effects of Low and Middle-educated individuals for labour market outcomes. On the contrary, evidence shows that the effects on Low and Middle-educated people differ from the High-educated, i.e., those with a university degree have better labour market outcomes with respect to the remaining subsample. Next, a relevant gender difference can be outlined: men have different effects on the probability of working. However, it is important to acknowledge that this approach has its limitations, as it only considers tests conducted on the entire subsamples, overlooking potential determinants of heterogeneity

[24]In detail, the Welch Two Sample t-test is run to avoid the assumption of equal variances of the two groups.

such as gender differences in "Hourly earnings." By not accounting for this heterogeneity, the analysis may fail to capture nuanced variations and distinct effects that could exist within subgroups.

6.2 Educational attainment

If education mitigates the long-term impact, then the intermediate outcomes of highly educated individuals may provide insights about which are the underlying mechanisms used to hedge from early shocks. To investigate heterogeneity according to educational attainment, I create three categories: less than high school diploma (Low), high school diploma (Middle), and more than high school diploma (High)[25].

Figure 13 displays the CAPEs density distributions of the long-term effects of experiencing worse initial labour market conditions according to educational attainment. Next, the CAPEs are used within the double robust estimator to recover the Group Average Treatment Effects (GAPEs) according to the achieved level of education. Table 9 presents the tabular results of GAPEs. The first insight from figure 13 is that the density distributions of highly-educated individuals tend to be always centred around the 0 effects. On the contrary, individuals with a high-school diploma or less than a high-school diploma (Low and Middle) are those who experienced larger long-term effects. Remarkably, the larger difference between educational attainment groups can be observed in terms of "Hourly earnings". As a matter of fact, the GAPEs shows that the effect is mainly driven by Low (-6.3%) and Middle-educated (-4%) individuals which are both statistically significant at the 10% level. Again, high-educated people starting with worse conditions do not suffer a statistical reduction in total and hourly earnings but work fewer hours[26]. Next, the long-term effects on "Self-perceived health" and "Mental health" seem to be entirely driven by Middle-educated people. The density distributions of the effects allow for observing a noteworthy peculiarity which would be more hidden in a traditional tabular average effect. The self-perceived and mental health measures for Low-educated have larger tails. As a consequence, the two tails signal further heterogeneity that should be investigated. One possible explanation (that is also investigated in section 7.2) claims that worse initial conditions forced some women to never enter the labour market. As a result, they might have avoided stressful working environments and so benefited in terms of long-term health (Maclean (2013)). The effect on the "Objective

[25]In detail, the assignment follows the International Standard Classification of Education (ISCED) classification. Low: ISCED $<= 2$; Middle: ISCED$= 3, 4$; HIGH: ISCED$= 5, 6$.

[26]Interpreting this reduction in working hours as a positive outcome can be questionable. However, the average working hours per week are respectively 34.85, 34.67, and 36.78 for respectively low, middle, and high-educated individuals. Being able to work fewer hours when they are already relatively high might signal an improvement in the personal life-work balance.

grip strength" is mainly driven by Low-educated people. Finally, family outcomes are statistically significant for Low-educated individuals who have a 3% lower probability of "Never married or divorced", whereas Middle-educated individuals have more children.

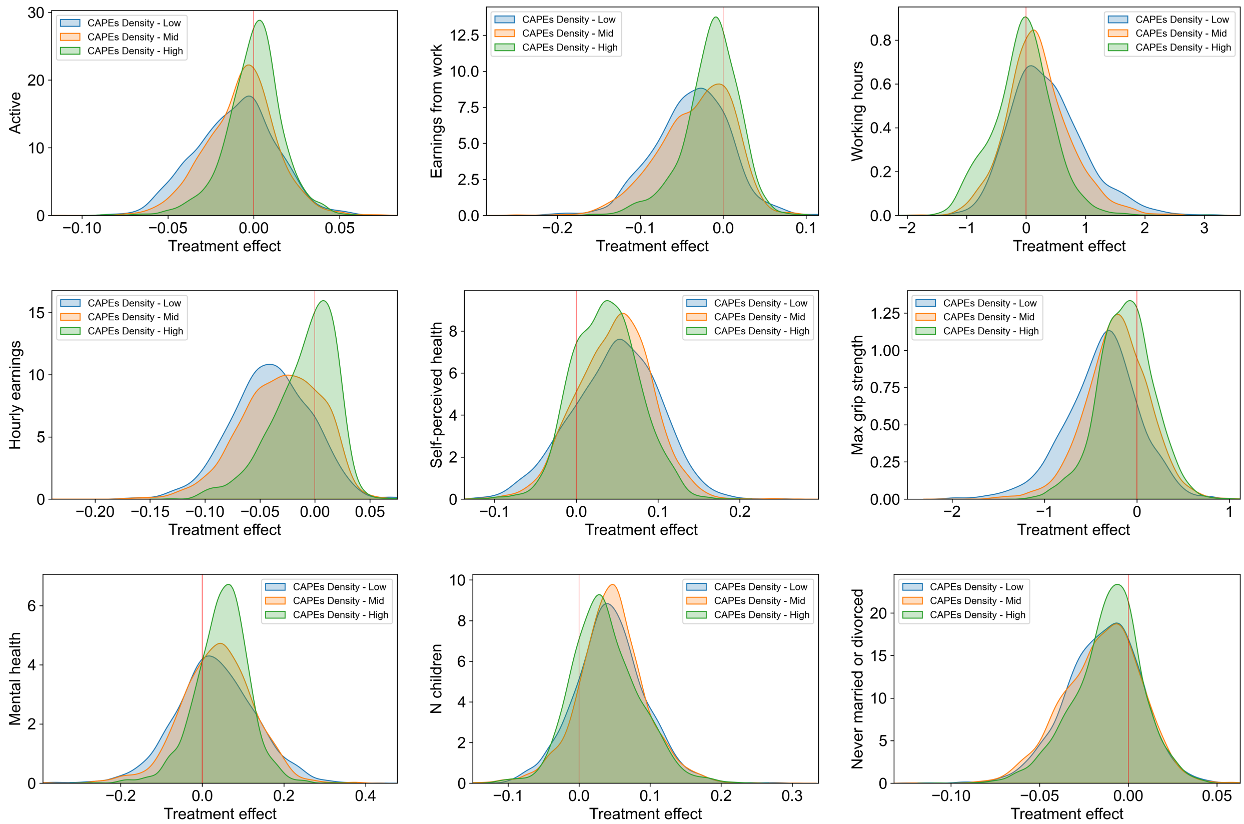


Figure 13: Distribution of CAPEs according to educational attainment. Low: less than a high school diploma; Mid: High school diploma; High: University degree.

Education	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Low	-0.014	-0.072	0.987	-0.063*	0.021	-1.119***	-0.187	0.033	-0.03*
	(0.02)	(0.046)	(0.701)	(0.036)	(0.044)	(0.338)	(0.125)	(0.038)	(0.017)
Middle	-0.009	-0.048*	0.54	-0.04*	0.089***	0.082	0.118*	0.077**	-0.014
	(0.013)	(0.029)	(0.424)	(0.021)	(0.026)	(0.244)	(0.061)	(0.035)	(0.011)
High	0.003	-0.036	-0.858**	-0.016	0.021	-0.127	0.014	0.058	-0.015
	(0.014)	(0.035)	(0.412)	(0.024)	(0.034)	(0.232)	(0.069)	(0.045)	(0.016)

Table 9: Average partial effects estimated by GRF capturing the heterogeneity by educational attainment. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate experienced at graduation. All estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender, born in a rural area, and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

To sum up, only individuals with less than a high-school diploma or those with a high-school diploma (Low and Middle) suffered long-term negative effects owing to worse initial conditions at the time of completing education. High-educated people are able to almost completely absorb the early shocks. A noteworthy implication is that the continuation of the schooling career in case of bad economic conditions may hedge from systematic deviations in life course trajectories.

6.3 Gender

Figure 14 and table 10 show respectively gender differences through the density distributions and the GAPEs (including educational attainment differences).

Figure 14 displays that women who started with worse initial conditions are less likely to be "Active in the labour market" (-2.8%) after more than 35 years after ending their education. Men, contrary to women, are hit in "Hourly earnings" (-5.1%) and "Hours worked" (0.817 hours). However, ignoring educational attainment within gender groups may hide important heterogeneity. Once educational attainment is considered, it can be pointed out that men's long-term effects are mainly driven by Low-educated males who receive almost 10% less "Hourly earnings" and at the same time are working more than 2 additional hours.

Noteworthy heterogeneity can be reported also for health outcomes. The comparison of "Self-perceived health" distributions does not show any clear gender heterogeneity. Although the point estimates for "Self-perceived health" are quite similar, the double robust estimator shows that the effect is statistically significant only for men. The interesting within gender heterogeneity is that "Self-perceived health" is not statistically significant for Low-educated women and for High-educated men. Next, the "Objective grip strength" is statistically different from 0 only for women. However, when focusing on educational attainment, it is clear that the effect is quite similar (roughly -1) and significant for both low-educated men and women. Further, table 10 shows that women with less than a high-school diploma (Low) that ended education in worse labour market conditions declare fewer depressive symptoms (approximately -16%)[27]. Again, a lower long-term labour market participation may have reduced stressful on-the-job duties (section 7.2).

Finally, long-term effects on family outcomes are quite similar between the 2 groups. The only difference is in the number of children which is driven by Middle-educated males.

Summing up, some gender differences in the labour market outcomes can be observed. The reason can be due to the fact that some women may be forced to never enter the labour market owing to worse initial conditions. University-educated men are never

[27]-16% is the proportion with respect to the mean of the "EURO-D depression scale", i.e., -0.346/2.10.

significantly affected in the long run.

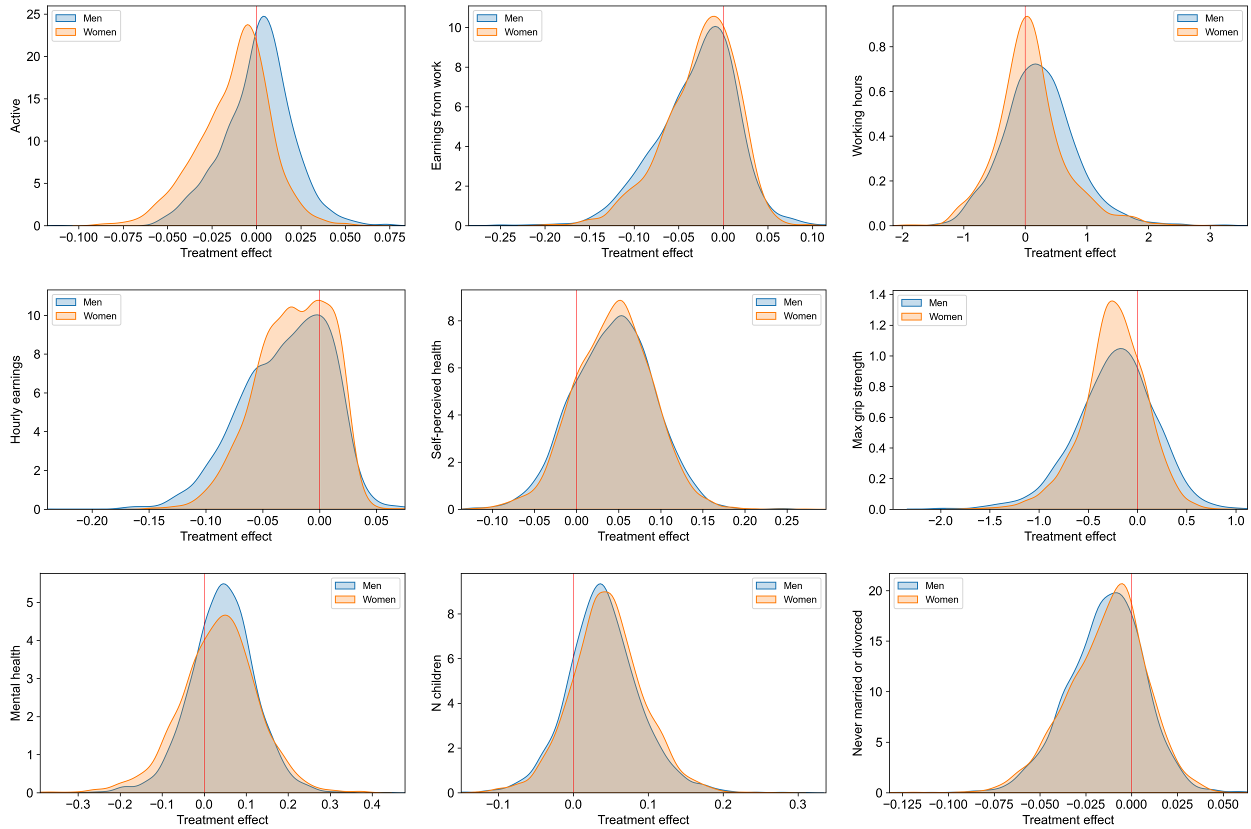


Figure 14: Distribution of CAPEs according to gender.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Women	-0.028**	-0.055*	-0.282	-0.026	0.039	-0.448**	-0.055	0.057**	-0.019
Women Low	-0.016	-0.082	-0.675	-0.009	-0.05	-1.073**	-0.346**	0.073	-0.039
Women Middle	-0.033*	-0.054	0.144	-0.045	0.081**	-0.12	0.119	0.044	-0.01
Women High	-0.034	-0.03	-0.782	-0.003	0.091**	-0.189	0.008	0.06	-0.004
Men	0.01	-0.049*	0.817**	-0.051**	0.066**	-0.239	0.054	0.061*	-0.02*
Men Low	-0.012	-0.065	2.028***	-0.096*	0.094*	-1.166**	-0.018	-0.008	-0.02
Men Middle	0.013	-0.044	0.854	-0.035	0.096**	0.277	0.116	0.109**	-0.018
Men High	0.035**	-0.041	-0.922	-0.027	-0.039	-0.074	0.018	0.056	-0.025
Women Obs	4063	2687	2882	2660	5495	5216	4332	5435	5495
Men Obs	3511	2863	3048	2834	4886	4662	3837	4824	4886

Table 10: Average partial effects estimated by GRF capturing the heterogeneity by gender and educational attainment. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate experienced at graduation. All estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender, born in a rural area, and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

6.4 Non-linearities

Reporting one single Average Partial Effect for each outcome as in table 5 may be insightful to provide the main idea of the results. However, assuming that the effect of worsening the initial conditions (by increasing the unemployment rate at the time of completing education) impacts long-term outcomes linearly may be too restrictive. The CAPEs at the individual level can discern the non-linear effect of completing education according to several dimensions.

6.4.1 Unemployment rate at ending education

Figure 15 shows the non-linear effect of increasing the unemployment rate at the time of completing education on the main outcomes under analysis. For the sake of interpretation, the effects are first ordered according to the dimension of interest, then are plotted the exponentially weighted moving averages (EWMA)[28]. Similarly, the 95% confidence intervals cover the variability of the EWMA within the same span. It is noteworthy to point out that the effects are not linear and follow different patterns according to the outcome of interest. However, one characteristic is recurrent, i.e., worsening early conditions is more detrimental at lower levels of unemployment rates. This is particularly true for "Active", "Earnings from work", "Hourly earnings", "Max grip strength", "N children", and "Never married or divorced". It is interesting to notice that the effect on "Self-perceived health" is the only one which can be broadly linearly approximated. Further, figure 16 shows the heterogeneity of the effects according to the country of completing education.

6.4.2 Other dimensions

Next, in figure 17, I show the non-linearity of the effects on "Hourly earnings"[29] for the age at the time of completing education and year of completing education. It is interesting to notice that the effect on "Hourly earnings" shows again that education hedges from early bad conditions. As a matter of fact, the impact of completing education during bad conditions is larger for individuals completing education at younger ages. The effect becomes gradually less negative up to approximately age 23, when the university degree is expected to be completed, then becomes zero. However, the effect is not statistically different from zero at around age 17. Next, the effects are strongly non-linear according

[28]The exponentially weighted moving average (EWMA) gives more weight to more recent values and allows to provide more interpretable figures.

[29]Only "Hourly earnings" are shown since it is the main outcome of interest and to avoid bulky graphical representations.

to the year of concluding education. Interviewees who concluded their education during the Seventies are also those who experience larger long-term detrimental effects.

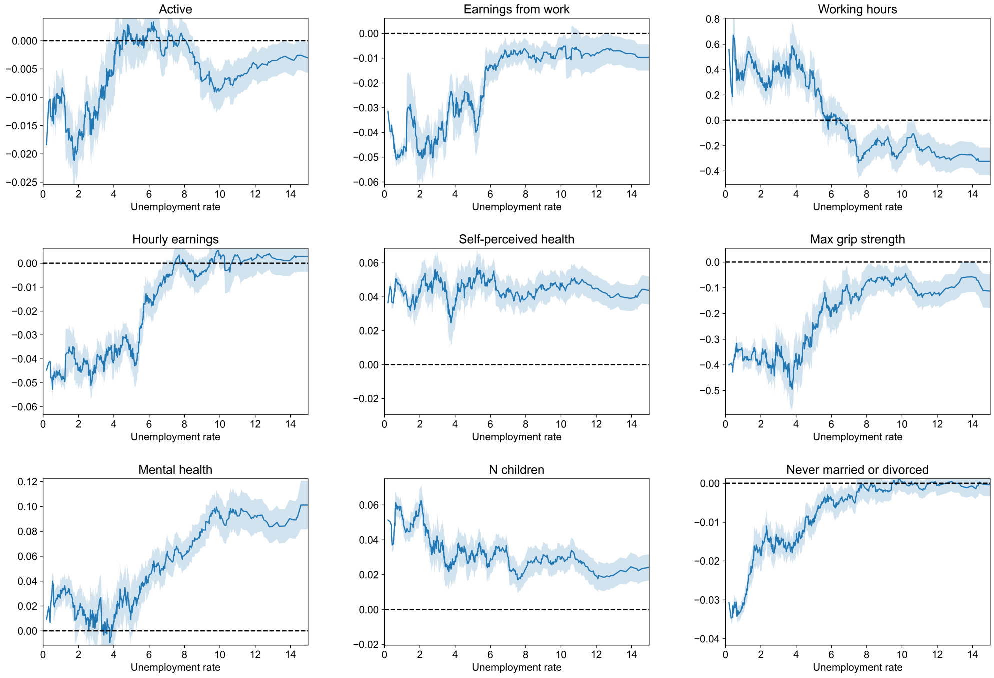


Figure 15: Non-linearity in worsening initial conditions. The blue line represents the exponentially weighted moving average (EWMA) with a span of 50 observations. 95% confidence interval.

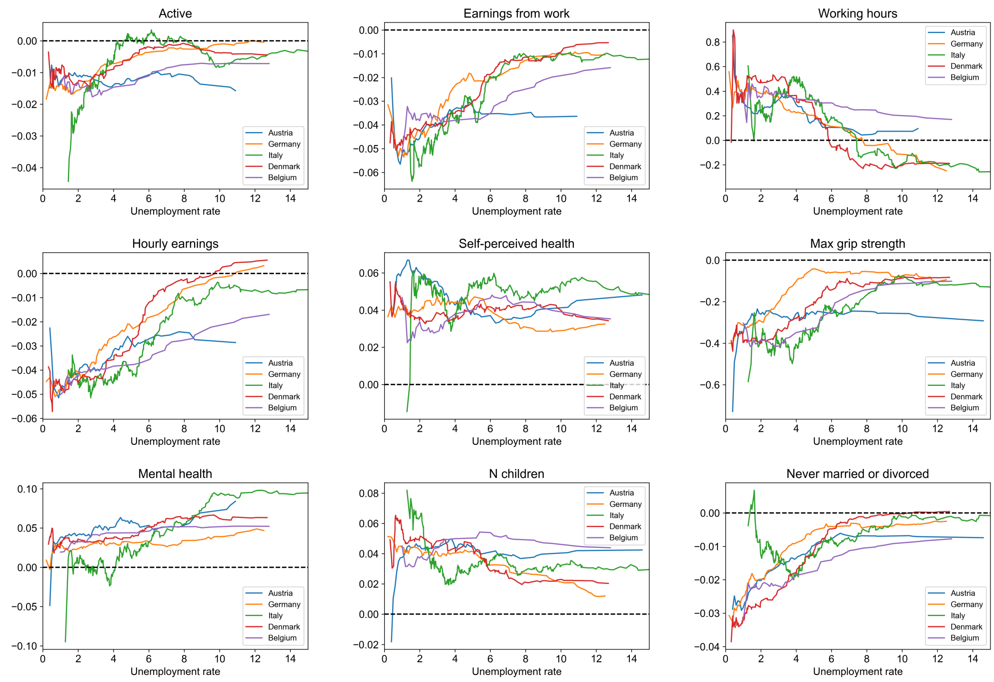


Figure 16: Non-linearity in worsening initial conditions according to country of completing education. The lines represent the exponentially weighted moving average (EWMA) with a span of 50 observations.

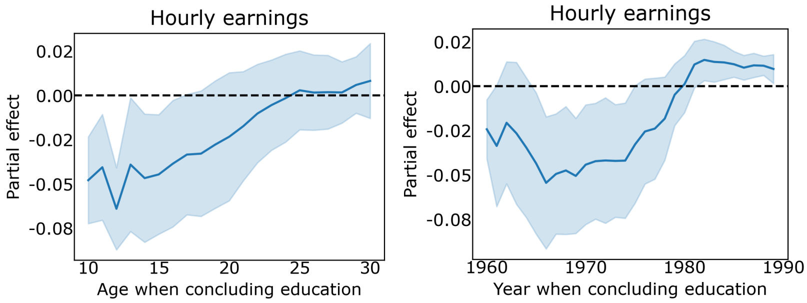


Figure 17: Non-linear effects regarding age at ending education and year of completing education for "Hourly earnings".

7 Theories explaining negative long-term outcomes of early labour market conditions

But why temporary shocks should permanently affect lifetime outcomes? The main channel under investigation relies on career development theories. There is substantial agreement on the relevance of initial opportunities on the job market in shaping long-term career success. Worse economic conditions at graduation are blamed to discourage labour market participation, increase unemployment, and induce job mismatching or under-employment. In this regard, recent evidence shows that during an economic downturn, the quality of vacancies shrinks and the likelihood to find a better job drastically declines (Moscarini and Postel-Vinay (2016), Haltiwanger et al. (2018)). The below-mentioned theories suggest that an initial relative disadvantage may cause a systematic divergence in life course trajectories.

The first stream of literature worthy to mention is the theory of human capital accumulation (Becker (1962)). In detail, during an economic downturn, younger graduates remain out of the labour force or become unemployed. As a consequence, they are not actively investing in both general or industry/firm-specific human capital. Again, in case of fewer opportunities available, young graduates might accumulate wrong/worse industry-specific skills if implicitly forced to accept a downgrading occupation position. Hence, no or wrong initial job matches shrink the productivity of the graduates who may permanently suffer a downgrade in career development.

Another relevant line of research deals with job search models (Topel and Ward (1992)). In brief, independent of labour market conditions, young workers search and switch jobs until the "correct" one is achieved. Hence, in case of economic downturns at the time of labour market entry, the initial gap is filled by changing jobs in search of better positions. In case of no frictions, workers would continuously search for better jobs and so no long-term effect on their career would be expected. However, empirical evidence shows that the cost of searching increases with age and job tenure (Oreopoulos et al. (2006)). Additionally, it is expected that highly educated workers may be more prone to migrate between regions and industries in search of better-paying jobs.

Human capital accumulation and search theories are not orthogonal and usually talk one with each other. On the one hand, mobility costs may increase with job tenure due to higher firm-specific human capital accumulation. On the other hand, the optimal job could be achieved by frequent job changes, but investing in specific human capital might not be worthwhile since the match happened too late and the expected net benefits are too low (Becker (1967)).

Furthermore, Lange (2007) points out that in an assortative matching model, persistent effects on job quality may be due to the differential speed of learning. So, low-skilled workers are doomed to perform undervalued tasks for a longer period of time. Next, Devereux (2002) presents a stigma model in which in case of imperfect information, employers perceive the unemployment condition and low wages as a signal of lower productivity. Hence, at the early stages of a career, exogenous economic downturns may have long-lasting effects through an unlucky signal. Besides, also institutional context matters. Harris and Holmstrom (1982) predict persistent downgraded job conditions if workers have little or no power to renegotiate the initial working conditions. Furthermore, the availability of top job positions is positively related to business cycles (Okun et al. (1973)).

To sum up, the above-mentioned theories agree that initial job placement has a crucial role in shaping long-term career development. A possible explaining mechanism relies on the quantity and quality of human capital accumulated during the early stages of a career. Catching up could take place with job mobility. Additionally, initial unfavourable conditions may become permanent also in case of stigma or institutional frictions. In the end, all theories agree that higher educated workers can adapt better to initial adverse conditions.

7.1 Testing hypothesis

In this section, I empirically test the above-mentioned theories using relevant intermediate labour market outcomes.

Even before the first job placement, unfavourable labour market conditions at the time of completing education may impact career development by discouraging labour market participation and fostering unemployment. In table 11a-b, I analyse whether worse early labour market conditions reduce the probability of finding at least one job in one and three years after completing education. Results show that for a point increase in the unemployment rate at the time of completing education, the probability of finding at least one job falls by 3% and 2.2% in one and three years. Table 11c-d shows that conditional on finding at least one job, the number of jobs declines in case of adverse early conditions (-0.019 and -0.031 respectively in the first 3 and 5 years). It is interesting to notice that the results are mostly driven by middle-educated people. Lower labour market attachment at the early stages of the careers reduces the accumulation of general and firm-specific human capital. Furthermore, since these individuals face fewer available job positions, they are not able to catch up by searching for better employers (Oreopoulos et al. (2012)). One possible response to worse starting conditions is moving between industry and occupations[30]. Table 11e-f shows that early economic conditions do not change industry and occupation mobility in the three years after ending full-time education. Finally, I test whether the earnings of the first job are impacted by the unemployment faced at the time of completing education. Table 11g shows insignificant effects. Nevertheless, the point estimates are negative and show a clear educational gradient (-1.45, -0.973, and -0.246 for Low, Middle, and High-educated). However, due to the restricted sample who reported this information, the statistical power is likely to be very low. Next, by employing the rich migration history of individuals, I check whether worse economic conditions at the time of completing education affect migrations. Evidence in 11h shows that individuals do not significantly move between regions in search of better labour markets. Surprisingly, the effect for university-educated people (High) is statistically significant and negative (-0.024).

To sum up, individuals who experience more adverse labour market conditions in the early stages of their careers have a lower probability to find a job and to move between different jobs. Furthermore, there is (tiny) evidence that they start with a lower wage and have fewer opportunities to catch up by moving between regions, job positions and occupations in search of better jobs. The effects are statistically significant mainly for individuals with high-school diplomas (Middle). Overall, the early labour market outcomes signal that individuals who started with worse early labour market outcomes are likely to accumulate less or the wrong human capital. Furthermore, the catch-up by searching for better employers is restrained by the adverse shock to the demand for workers. Additionally, early adverse conditions may allocate workers toward downgraded working conditions and generate stigma regarding their productivity through no fault of their own. Then, the downward spiral may be continued by employers who learn only slowly about the productivity of lower-skilled workers. Eventually, renegotiation starts from an underrated baseline. To sum up, the starting unfavourable conditions may be cumulatively disadvantaged by all the above-mentioned additional factors.

To corroborate the hypothesis that early unfavourable labour market conditions permanently affect the lifetime trajectory, I estimate the effect of initial labour market

[30]For industry NACE levels are used, whereas for occupation the 1 digit ISCO code.

conditions on the quality of the main job[31]. In detail, questions regard effort, demand, control and job circumstances[32]. Respondents convey their agreement on a 4-point scale from "1 Strongly disagree" to " 4 Strongly agree". Table 11 shows the results (positive signs refer to larger agreement).

	(a) Job immediately	(b) Job in three years	(c) Number of jobs in first three years	(d) Number of jobs in first five years	(e) Industry mobility	(f) Occupation mobility	(g) First wage
All	-0.03***	-0.022***	-0.019*	-0.031*	-0.002	-0.009	-0.888
Low	-0.022	-0.014	-0.024	-0.027	-0.014	-0.02	-1.425
Middle	-0.045***	-0.04***	-0.026	-0.051**	-0.001	-0.006	-0.973
High	-0.008	0.009	0	0.008	0.01	-0.001	-0.246
Obs All	10381	10381	9025	9382	9025	9025	4859
	(h) Migrated in first three years	(i) Physically demanding	(j) Environment uncomfortable	(k) Time pressure	(l) Emotionally demanding	(m) Recurrent conflicts	(n) Little freedom to decide
All	0	0.04*	0.018	0.002	0.043**	0.033*	0.013
Low	0.009	0.015	-0.01	-0.028	-0.016	-0.012	-0.047
Middle	0.005	0.078***	0.035	0.01	0.088***	0.06**	0.044*
High	-0.024***	-0.012	0.018	0.027	0.024	0.038	0.027
Obs All	9025	9268	9268	9269	9266	9266	9265
	(o) Opportunities to develop skills	(p) Received recognition	(q) Adequate salary	(r) Received adequate support	(s) Good atmosphere with colleagues	(t) Employees treated fairly	(u) State protected health hazards
All	-0.042**	-0.043***	-0.002	-0.044***	-0.007	-0.042***	-0.009
Low	-0.054	-0.034	-0.01	-0.022	0.017	-0.027	-0.025
Middle	-0.047*	-0.058**	0.011	-0.067***	-0.018	-0.041**	-0.025
High	-0.017	-0.021	-0.023	-0.025	-0.016	-0.064**	0.05**
Obs All	9263	9232	9248	9219	9142	8226	9080

Table 11: Average partial effects estimated by GRF capturing the heterogeneity by educational attainment. The dependent variables are the explaining mechanism outcomes. The treatment variable is the unemployment rate experienced at graduation. All estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender, born in a rural area, and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

In detail, workers who faced higher unemployment rates at the time of completing education experienced more physically and emotionally demanding workplaces with recurrent conflicts. The effects are mainly driven by individuals with high-school diplomas (Middle). Furthermore, individuals starting their careers with worse initial labour market conditions report also an inappropriate working environment for the accumulation of human capital. The effects on the opportunities to develop skills are statistically significant only for Middle-educated individuals, although the point estimates are larger for Low.

[31]The main job refers to the current job position for working individuals (since are 50+ it is likely that the current position corresponds to the main one) and the main reported one for not active individuals. Respondents are asked to convey their agreement on a 4-point scale.

[32]The questions are asked in the following way: "Work was physically demanding?", "Work had heavy time pressure?", "Work had adequate salary?", "Work employees treated fair?"

The worse placement is evident also from the measures of the received "recognition", "support" and the "fairness" in the work environment. It is interesting to notice that results are mainly driven by Middle-educated workers.

To sum up, evidence shows that worse labour market conditions at graduation negatively impact the starting conditions of unlucky graduates who stick to jobs of inferior quality. As a result, unlucky individuals suffer a systematic divergence in life course trajectories outcomes.

7.2 Intermediate outcomes: gender analysis

Do women respond differently in the first stages of their careers to bad initial conditions with respect to men? The previous sections provided evidence that women are less active in the labour market after more than 35 years after ending their education. In this section, I aim to investigate whether poor initial conditions are causing women to leave the labour market and engage in housekeeping activities instead. Table 12 shows that early worse conditions reduce the likelihood of women entering the labour market (-4% and -5.3% in the first and first three years respectively). Next, worse initial conditions, reduce the number of jobs women can find and reduce the probability to migrate towards other regions. Again, worse early conditions reduce the likelihood of ever entering into the labour market (-2.1%) and the total number of years spent in the labour market during the entire career (-1.36 years)[33]. To sum up, worse initial conditions permanently reduce the labour market attachment of women. As a result, only a selected subgroup of women enter and remain in the labour market. This can be one reason why women are less affected in the labour market outcomes (hours worked and hourly earnings) if compared with men. Additionally, table 12 highlights that men suffer less in terms of achieving jobs but are forced to accept jobs that are worse off, i.e., significantly lower first wage (for uneducated individuals the effect is -3.659 percentiles). Again, the educational gradient is always important.

	Job immediately	Job in three years	Number of jobs in first five years	Migrated first three years	First wage	Ever worked	Total years worked
Women	-0.041***	-0.053***	-0.04*	-0.01**	-0.161	-0.021***	-1.362***
Women Low	-0.055**	-0.077***	-0.025	-0.01	1.034	-0.054***	-2.086**
Women Middle	-0.045***	-0.055***	-0.056*	-0.001	-0.61	-0.009	-1.127***
Women High	-0.009	-0.004	-0.023	-0.03**	-0.121	0.004	-0.655
Men	-0.019	0.009	-0.023	0.008	-1.547	0	-0.187
Men Low	0.013	0.052**	-0.028	0.025	-3.659*	-0.005	-0.264
Men Middle	-0.046***	-0.025**	-0.046	0.011	-1.316	0.002	-0.130
Men High	-0.007	0.02*	0.032	-0.018**	-0.348	0.002	-0.199

Table 12: Selected intermediate labour market outcomes according to gender.

[33] Section C.1 in the appendix provides a more detailed heterogeneity analysis for total years worked.

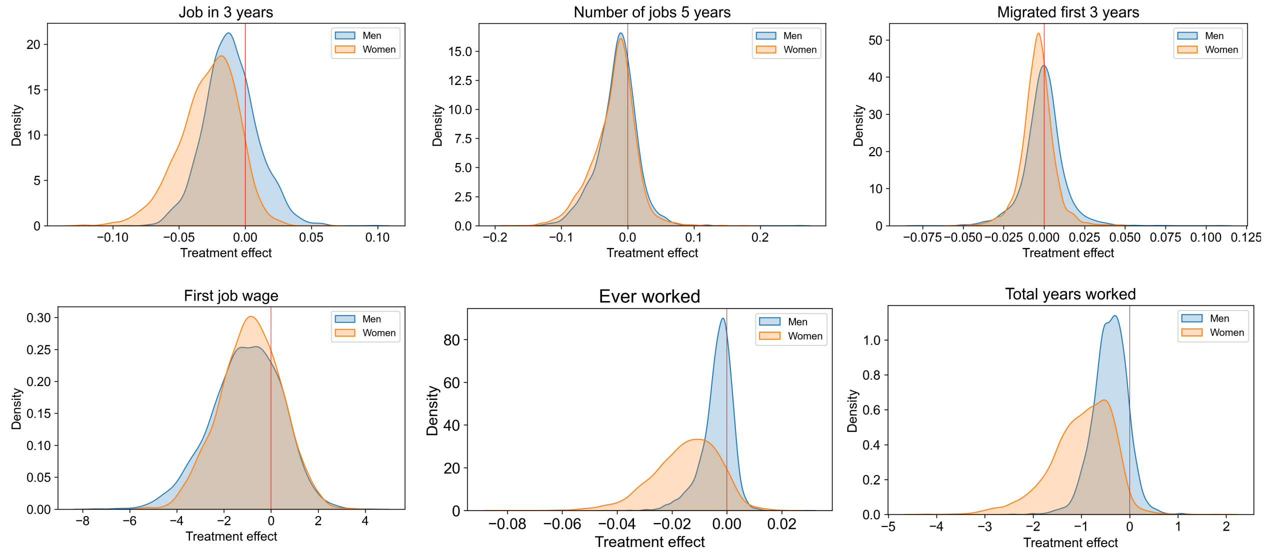


Figure 18: Probability density distribution of selected intermediate labour market outcomes according to gender.

8 Validity and robustness checks

It is worth mentioning that the main results remain robust to different identification strategies and estimation procedures. Furthermore, the estimated effects are supported by intermediate labour market channels that explain the negative old-age impacts. To further support the validity of the findings, I conduct a series of robustness checks.

Placebo test

First, I conduct a placebo test to show that assigning a treatment, which theoretically should not impact the outcome, does not yield significant relationships. This helps eliminate the possibility that the main results are influenced by spurious correlation. In detail, in section D.1 I assign random unemployment rates[34], drawn from a similar distribution as the actual data, to each year-region of completion of education. Specifically, the results in figure 28 provide evidence about the absence of an effect in instances where none is expected.

Analysis by gender

Next, to provide evidence that the gender results are not driven by complex patterns, inequalities, and structurally different labour market careers that might remain hidden

[34]Different distributions have been used: Chi-squared distribution with different degrees of freedom to mimic the unemployment rate distribution in the data and uniform distribution in the min-max interval observed in the data.

when examining data as a whole, I run the analysis separately by gender. Section D.2 shows that the results almost completely overlap with the main analysis.

Alternative treatments

A further concern is that the unemployment rate measured in the year of completing education captures only partially the labour market conditions in the early stages of the career. To control for this aspect, in section D.3 I run the analysis on different combinations of the unemployment rates in the years close to completing education. Precisely, the results remain robust if I make use of the mean of the first three years after completing education (the year before and after, etc.) as the treatment variable. Next, in almost all the literature the treatment variable is the level of the unemployment rate. To check the robustness of this measure, I additionally use the deviation from unemployment trends using Hodrick-Prescott filtering (table 25). Results remain robust.

Unemployment rates at country level

Next, in table 26 I show the effects of using the unemployment rates at the country level instead of the regional one. As expected, most of the estimates lose their statistical significance due to substantial attenuation bias caused by mismeasurement in early labour market conditions. This underscores the importance of employing unemployment rates at the most suitable level and emphasises the valuable contribution of recovering regional unemployment rates for the 1960-1990 period.

More about the childhood conditions

Following that, I use the Principal Component Analysis to extract the most important socio-economic and ability dimensions during childhood. By using the PCA instead of the original childhood conditions, the results remain robust (see table 27). Next, as discussed in section 4.1, the direction of the bias due to possible self-selection into timing or location of concluding education is not known a priori. Here, I investigate which is the direction of the bias in the case of the removal of the childhood controls. As it is clear from table 28, the effects are now larger in absolute values, meaning that the exclusion of the childhood conditions introduces now a negative bias for earnings (and a positive bias for self-perceived health). Hence, childhood conditions are essential to controlling for unobserved factors. Specifically, it seems that the direction of the bias is driven mostly by individuals migrating towards other regions rather than those deciding the timing of the conclusion of the educational pathway.

More about clustering

In continuation, since the level of clustering impacts the point estimates of the Causal Forest, I run the analysis without any clustering. Table 29 shows that the results are not driven by the clustering choice. Again, in table 30 I show that by assigning an equal weight in the forest for each cluster, the results remain robust.

Excluding specific countries from the analysis

Next, I undertake the analysis by systematically excluding one country at a time to show that there is no singular country driving the results (see table 20). Again, in table 21, I remove Austria and Germany to control if the "Germanic" educational system drives the findings and the results hold.

Robustness checks and simulations for Causal Forest

In this paper, I extensively exploited the latest advancement in the machine learning literature in the policy evaluation literature. However, since it is a relatively new methodology in applied work and requires more specific assessments, in section E, I provide further checks. In section E.1.1, I extensively discuss the hyperparameters tuning procedure. Next, I show that the flexibility of the Causal Forest can effectively address biases stemming from model misspecification in conventional OLS methods. Again, in section E.2.2, I perform several simulations to investigate the finite sample properties of the results. Further, in section E.2.3, I propose additional simulations to investigate the ability of the Causal Forest to capture the underlying heterogeneity.

9 Conclusion

This paper studies the very long-term effects of adverse labour market conditions experienced during a crucial stage of career development, i.e., the time of completing education. The focus is on European cohorts who completed education from 1960 to 1990, a time period containing the stagflationary environment during the 1970s. It is well known that adverse labour market conditions affect especially the most vulnerable groups, including young people (Elsby et al. (2016)). However, the existing literature does not offer a systematic approach to identifying the most vulnerable groups. Moreover, there is little evidence concerning the effects during their late adulthood and how they propagate during their entire life cycle.

To fill this gap, I estimate the effects of the exposure to local unemployment rates at the time of completing education on long-term labour market, health, and family outcomes measured more than 35 years after completing education in five European countries (Austria, Belgium, Denmark, Italy, and (West-)Germany from SHARE). Estimation is performed through Causal Forests (Athey et al. (2019)), a Causal Machine Learning method specifically designed for causal inference, which combines the predictive power of statistical learning with microeconometric approaches to identifying causal effects. By reducing the risk of model misspecification and focusing on discovering heterogeneity, the Causal Forest offers a more comprehensive understanding of the phenomenon under study. To further support the baseline results, a standard instrumental variable approach is employed, using exogenous unemployment rates based on year of birth and compulsory minimum school-leaving age laws as instruments. However, the local nature of the IV does not allow for fully exploring the heterogeneity in the treatment effects.

I find that individuals who complete their education during worse economic conditions experience permanent and economically significant effects. One percentage point increase in the unemployment rate at the time of completing education reduces earnings (approximately -5.2%) and health (grip strength (-1%) and self-perceived health (-2.23%)), even after more than 35 years since the completion of their educational pathway.

Importantly, effects are heterogeneous. University-educated individuals can almost completely protect themselves from adverse initial conditions. Men are relatively more affected in terms of hourly earnings (-5.1%), whereas women are more impacted in their labour market participation (-2.8%). The intermediate labour market outcomes provide clear explaining mechanisms. On the one hand, women face early challenges that result in persistent difficulties in labour market participation. On the other hand, men, although able to enter the labour market, often find themselves compelled to accept significantly lower-quality job positions. However, both genders tend to end up in main job roles characterised by undesirable features, including higher physical and emotional demands, increased conflicts, and reduced recognition. Additionally, the accumulation of valuable skills is hindered by work environments that do not facilitate skill development.

These results shed light on current policy issues regarding bad economic outlooks, and how these will impact the long-term careers of young graduates. From a policy perspective, mitigating the early labour market shocks could be extremely valuable for individuals who are more severely affected, i.e., the less-educated individuals. One way to deal with it may be through reducing university enrolment barriers. A positive effect would be observed only if a university education provides adequate skills or the value of the signalling of a university degree is large enough. Alternatively, in case of an extremely negative economic outlook, job availability could be fostered by tax relief schemes for hiring less-educated young people. Moreover, focusing on less-educated women may also help alleviate the long-term challenges they face in labour market participation.

References

J. D. Angrist, G. W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

S. Athey and G. Imbens. Machine learning methods economists should know about. *arXiv preprint arXiv:1903.10075*, 2019.

S. Athey and G. W. Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32, 2017.

S. Athey and S. Wager. Estimating treatment effects with causal forests: An application. *Observational Studies*, 5(2):37–51, 2019.

S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

A. Bald, E. Chyn, J. S. Hastings, and M. Machelett. The causal impact of removing children from abusive and neglectful homes. Technical report, National Bureau of Economic Research, 2019.

G. S. Becker. Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49, 1962.

G. S. Becker. Human capital and the personal distribution of income, woytinsky lecture no. 1. *Ann Arbor, Michigan: University of Michigan Press*, 1967.

P. Bingley and A. Martinello. Measurement error in the survey of health, ageing and retirement in europe: A validation study with administrative data for education level, income and employment. *Work Pap Ser*, 16:2014, 2014.

A. Börsch-Supan, M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber. Data resource profile: the survey of health, ageing and retirement in europe (share). *International journal of epidemiology*, 42(4):992–1001, 2013.

J. Bound. Self-reported vs. objective measures of health in retirement models, 1989.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

A. Brugiavini, C. E. Orso, M. G. Genie, R. Naci, and G. Pasini. Combining the retrospective interviews of wave 3 and wave 7: the third release of the share job episodes panel. *Munich Center for the Economics of Aging (MEA), Munich, SHARE Working Paper Series*, pages 36–2019, 2019.

B. Brunner and A. Kuhn. The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics*, 27(3):705–738, 2014.

A. Case, D. Lubotsky, and C. Paxson. Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5):1308–1334, 2002.

A. Case, A. Fertig, and C. Paxson. The lasting impact of childhood health and circumstance. *Journal of health economics*, 24(2):365–389, 2005.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters, 2018a.

V. Chernozhukov, M. Demirer, E. Duflo, and I. Fernandez-Val. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research, 2018b.

B. Cockx and C. Ghirelli. Scars of recessions in a rigid labor market. *Labour Economics*, 41:162–176, 2016.

J. Currie. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *JoUrnal of economIc lIteratUre*, 47(1):87–122, 2009.

D. M. Cutler, W. Huang, and A. Lleras-Muney. When does education matter? the protective effect of education for cohorts graduating in bad times. *Social Science & Medicine*, 127:63–73, 2015.

G. De Luca, M. Celidoni, and E. Trevisan. Item non response and imputation strategies in share wave 5. *SHARE Wave*, 5:85–100, 2015.

P. J. Devereux. The importance of obtaining a high-paying job. 2002.

T. A. DiPrete and G. M. Eirich. Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual review of sociology*, pages 271–297, 2006.

M. W. Elsby, D. Shin, and G. Solon. Wage adjustment in the great recession and other downturns: Evidence from the united states and great britain. *Journal of Labor Economics*, 34(S1):S249–S291, 2016.

J. Fan, M. Farmen, and I. Gijbels. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):591–608, 1998.

D. Fernández-Kranz and N. Rodríguez-Planas. The perfect storm: graduating during a recession in a segmented labor market. *ILR Review*, 71(2):492–524, 2018.

C. Garrouste and O. Paccagnella. Data quality: Three examples of consistency across share and sharelife data. *Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE Methodology. Mannheim Research Institute for the Economics of Ageing (MEA): Mannheim*, pages 62–72, 2011.

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

J. C. Haltiwanger, H. R. Hyatt, L. B. Kahn, and E. McEntarfer. Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics*, 10(2):52–85, 2018.

M. Harris and B. Holmstrom. A theory of wage dynamics. *The Review of Economic Studies*, 49(3):315–333, 1982.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

J. J. Heckman and E. Vytlacil. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738, 2005.

K. Hirano and G. W. Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

G. W. Imbens and J. D. Angrist. Identification and estimation of local average treatment effects. *Econometrica: journal of the Econometric Society*, pages 467–475, 1994.

J. Johannemann, V. Hadad, S. Athey, and S. Wager. Sufficient representations for categorical variables. *arXiv preprint arXiv:1908.09874*, 2019.

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

F. Lange. ‘the speed of employer learning. In *Journal of Labor Economics*. Citeseer, 2007.

K. Liu, K. G. Salvanes, and E. Ø. Sørensen. Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84:3–17, 2016.

J. C. Maclean. The health effects of leaving school in a bad economy. *Journal of Health Economics*, 32(5):951–964, 2013.

F. Mazzonna and E. Havari. Can we trust older people's statements on their childhood circumstances? evidence from sharelife. 2011.

R. Molloy, C. L. Smith, and A. Wozniak. Internal migration in the united states. *Journal of Economic perspectives*, 25(3):173–196, 2011.

G. Moscarini and F. Postel-Vinay. Did the job ladder fail after the great recession? *Journal of Labor Economics*, 34(S1):S55–S93, 2016.

B. R. Moulton. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The review of Economics and Statistics*, pages 334–338, 1990.

X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

A. M. Okun, W. Fellner, and A. Greenspan. Upward mobility in a high-pressure economy. *Brookings Papers on Economic Activity*, 1973(1):207–261, 1973.

P. Oreopoulos, T. M. von Wachter, and A. Heisz. The short-and long-term career effects of graduating in a recession: Hysteresis and heterogeneity in the market for college graduates, 2006.

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

M. Päällysaho. The short-and long-term effects of graduating during a recession: Evidence from finland. *VATT Institute for Economic Research Working Papers*, 96, 2017.

P. Probst, A.-L. Boulesteix, and B. Bischl. Tunability: Importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research*, 20(1): 1934–1965, 2019.

P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

H. Schwandt and T. Von Wachter. Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets. *Journal of Labor Economics*, 37(S1):S161–S198, 2019.

H. Schwandt and T. M. Von Wachter. Socioeconomic decline and death: Midlife impacts of graduating in a recession. Technical report, National Bureau of Economic Research, 2020.

R. H. Topel and M. P. Ward. Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479, 1992.

W. Van den Berge. Bad start, bad match? the early career effects of graduating in a recession for vocational and academic graduates. *Labour Economics*, 53:75–96, 2018.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Appendix

A Regional unemployment rates, by Country

A.1 Austria

Data about the regional unemployment rate in Austria for the years 1960-1990 is gathered from The Austrian Chamber of Commerce.

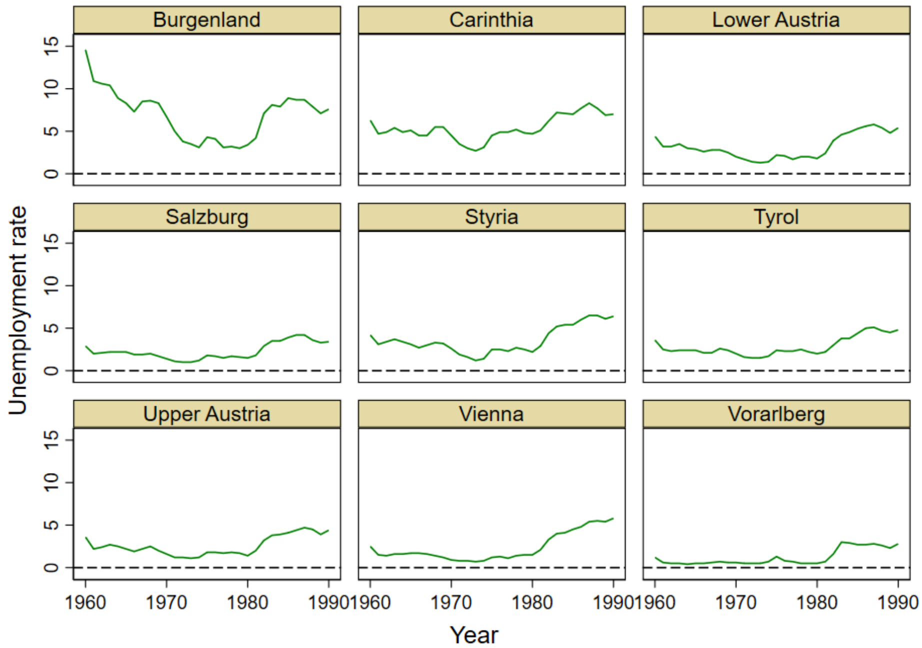


Figure 19: Unemployment rate at the regional level, Austria 1960-1990. Source: The Austrian Chamber of Commerce.

A.2 Belgium

Data for Belgium are retrieved from "Dossier Statistique de Population active d'emploi et de chomage" from 1960 to 1982.[35]

[35]Most of the sample is however graduating before 1982.

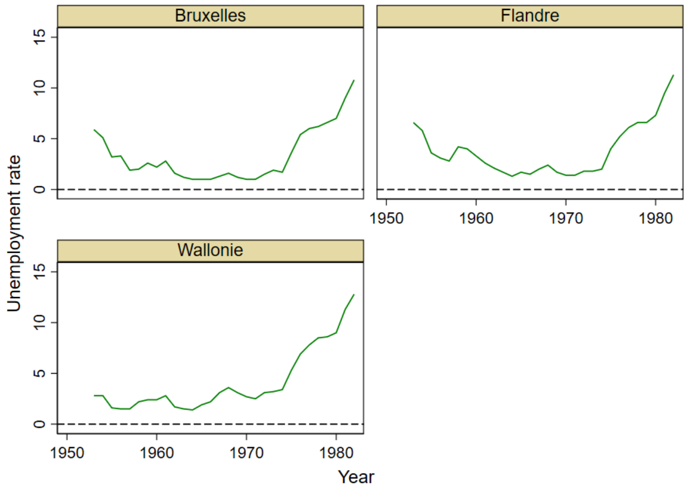


Figure 20: Unemployment rate at regional level, Belgium 1953-1982. Source: Dossier Statistique de Population active d'emploi et de chomage

A.3 Denmark

Municipal data about the unemployment rate are gathered from Statistical Yearbooks from 1961 to 1990. Then, municipal data are averaged to match the regional levels in SHARELIFE.

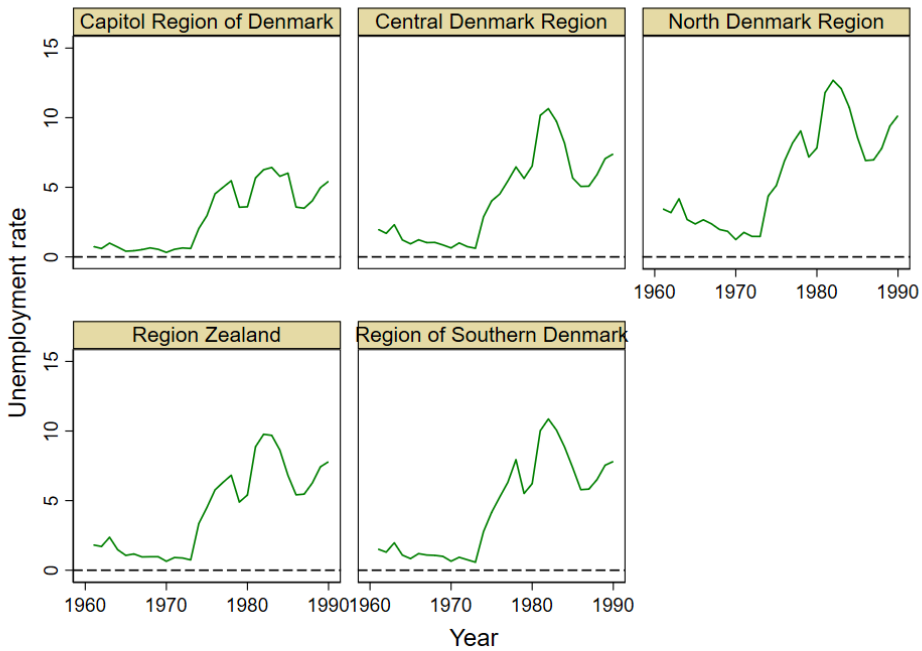


Figure 21: Unemployment rate at regional level, Denmark 1961-1990. Source: Statistical Yearbook. (By weighting for size of municipalities).

Two main issues arise when dealing with Denmark's unemployment rates. The first one is that unemployment rates are provided at different finer geographical levels and so the unemployment rates are computed using a weighted average by keeping into account the number of workers enrolled in the insurance companies. The underlying assumption is that the share of enrollment into unemployment insurance companies is similar across the municipalities. The second issue refers to the fact that workers who are enrolled in an insurance company may not be a random sample of the population. In fact, it is likely that those enrolled could be a self-selected group which is most at risk of unemployment. To test this, the weighted average of the regional unemployment rate is compared with the national data. As it can be observed in figure 22, regional measures tend to be slightly upward biased.

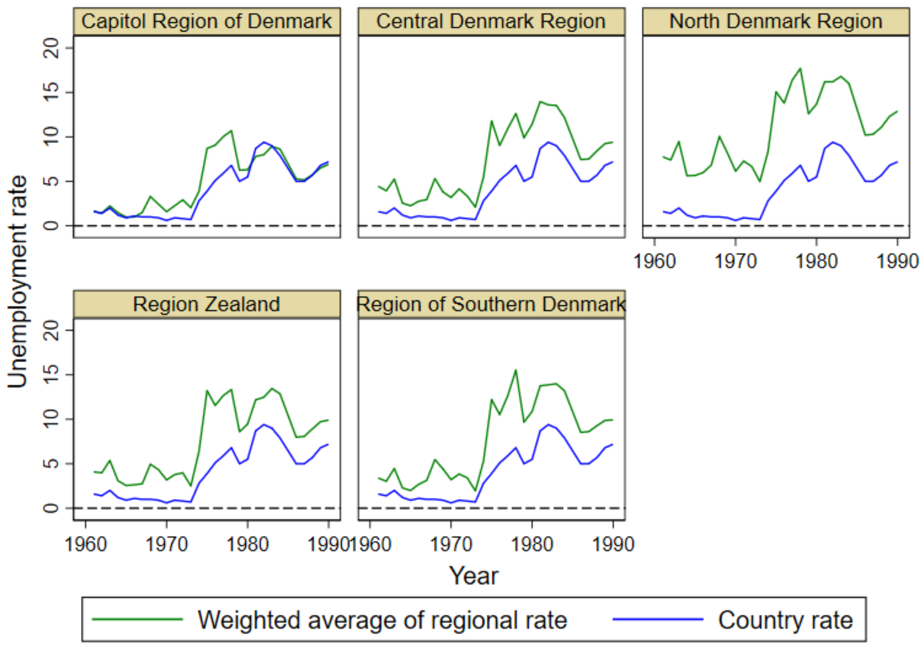


Figure 22: Unemployment rate at regional level, Denmark 1961-1990. (By weighting for size of municipalities).

To avoid this second problem, each regional unemployment rate is proportionally re-scaled to have an average unemployment rate at the regional level close to the country level[36].

[36]To re-scale each unemployment rate the following steps are proposed. First, for each year compute the weighted average (using 1985 "population size") of unemployment given the five regions. Then relate the country's unemployment rate gathered from AMECO with the one computed in the previous point. Finally, use this adjustment vector to shrink the regional level towards the country one.

A.4 (West-)Germany

Data are available at the Lander level from 1961 to 1986. Data retrieved from old publications of "Official news from the Federal Employment Agency" (Amtliche Nachrichten der Bundesanstalt für Arbeit). Some Landers data are presented together in the source for far away time periods. Schleswig-Holstein is aggregated with Hamburg, Lower Saxony is aggregated with Bremen and Rhineland-Palatinate is aggregated with Saarland from 1961 to 1976.

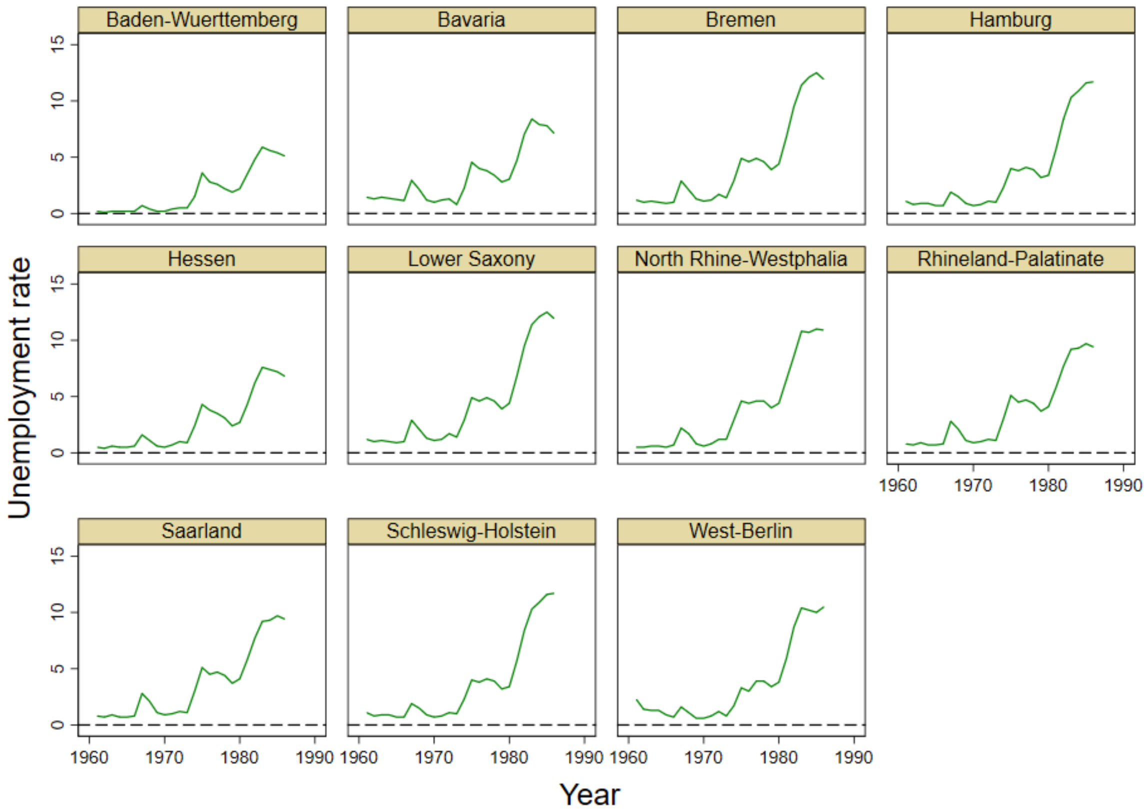


Figure 23: Unemployment rate for German Lander from 1961 to 1986. Source: Official news from the Federal Employment Agency.

A.5 Italy

Data about employed and unemployed in each region are retrieved for each year from 1963 to 1990 from the Statistical Yearbook (Annuario Statistico Italiano). In figure 24 are shown the unemployment rates experience in each Italian region from 1963 to 1990.

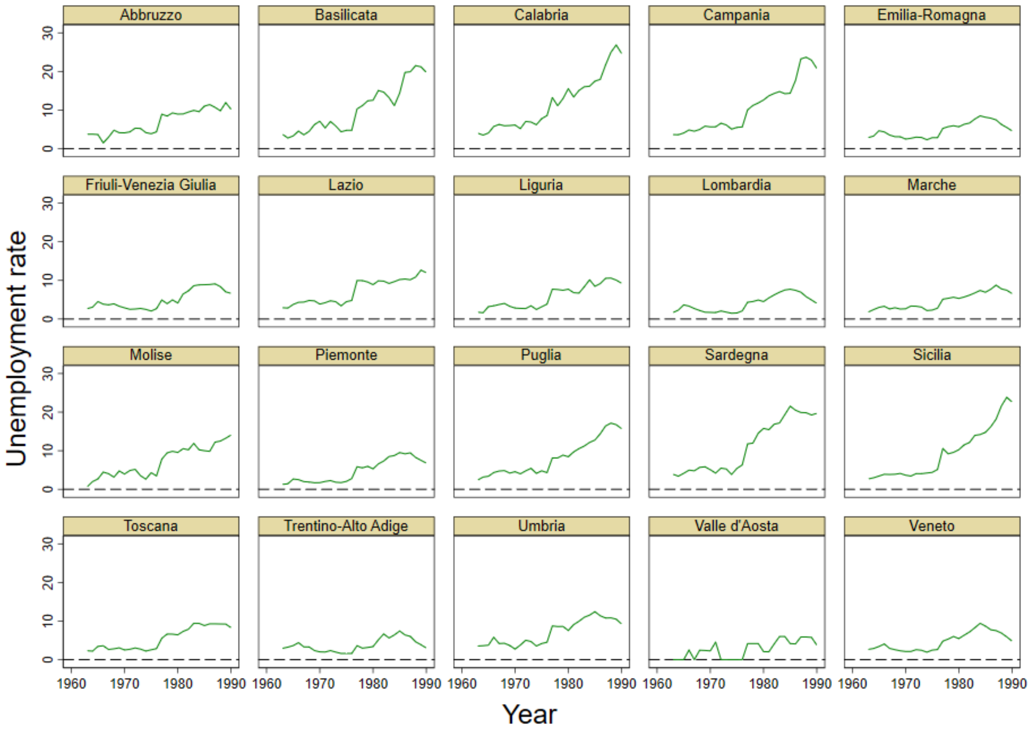


Figure 24: Unemployment rate at regional level, Italy 1963-1990. Source: Statistical Yearbooks.

A.6 Region of graduation FE

One noteworthy topic to discuss is the way the region of graduation effects are considered in the model. In my specific setting, the usual way to model the long-term outcomes of graduates from different regions is to acknowledge that each region might generate idiosyncratic effects that are not explained by other controls. This is usually done by encoding categorical variables as dummy vectors. However, this representation can be wasteful since it adds many low-signal regressors, especially when the number of unique categories is large Johannemann et al. (2019). The reason why can be found in the fact that unless the is strong evidence that those dummies are strong predictors, each dummy contribution is set to 0. As a consequence, regional dummies should be carefully defined considering the trade-off between capturing local idiosyncratic effects and avoiding weak predictors. To do so, for each country, I aggregate the regional observations in more aggregated entities which still maintain local idiosyncrasies. For Austria, regional data are available at the NUTS 2 level, whereas the dummies are aggregated at the NUTS 1 level. For Germany, only a few aggregations are performed using older Statistical Yearbooks' aggregation. In detail, Schleswig-Holstein is merged with Hamburg, Lower Saxony with Bremen, and Rhineland-Palatinate with Saarland. Next, for Italy is followed the NUTS-1 geographical aggregation: North-West, North-East, Centre, South, Islands.

Finally, the aggregation level of Denmark and Belgium are kept respectively at NUTS-2 and NUTS-1.

B Instrumental variable

B.1 First-stage

	Labour market				Health			Family	
	(Active into labour market) UR grad	(Earnings from work) UR grad	(Working hours) UR grad	(Hourly earnings) UR grad	(Self-perceived health) UR grad	(Objective grip strength) UR grad	(EURO-D) depression scale UR grad	(Number of children) UR grad	(Never married or divorced) UR grad
Instrument	0.38***	0.30***	0.31***	0.31***	0.44**	0.43***	0.42***	0.44***	0.44***
Std err	(0.041)	(0.046)	(0.045)	(0.046)	(0.040)	(0.039)	(0.043)	(0.039)	(0.039)
N	7196	5274	5631	5222	9509	9050	7356	9389	9509
F-test	87.47	42.44	48.02	43.67	123.1	121.45	99.18	123.15	123.1

Table 13: 2SLS first stage point estimates, standard errors, sample sizes, and F-tests. UR grad is the unemployment rate faced at graduation (treatment) and the instrument is the unemployment rate determined by the exogenous minimum school-leaving age. The first stage is different across different outcomes due to the different sample sizes.

B.1.1 Relevance

In this subsection, I check the relevance of the IV by educational attainment level.

In table 14 I show that the instrument is not relevant for high-educated individuals (F-statistics are always below 10). In table 15 I estimate the long-term effects by removing the high-educated individuals. No relevant differences can be noted. Hence, the baseline IV estimands of interest are likely capturing especially the long-term effect of low and middle-educated individuals.

	Labour market				Health			Family	
	(Active into labour market) UR grad	(Earnings from work) UR grad	(Working hours) UR grad	(Hourly earnings) UR grad	(Self-perceived health) UR grad	(Objective grip strength) UR grad	(EURO-D) depression scale UR grad	(Number of children) UR grad	(Never married or divorced) UR grad
Low-educ									
F-test	176.41	64.76	81.96	87.50	253.78	249.99	204.79	245.75	253.78
Mid-educ									
F-test	35.71	30.53	26.45	32.19	55.25	49.49	40.33	58.88	55.25
High-educ									
F-test	5.89	5.01	6.77	6.30	8.79	9.25	9.60	9.80	8.78

Table 14: 2SLS first stage point estimates, standard errors, sample sizes, and F-tests. UR grad is the unemployment rate faced at graduation (treatment) and the instrument is the unemployment rate determined by the exogenous minimum school-leaving age. The first stage is different across different outcomes due to the different sample sizes.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
2SLS, Baseline	-0.059***	-0.151**	-0.182	-0.100**	0.089**	0.206	0.034	0.042	-0.010
	(0.022)	(0.068)	(0.923)	(0.048)	(0.037)	(0.308)	(0.096)	(0.040)	(0.015)
2SLS, No high-educ	-0.066***	-0.141**	0.350	-0.094**	0.111***	0.075	0.095	0.052	-0.014
	(0.0203)	(0.0649)	(0.880)	(0.0458)	(0.0373)	(0.306)	(0.0941)	(0.0375)	(0.0138)
Observations	5,157	3,526	3,790	3,485	6,812	6,450	5,249	6,725	6,812
R-squared	0.026	0.052	0.090	0.016	0.038	0.567	0.035	0.020	0.017
First stage F-stat	105.2	52.31	58.62	54.37	146.9	145.7	115.1	152.5	146.9

Table 15: 2SLS estimates by removing the high-educated people. Baseline specification. Standard errors clustered by year-region of completing education.

B.2 Reduced form

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Instrument	-0.023***	-0.048**	-0.060	-0.028*	0.089**	0.084	0.017	0.015	-0.04
Std err	(0.008)	(0.21)	(0.307)	(0.015)	(0.038)	(0.138)	(0.042)	(0.018)	(0.007)
N	7196	5274	5631	5222	9509	9050	7356	9389	9509

Table 16: OLS regression of long-term outcomes on the instrument. Baseline specification.

B.3 Monotonicity

Previous literature points out that IV estimates can be interpreted as a Local Average Treatment Effect only if monotonicity is satisfied (Imbens and Angrist (1994); Angrist et al. (1996)). In brief, monotonicity implies that a change in the instrument affects the treatment of all individuals in the same direction. In the context of non-monotonicity, the instrumental variable estimate would represent a weighted mean of marginal treatment effects, with the caveat that the weights do not sum up to one (Angrist et al. (1996); Heckman and Vytlacil (2005)). To understand why the monotonicity assumption is valid in my framework, two subtle factors must be discussed. First, during the period under analysis, the unemployment rates have an increasing trend. Consequently, individuals graduating within the same region but at a later point will encounter both higher instrumental and treatment unemployment rates. Second, the unemployment rates are highly path-dependent within each region. As a result, regions marked by high initial unemployment rates are prone to maintain higher rates as time progresses. To provide empirical evidence about the validity of the monotonicity assumption in my setting, I follow the insights in Bald et al. (2019). They claim that in the case of monotonicity, the first-stage coefficients for any sub-sample should be always non-negative. In table 17, I provide evidence about the fact that for different sub-samples, the coefficients are always non-negative.

	Education			Country					Year completing education			Age at interview		
	Low	Mid	High	Austria	Germany	Italy	Denmark	Belgium	1960-1970	1970-1980	1980-1990	50-55	55-60	60-65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Women														
	0.675***	0.159**	0.230**	0.776***	0.372***	0.686***	0.263**	0.281***	0.499***	0.294***	0.102	0.423***	0.371***	0.657***
	(0.067)	(0.063)	(0.110)	(0.064)	(0.080)	(0.096)	(0.120)	(0.064)	(0.084)	(0.056)	(0.098)	(0.084)	(0.064)	(0.079)
Men														
	0.660***	0.452***	0.168*	0.689***	0.329***	0.794***	0.155	0.212***	0.489***	0.239***	0.369***	0.427***	0.439***	0.481***
	(0.060)	(0.064)	(0.088)	(0.110)	(0.088)	(0.081)	(0.100)	(0.044)	(0.081)	(0.063)	(0.105)	(0.086)	(0.062)	(0.073)

Table 17: 2SLS First Stage for various Sub-samples. The outcome variable is the unemployment rate experienced at the time of completing education. The main explanatory variable is the instrumental variable determined by the exogenous year of birth and compulsory minimum school leaving age laws. Controls contain the baseline specification. Standard errors clustered by year-region of completing education. *** denotes significance at 1%, ** at 5%, and * at 10%.

B.4 Functional forms

In table 18 I check the robustness of the IV estimates to different functional form specifications. The estimated results remain robust.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Linear trend	-0.045**	-0.141***	-0.618	-0.081**	0.058	-0.063	0.056	-0.015	-0.004
	(0.017)	(0.054)	(0.712)	(0.037)	(0.044)	(0.350)	(0.105)	(0.050)	(0.018)
Quadratic trend	-0.059***	-0.071***	-0.548	-0.052***	0.017	-0.730**	0.163**	-0.010	-0.043***
	(0.014)	(0.023)	(0.443)	(0.020)	(0.026)	(0.287)	(0.077)	(0.026)	(0.010)
FE (Baseline)	-0.059***	-0.151**	-0.182	-0.100**	0.089**	0.206	0.034	0.042	-0.010
	(0.022)	(0.068)	(0.923)	(0.048)	(0.037)	(0.308)	(0.096)	(0.040)	(0.015)
FE, FExFE	-0.100***	-0.136	0.199	-0.096	0.127***	0.003	0.054	0.014	0.001
	(0.031)	(0.086)	(1.223)	(0.066)	(0.038)	(0.276)	(0.097)	(0.041)	(0.012)

Table 18: 2SLS estimates according to different specifications. FE contains only FE and FExFE contain the interactions between the region of graduation and the year of graduation.

B.5 Baseline GRF vs IV-GRF

The lack of statistical properties to properly conduct the inference analysis in the case of continuous treatments and the difficulty of interpreting local average treatment effects at the individual level impede fully exploiting the instrumental variable in my work. However, since GRF allows estimating CAPE exploiting an IV ($\tau^{IV}(X) = \frac{\text{Cov}[Y,T|X=x]}{\text{Cov}[T,Z|X=x]}$), I compare the IV estimates with the baseline ones.

Figure 25 shows that although the IV distributions of CAPEs are more dispersed for both hourly earnings and total years worked, the baseline analysis is more conservative being likely the effects are overestimated.

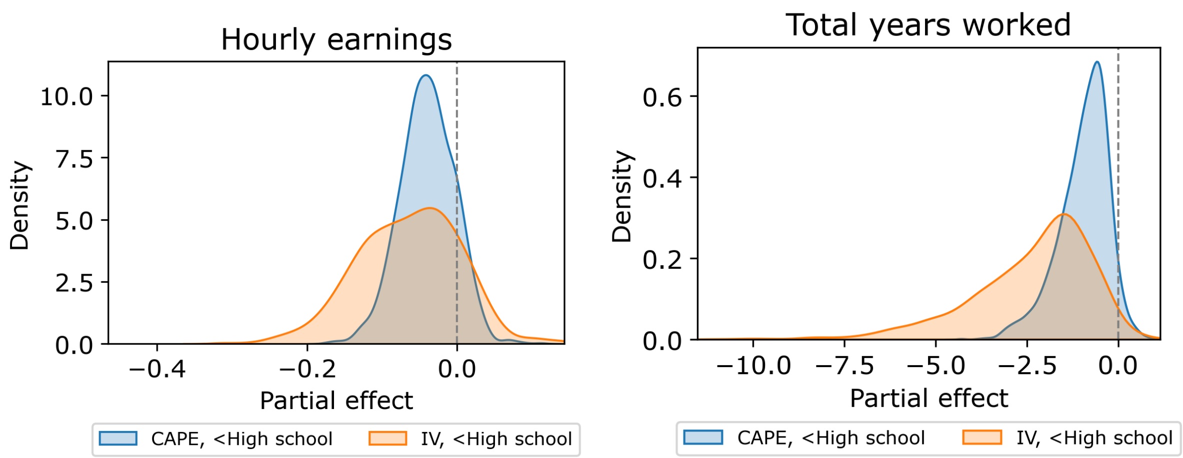


Figure 25: Baseline CAPEs and CAPEs estimated through an IV for "Hourly earnings" and "Total years worked".

C Further heterogeneity

C.1 Heatmap for Total Years Worked

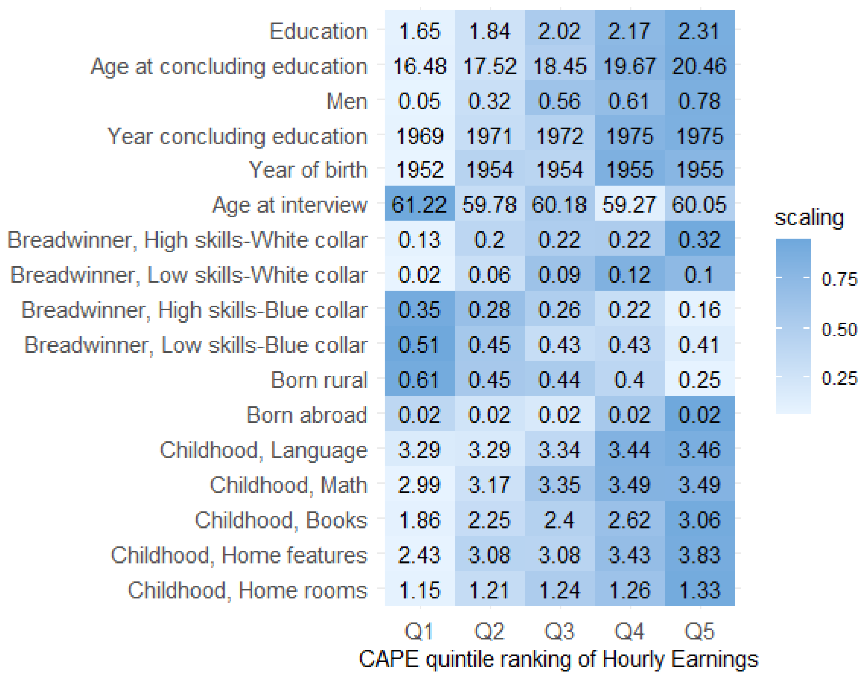


Figure 26: Average covariate values within quintile groups (based on CAPE estimate ranking) of "Total years worked". The scaling is achieved through a standardisation process where the proportion of a normal population is less than or equal to any given quantile.

The heatmap in figure 26 provides insights into the profiles of individuals who experience heterogeneous levels of long-term impacts on their total years worked. At first glance, the quintiles may appear similar to those in figure 12, which pertains to hourly earnings. However, a closer examination reveals striking disparities. One striking observation is that the majority (95%) of individuals in the most affected quintile (Q1) are women. This gender disparity persists across subsequent quintiles, with men being significantly overrepresented (78%) in the least affected quintile (Q5). Another notable difference pertains to the birthplace of individuals, particularly those born in rural areas, who are more prevalent in the quintile associated with worse-off outcomes. Additionally, it's worth mentioning that differences in childhood ability proxies, while present, are relatively less pronounced when compared to the patterns observed in the hourly earnings heatmap."

In summary, a shortage of early opportunities disproportionately hinders women, particularly those from rural areas, from accessing the labour market.

C.2 Country

Figure 27 displays the density distribution of CAPEs and table 19 presents the double robust estimates according to the country of completing education. The first point to raise is that Italy and Austria have a probability distribution of the CAPEs which is more centred on the more negative effects on "Hourly earnings". However, once the CAPEs are inserted into the double robust estimator, the estimates are no longer significant, with the exception of Denmark. It is important to stress that the statistical properties of the double-robust estimator employed to compute the Group Average Partial Effects are valid only asymptotically. As a result, the small sample size which is kept when each country is individually analysed does not allow to provide reliable estimates. This is particularly true for Denmark for which very few low-educated individuals are available in the dataset.

To fade away the doubts that the effects are driven by a single specific country (e.g. Denmark), I perform the analysis by removing one country per time.

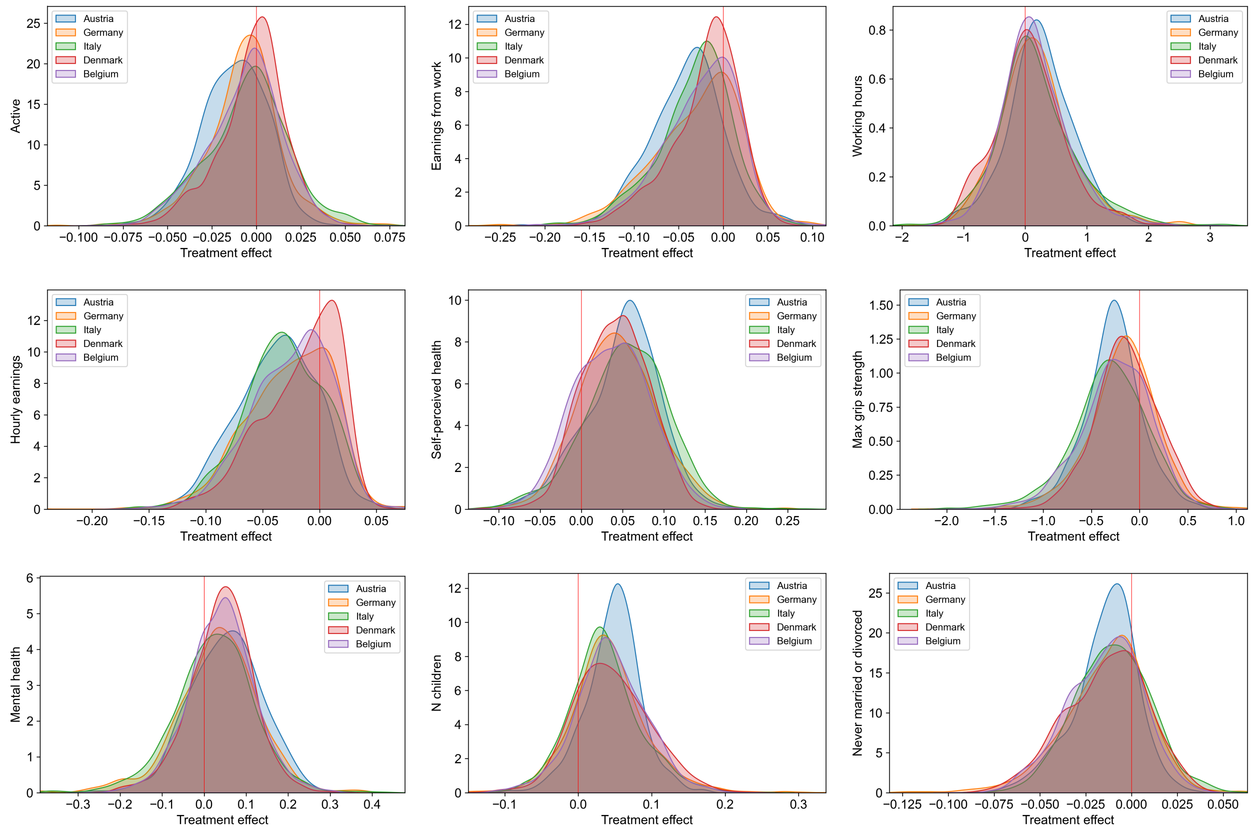


Figure 27: Distribution of CAPEs according to country of completing education.

	Active into labour market	Earnings from work	Working hours
Baseline	-0.007	-0.048**	0.422
Austria	0.007	-0.068	-0.827
Germany	0.014	-0.046	0.406
Italy	-0.018	-0.041	0.696
Denmark	-0.061***	-0.181***	-0.139
Belgium	-0.004	0.011	-0.201
	Hourly earnings	**Subjective health**	**Max grip strength**
Baseline	-0.043***	0.066***	-0.365**
Austria	0.008	-0.007	-0.048
Germany	-0.034	0.021	-0.145
Italy	-0.05*	0.096***	-0.646**
Denmark	-0.141***	0.234***	0.012
Belgium	0.013	0.053**	-0.302*
	Mental health	**Number of children**	**Never married or divorced**
Baseline	0.013	0.027	-0.011
Austria	-0.147	-0.032	-0.047**
Germany	0.006	0.064	-0.016
Italy	-0.015	0.006	-0.003
Denmark	0.135*	-0.087	-0.04**
Belgium	0.22***	0.062	0.004

Table 19: Heterogeneity according to the country by GRF.

To show that the effects are not driven by one specific country, I run the main estimations by removing one country's observations at a time. The results displayed in table 20 provide strong evidence of the robustness of the main results.

	Active				Earnings				Hours worked			
	All	Low	Mid	High	All	Low	Mid	High	All	Low	Mid	High
Main	-0.008	-0.014	-0.009	0.003	-0.052***	-0.072	-0.048*	-0.036	0.346	0.987	0.54	-0.858**
No Austria	-0.008	-0.025	0	0.002	-0.045*	-0.066	-0.019	-0.074**	0.422	0.965	0.795	-0.974**
No Germany	-0.013	-0.014	-0.017	-0.001	-0.033	-0.038	-0.037	-0.015	0.621*	1.269**	0.839	-1.095***
No Italy	-0.009	0.014	-0.017	-0.003	-0.065	-0.117	-0.049	-0.077	0.215	-1.82	0.825	-0.132
No Denmark	0	-0.011	0.004	0.005	-0.038*	-0.054	-0.021	-0.058	0.444	1.034	0.783	-1.049**
No Belgium	-0.005	-0.016	0	0.002	-0.055**	-0.067	-0.043	-0.066	0.504	0.936	0.853	-0.782*

	Hourly earnings				Subjective health				Max grip strength			
	All	Low	Mid	High	All	Low	Mid	High	All	Low	Mid	High
Main	-0.04***	-0.063*	-0.04*	-0.016	0.053**	0.021	0.089***	0.021	-0.342**	-1.119***	0.082	-0.127
No Austria	-0.042**	-0.052	-0.039	-0.036	0.085***	0.025	0.133***	0.068*	-0.455**	-1.112***	-0.064	-0.402
No Germany	-0.051**	-0.056	-0.077**	0.014	0.099***	0.091**	0.114***	0.084***	-0.724***	-1.169***	-0.361	-0.449**
No Italy	-0.027	0.032	-0.031	-0.04	0.017	-0.192	0.068	0.019	0.18	0.035	0.292	0.038
No Denmark	-0.037**	-0.06	-0.039	-0.003	0.069***	0.025	0.106***	0.05	-0.425**	-1.097***	-0.009	-0.401
No Belgium	-0.055***	-0.07*	-0.061**	-0.026	0.085***	0.032	0.117***	0.087*	-0.385*	-1.09**	0.038	-0.391

	Mental health				Number of children				Never married or divorced			
	All	Low	Mid	High	All	Low	Mid	High	All	Low	Mid	High
Main	-0.001	-0.187	0.118*	0.014	0.059**	0.033	0.077**	0.058	-0.019**	-0.03*	-0.014	-0.015
No Austria	0.024	-0.231	0.211***	-0.035	0.043	0.023	0.078*	-0.005	-0.01	-0.026	0.001	-0.012
No Germany	-0.013	-0.129	0.079	0.059	-0.02	-0.004	-0.031	-0.037	-0.005	-0.002	-0.01	-0.001
No Italy	0.028	-0.433	0.229**	-0.131	0.148**	0.227	0.175**	0.062	-0.019	-0.132	0.009	-0.018
No Denmark	-0.014	-0.282*	0.163**	-0.033	0.041	0.032	0.069*	-0.012	-0.011	-0.029	0.001	-0.011
No Belgium	-0.024	-0.307*	0.156**	-0.056	0.032	0.005	0.071*	-0.019	-0.01	-0.027	0.005	-0.017

Table 20: Robustness check: results are run by removing one country at a time.

C.2.1 Removing Germany and Austria

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.013	-0.033	0.621*	-0.051**	0.099***	-0.724***	-0.013	-0.02	-0.005
Low	-0.014	-0.038	1.269**	-0.056	0.091**	-1.169***	-0.129	-0.004	-0.002
Middle	-0.017	-0.037	0.839	-0.077**	0.114***	-0.361	0.079	-0.031	-0.01
High	-0.001	-0.015	-1.095***	0.014	0.084***	-0.449**	0.059	-0.037	-0.001

Table 21: Average partial effects estimated GRF removing Germany and Austria. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate experienced at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

D Robustness

D.1 Placebo analysis

To show that the early labour market conditions experienced after completing education are really driving the long-term effects, I run a placebo analysis in which the unemployment rate faced after completing education is randomly assigned. In detail, the unemployment rate (treatment) is assigned to each region-year from a random draw from a Chi-square distribution with 4 degrees of freedom (similar to real distribution)[37]. Since the unemployment rate is randomly assigned, whereas the theoretical relationship is between the true unemployment rate and the long-term outcome, no effect is expected.

Figure 28 shows the distribution of the Treatment Effect for 1000 simulations for ”Hourly earnings” by using the baseline estimation framework with the only difference that now the treatment is randomly assigned. As expected, the placebo distribution is centred around 0, and it is far from the estimated treatment effect -4%. Hence, the placebo analysis shows that the long-term effects are not there if the unemployment rates are randomly assigned and the original assignment of the initial unemployment rates is likely to drive the long-term results.

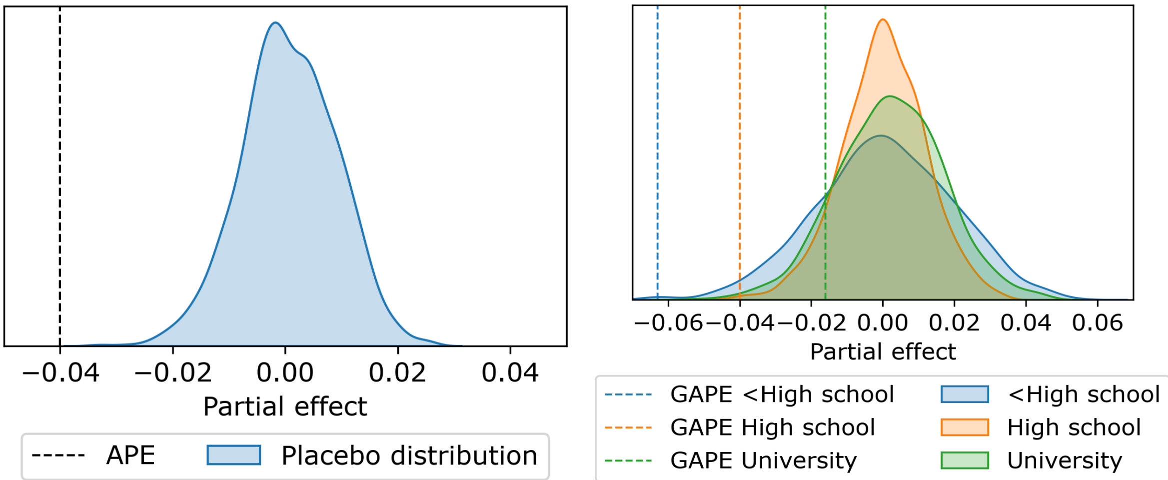


Figure 28: Distribution of the partial effects in the placebo analysis using GRF. 1000 simulations. The outcome is ”Hourly earnings”. The unemployment rate is assigned to each region-year from a random draw from a Chi-square distribution with 4 degrees of freedom (similar to real distribution). The vertical dotted lines are the baseline partial effects.

Furthermore, this framework can be extended to test if GRF is estimating heterogeneous effects when no heterogeneity is expected. Again, figure 28 (on the right) shows that the distributions of the group partial effects according to the level of education are centred around zero and are far from the baseline estimates represented by the vertical dotted lines. As a consequence, the estimation framework does not spot any heterogeneity. Another insight from this exercise is the impact of the sample size. High school graduates are the most numerous (2299 individuals), followed by university degree holders (1738), and finally, individuals with less than a high school education (1186). The distribution of the partial effect of ”<High school” has fatter tails, so, small sample sizes are more likely to detect biased point estimates. Next, table 22 shows the empirical proportion of

[37]Using different random assignment strategies such as changing to a uniform distribution provides similar results.

cases in which the null hypothesis is rejected out of the 1000 placebo simulations given that it is true (type I error). From this table, it can be seen that the proportion of cases in which the null hypothesis is rejected when the true is null is not relevantly guided by the sample size.

Sample	Sample size	Significance level		
		0.10	0.05	0.01
All	5223	0.093	0.052	0.009
<High school	1186	0.106	0.043	0.014
High school	2299	0.097	0.049	0.015
University	1738	0.092	0.045	0.008

Table 22: Proportion of cases out of 1000 placebo simulations in which the null hypothesis is rejected according to the significance level and sample.

D.2 Analysis by gender

In this section, I check if the results change if the analysis is performed separately by gender. Table 23 shows that the results remain almost identical to the baseline heterogeneity analysis in which the entire sample is used (table 10).

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Women	-0.03**	-0.049*	-0.426	-0.029	0.036	-0.402**	-0.007	0.057**	-0.019*
Women Low	-0.033	-0.069	-0.995	0.012	-0.023	-1.049***	-0.236*	0.072	-0.039*
Women Middle	-0.03*	-0.055	-0.042	-0.061**	0.059*	-0.073	0.127	0.044	-0.01
Women High	-0.023	-0.019	-0.654	-0.001	0.085*	-0.11	0.054	0.065	-0.007
Men	0.01	-0.05*	0.885**	-0.047**	0.051**	-0.165	0.04	0.05	-0.02**
Men Low	-0.003	-0.07	1.982***	-0.084*	0.087*	-0.893	-0.035	-0.006	-0.022
Men Middle	0.01	-0.043	0.967	-0.029	0.069**	0.246	0.098	0.092*	-0.021
Men High	0.029**	-0.039	-0.8	-0.038	-0.042	-0.051	0.02	0.039	-0.015
Women Obs	3887	2578	2764	2552	5110	4859	3974	5050	5110
Men Obs	3310	2697	2868	2671	4401	4191	3382	4341	4401

Table 23: Average partial effects estimated GRF using separately only the subsample containing women and men. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate experienced at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

D.3 Alternative treatments

Alternative unemployment rate definitions

The unemployment rate at the time of completing education is the most used proxy in the literature for capturing the economic opportunities encountered at the beginning of a career. However, alternative measures of unemployment can be used. Table 24 shows that using the average unemployment rate experienced in the first three years after completing education provides results similar to the baseline.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.03***	-0.026	0.371	-0.03**	0.011	-0.535***	-0.037	0.035	-0.025**
Low	-0.038*	-0.06	1.611*	-0.062*	0.011	-0.996***	-0.177	0.074*	-0.06***
Middle	-0.037***	-0.006	0.072	-0.024	0	-0.323	0.019	0.001	-0.009
High	-0.003	-0.031	-0.489	-0.009	0.036	-0.317	0.05	0.05	-0.008
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 24: Average partial effects estimated GRF. The dependent variables are the long-term outcomes. The treatment variable is the average unemployment rate experienced in the first 3 years after completing education. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

Deviation from unemployment trend

I investigate the robustness of the results using the deviation from unemployment trends. In detail, I use the Hodrick-Prescott filter, with the smoothing parameter equal to 100, to separate the trend and cyclical components of the unemployment rate within each region. Table 25 shows that results remain broadly robust.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.015	-0.093***	0.248	-0.074***	0.087**	-0.258	0.17*	0.037	-0.025*
Low	0.006	-0.094	1.791	-0.09	0.029	-0.633	0.111	-0.029	-0.021
Middle	-0.02	-0.088*	-0.142	-0.07*	0.136***	-0.114	0.271**	0.085	-0.032*
High	-0.039	-0.103	-0.766	-0.063	0.064	-0.028	0.021	0.028	-0.016
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 25: Average partial effects estimated GRF. The dependent variables are the long-term outcomes. The treatment variable is the deviation of the unemployment rate experienced at graduation from the trend determined by Hodrick-Prescott filtering ($\lambda = 100$). GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

D.4 Other robustness

Country unemployment rate

Table 26 shows that using unemployment rates at the country level adds relevant attenuation bias. Most of the results become not significant and the "Self-perceived health" changes also the direction of the effect.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.006	-0.012	0.193	-0.017	-0.088***	-0.461**	0.012	-0.003	-0.022**
Low	-0.019	-0.032	1.057*	-0.05	-0.156***	-0.99***	-0.189	-0.052	-0.004
Middle	-0.006	0.016	0.136	-0.001	-0.058*	-0.158	0.126	0.026	-0.039***
High	0.016	-0.047	-0.737	-0.014	-0.047	-0.348	0.056	0.01	-0.013
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 26: Average partial effects estimated GRF. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation measured at the country level. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

PCA for childhood conditions

	Labour market					Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Years worked	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.008	-0.056**	0.246	-0.038***	-0.758***	0.05**	-0.413**	0.003	0.066**	-0.021**
<High school	-0.012	-0.082*	0.759	-0.054	-1.234**	0.023	-1.197***	-0.159	0.038	-0.030*
High school	-0.010	-0.050	0.439	-0.039*	-0.529*	0.080***	-0.016	0.112*	0.089**	-0.017
University	0.004	-0.041	-0.800**	-0.017	-0.525**	0.024	-0.125	0.003	0.058	-0.017
Observations	7574	5550	5930	5494	9511	10381	9878	8169	10770	10381

Table 27: Average partial effects estimated GRF using a PCA for childhood conditions. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, year of interview fixed effects, and gender. Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

No childhood conditions controls

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.009	-0.071***	0.397	-0.06***	0.081***	-0.444**	0.017	0.035	-0.012
Low	-0.02	-0.11*	1.146	-0.101**	0.038	-1.193***	-0.18	0.034	-0.029*
Middle	-0.007	-0.054	0.642	-0.053**	0.112***	-0.005	0.169***	0.056	-0.003
High	0	-0.066*	-0.956**	-0.028	0.073**	-0.394	-0.05	-0.009	-0.009
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 28: Average partial effects estimated GRF without childhood controls. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, year of interview fixed effects, and gender. Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

No clustering

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.007	-0.051**	0.328	-0.046***	0.073***	-0.409*	-0.013	0.062*	-0.033***
Low	-0.027	-0.057	0.803	-0.063**	0.009	-1.349***	-0.226*	0.045	-0.052**
Middle	0.002	-0.047	0.545	-0.048*	0.127***	0.051	0.135*	0.078	-0.022
High	0.002	-0.05	-0.726*	-0.022	0.049	-0.028	-0.026	0.054	-0.027
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 29: Average partial effects estimated GRF. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation. No clustering. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

Equalize cluster

Here, instead of weights, each cluster has the same weights regardless the sample size of each cluster.

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
All	-0.024***	-0.034*	0.23	-0.037**	0.031*	-0.267**	0.067*	-0.001	-0.008
Low	-0.029	-0.01	1.341**	-0.044	0.051	-0.623**	0.031	-0.015	-0.01
Middle	-0.03***	-0.054	0.196	-0.057**	0.001	-0.07	0.086*	0.009	-0.002
High	-0.008	-0.024	-0.737***	0	0.057***	-0.214*	0.074*	-0.004	-0.017**
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 30: Average partial effects estimated GRF. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation. All clusters receive the same weights regardless the sample size of each cluster. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, and year of interview fixed effects. Additionally, the following controls are included: gender and a rich set of childhood conditions at age 10 (health, ability, and socioeconomic status). Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

E More about the Generalized Random Forest

Since the GRF is an extension of "standard" random forests (Breiman (2001)), I briefly outline the random forest framework and explain the differences with the GRF framework. Next, I discuss the hyper-parameter tuning procedure. Finally, I show which variables are more important in determining the heterogeneity in the adaptive weighting function derived from a forest.

E.1 (Generalized) Random Forest

Random Forests (Breiman (2001)) are a class of supervised learning algorithms which are based on an ensemble of several decision trees. Below, I briefly outline the baseline framework of a Random Forest (RF). For a full description of Random Forest please refer to Hastie et al. (2009).

Let N be independent and identically distributed samples, indexed $i = 1, ..., n$. $Y_i \in \mathcal{R}$ is the outcome of interest. The quantity of interest is the conditional mean function $\mu(x_i) = \mathbb{E}[Y_i|X_i = x_i]$. Let D be the full dataset. X is a $n \times p$ matrix, where n is the number of observations and p is the total number of covariates. First, the full dataset is split into training (D_{train}) and test data (D_{test}). During the training process, B trees are constructed following the next steps. For each decision tree T_b, a single root node is created by randomly drawing a $sample.fraction \in (0, 1]$ without replacement from the training dataset. Next, the root node is greedily partitioned into smaller child nodes (C_1, C_2), which are then repeatedly divided to construct a tree. Before each partitioning, a random subset of variables $mtry < p$ is chosen as potential candidates for splitting. For each candidate x, are evaluated all its possible values v to determine a split into two children leaves. The quality of a split (x, v) depends on the loss function. In a "standard" decision tree, the split aims for minimizing the prediction error in the outcome of interest using the "Mean squared error" for regression and "Gini Index" for classification tasks. The splitting procedure stops when a certain rule is met, e.g., if the reduction in means squared error is lower than a given threshold.

To predict the outcome of any observation $i \in D_{test}$ for each decision tree b, the following equation is used

$$\hat{Y}_i^b = \sum_{m=1}^{M^b} \hat{Y}^b_{R^b_m} I\{x_i \in R^b_m\}, \tag{15}$$

where each observation x_i belongs to exactly one subset R^b_m. I is an indicator function equal to one if x_i is in R^b_m and 0 otherwise. $\hat{Y}^b_{R_m}$ is the mean of all training observations in partition R^b_m. M^b is the number of total final leaves of the decision tree.

Figure 29 shows graphically how a single decision tree works. If $x_{1,i} < 3$ and $x_{2,i} < 2$, then i will fall in the lower-left partition R_1 in figure 29 (on the right). By averaging the outcome of all observations falling in partition R_1, $\hat{Y}_i = \hat{Y}_{R_1}$.

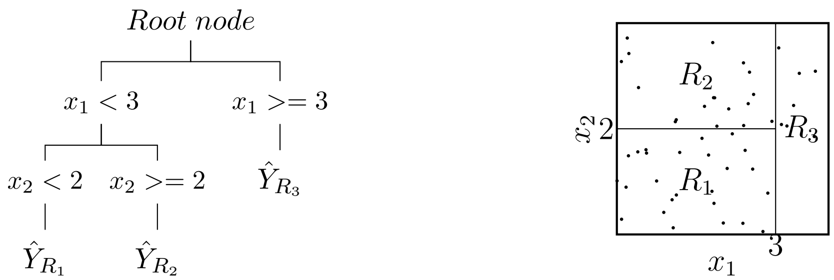


Figure 29: On the left, a single decision tree. On the right, partitioning of the covariate space of the single decision tree.

Since individual decision trees have low bias but high variance, averaging over several trees meaningfully stabilises predictions. Consequently, this procedure is repeated B times to achieve B different decision trees. The final prediction of $\hat{Y}_i$ is achieved by averaging the predictions of all decision trees (equation 16).

$$\hat{Y}_i = \frac{1}{B} \sum_{b=1}^{B} \hat{Y}^b \tag{16}$$

The Generalized Random Forest (Athey et al. (2019)) is a generalisation of the "standard" Random Forest (Breiman (2001)). As a matter of fact, GRF preserves several core elements of RF such as recursive partitioning, subsampling, and random split selection. However, the final estimate is no more obtained by averaging estimates from each decision tree. In detail, to facilitate the incorporation of statistical extensions, GRF interprets forests as a type of adaptive nearest neighbour estimator. Section 4.2.2 outlines the statistical framework of the adaptive nearest neighbour estimator framework. In brief, the forest calculates a weighted set of neighbours for each test point x and then solves a plug-in version of the estimating equation 8 using these neighbours. The term "Generalized" derives from the fact that the standard conditional mean estimation obtained by the averaging step in RF and the weighting views of forests of GRF are equivalent (see section 2.1 Athey et al. (2019)). However, GRF can be further extended to different quantities of interest such as quantile regression, conditional average partial effect estimation and heterogeneous treatment effect estimation by maintaining stability, ease of use, and flexible adaptation to different functional forms.

GRF can be expanded to estimate Causal Forests, i.e., it allows estimating average partial effects in the case of a treatment variable. In this case, the quality of a split (x, v) is measured by the degree to which it increases the heterogeneity in the quantity of interest, i.e., the quality of a split depends on how different the treatment effect estimates are in each node. Next, splits that lead to child nodes that are excessively imbalanced in terms of treated and control or too small are discarded. In addition, GRF uses different subsamples for constructing the tree and for making predictions ("honesty"). The motivation behind honesty is to reduce bias in tree predictions Wager and Athey (2018). Directly optimising the heterogeneity criterion is computationally expensive. To reduce computational costs, GRF computes the gradient of the objective and optimises a linear approximation to the criterion.

Hyperparameters

GRF uses the following hyperparameters: sample fraction, number of trees, honesty, honesty fraction, honesty prune leaves, mtry, minimum node size, alpha, and imbalance penalty.

"Sample fraction" is a value in the range of $(0, 1]$, which regulates the proportion of instances employed to construct each tree. The default value is 0.5.

The "number of trees" simply defines the number of trees produced during the training process. The default is 2000. In the baseline analysis, I use 5000 trees to increase the accuracy of predictions.

"Honesty" reduces bias in tree predictions, by using different subsamples for constructing the tree and for making predictions. Honest forests randomly partition this subsample into two halves and utilize only the first half for splitting. The second half is reserved for populating the tree's leaf nodes: as new instances are introduced, they are assigned to the appropriate leaf node. After splitting, the leaf nodes are "repopulated" using a new set of examples.

The "honesty fraction" parameter enables the adjustment of the proportion of instances employed for selecting tree splits. The default value is 0.5.

"Honesty prune leaves" is "True" per default meaning that it prunes away empty leaves so that each tree is able to handle all test points.

The "mtry" parameter controls the number of variables that are evaluated during each split. The default value for "mtry" is calculated as $min(sqrt(p) + 20, p)$, where p represents the number of variables (columns) in the dataset.

The parameter "alpha" determines the maximum allowed imbalance of a split. Specifically, when a parent node is split, the size of each resulting child node must not be less than alpha times the size of the parent node. Alpha must be a value between 0 and 0.25, and the default value is 0.05. In the case of Causal Forest, it is necessary to take into account not only the number of examples in each node but also the number of treatment and control examples. Specifically, the algorithm first computes the average of the treatment values in the parent node. Then, when considering a split, each child node must have at least "min node size" samples with treatment values less than the average, and at least that many samples with treatment values greater than or equal to the average. The default is 5.

The imbalance penalty parameter in GRF controls how imbalanced splits are penalized during tree growth. When choosing which variable to split on, each potential split is assigned a measure of "goodness" that reflects how much it increases heterogeneity between the resulting child nodes. However, GRF also applies a penalty to this measure to discourage splits that produce child nodes of vastly different sizes. The penalty is calculated by multiplying the imbalance penalty parameter by the inverse of the size of each child node (left and right). This penalty complements the hard restriction imposed by the alpha parameter, which sets a minimum size threshold for each child node relative to the size of the parent node. By default, the imbalance penalty is set to 0, indicating that no split penalty is applied.

E.1.1 Hyperparameters tuning

In the baseline analysis, I estimate the partial effects by using the above-mentioned default hyper-parameters to avoid long-computational costs. As a matter of fact, literature shows that random forest algorithms are less sensitive to hyperparameter tuning when the covariate space is not too large (Probst et al. (2019)). Furthermore, the default hyperparameters are calibrated to values that often perform reasonably well Athey and Wager (2019). To check the robustness of the main results to different hyperparametrizations, in table 32 I report the main results using a cross-validation procedure to select more optimized values of the hyperparameters. Specifically, 100 distinct random sets of parameter values are drawn. Next, for each set of parameter values, train a forest with these values and compute the out-of-bag error. For causal forests, GRF uses a measure of error developed by Nie and Wager (2021) motivated by residual-on-residual regression (Robinson (1988)).

Table 31 shows the optimal hyperparameters chosen following the above-mentioned procedure. Table 32 shows the results using the above-mentioned optimized hyperparameters. The variability of hyperparameter values is quite large. However, results remain almost identical to the baseline for all the outcomes. The first reason why is related to the fact that the default hyperparameters are chosen to work well in a variety of contexts. Again, in the case of random forest algorithms, the estimated outcomes are not so sensitive if the covariate space is not too large.

		Labour market				Health			Family	
		Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
	Default									
sample.fraction	0.5	0.48	0.50	0.08	0.50	0.40	0.20	0.19	0.17	0.28
mtry	27	21.00	27.00	20.00	27.00	16.00	5.00	14.00	18.00	1.00
min.node.size	5	21.00	5.00	4.00	5.00	23.00	15.00	55.00	465.00	11.00
honesty.fraction	0.5	0.57	0.50	0.69	0.50	0.56	0.66	0.72	0.74	0.50
honesty.prune.leaves	1	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00
alpha	0.05	0.08	0.05	0.18	0.05	0.03	0.02	0.05	0.10	0.06
imbalance.penalty	0	0.68	0.00	0.38	0.00	0.65	0.67	0.26	1.13	0.51

Table 31: Optimized hyperparamters. The number of trees is set to 5000 as in the baseline analysis and honesty is set to "Yes".

	Labour market				Health			Family	
	Active into labour market	Earnings from work	Working hours	Hourly earnings	Self-perceived health	Objective grip strength	EURO-D depression scale	Number of children	Never married or divorced
Hyper opt	0.002	-0.054**	0.374	-0.041**	0.064***	-0.259	0.012	0.068**	-0.015*
Baseline	-0.008	-0.052***	0.346	-0.040***	0.053**	-0.342**	-0.001	0.059**	-0.019**
Observations	7574	5550	5930	5494	10381	9878	8169	10770	10381

Table 32: Average partial effects estimated by GRF with optimized hyperparameters. The dependent variables are the long-term outcomes. The treatment variable is the unemployment rate at graduation. GRF estimations include the region of graduation, year of graduation, age at interview, year of birth, year of interview fixed effects, and gender. Standard errors in brackets clustered by graduation year by region. *** denotes significance at 1%, ** at 5%, and * at 10%.

To further show the sensitivity of the estimate outcomes to hyperparameters tuning, now I focus only on "Hourly earnings" and a grid search. In detail, to reduce time costs, I will use a grid search for 3 hyperparameters while keeping the others at default values. Table 33 shows that regardless of the choice of hyperparameters, the point estimates and the statistical significance remain always robust to the baseline results.

	Sample fraction = 0.2				Sample fraction = 0.3				Sample fraction = 0.4				Sample fraction = 0.5			
Min node size	15	25	50	100	15	25	50	100	15	25	50	100	15	25	50	100
mtry = 10	-0.029	-0.028	-0.028	-0.028	-0.027	-0.027	-0.027	-0.026	-0.031	-0.031	-0.03	-0.03	-0.027	-0.027	-0.027	-0.026
Min node size	15	25	50	100	15	25	50	100	15	25	50	100	15	25	50	100
mtry = 20	-0.034	-0.034	-0.033	-0.033	-0.032	-0.032	-0.032	-0.031	-0.033	-0.033	-0.032	-0.032	-0.037	-0.037	-0.036	-0.036
Min node size	15	25	50	100	15	25	50	100	15	25	50	100	15	25	50	100
mtry = 30	-0.035	-0.035	-0.034	-0.034	-0.036	-0.036	-0.035	-0.035	-0.04	-0.039	-0.039	-0.038	-0.036	-0.036	-0.036	-0.035
Min node size	15	25	50	100	15	25	50	100	15	25	50	100	15	25	50	100
mtry = 40	-0.04	-0.04	-0.039	-0.039	0.041	-0.041	-0.04	-0.04	-0.038	-0.038	-0.038	-0.037	-0.038	-0.039	-0.038	-0.038

Table 33: Point estimates for "Hourly earnings" by fine-tuning hyperparameters. All point estimates are statistically significant at least at the 5% level. Contrary to the baseline, here the number of trees is set to the default value of 2000 for time constraints at the cost of precision in estimation.

E.2 Stability of Causal Forest estimates

In this section, I show that the flexibility of the Causal Forest can overcome biases arising from model misspecification in standard OLS approaches. Next, I provide evidence about the validity of the Causal Forest estimates in a finite sample.

E.2.1 Functional form simulation

In this section, I perform two simulations to show that the flexibility of machine learning estimators better copes with model specification errors.

The idea is to run equation 17 100 times by making use of OLS and Causal Forest to observe the deviation of the distribution of β from the "true" value according to different data-generating functions.

$$Y_i = \beta T_i + g(X) + \nu_i \tag{17}$$

Y_i and T_i are constructed below. X contain all the confounding variables used in the main analysis. $g()$ defines the functional form used in the estimation: linear specification in the case of OLS and flexible adaptation in the case of the Causal Forest. The sample size is 5223, i.e., the sample available for "Hourly earnings"[38].

The (linear) data-generating process is as follows:

$$\begin{aligned} Y_i &= \beta T_i + X_1 + X_2 + \epsilon_i \\ T_i &= X_1 + X_2 + \eta_i, \end{aligned} \tag{18}$$

where ϵ_i and η_i are *i.i.d.* response noises drawn from $\mathcal{N}(0,1)$. X_1 and X_2 are respectively the "Number of books" and "Mathematical ability" during childhood from the real dataset. β is the causal parameter of interest and is set to -1 by construction.

[38]Since it represents one of the most important variables under analysis.

The outcome and the treatment follow very similar data-generating processing to mimic the interesting case in which the most important confounders are also important for predicting the outcome Y. Specifically, following the relations in the main analysis, the outcome variable and the treatment are positively related to the "Number of books" and "Mathematical ability".

Figure 30 shows that when the DGP is linear, OLS slightly outperforms the Causal Forest estimator. However, both models are able to correctly estimate the parameter of interest.

Now, I modify the data-generating process to add non-linearity and interaction terms:

$$\begin{aligned} Y_i &= \beta T_i + X_1 + 2X_1^2 + X_2 - 2X_1X_2 + \epsilon_i \\ T_i &= X_1 + 2X_1^2 + X_2 - 2X_1X_2 + \eta_i, \end{aligned} \tag{19}$$

where ϵ_i and η_i are *i.i.d.* response noises drawn from $\mathcal{N}(0,1)$. X_1 and X_2 are respectively the "Number of books" and "Mathematical ability" during childhood. Again, β is set to -1 by construction.

Figure 30 shows that even though the Causal Forest presents a small bias, its flexibility allows it to provide estimates close to the true value of the parameter β. On the contrary, once the OLS model is not correctly specified, the estimates are severely biased.

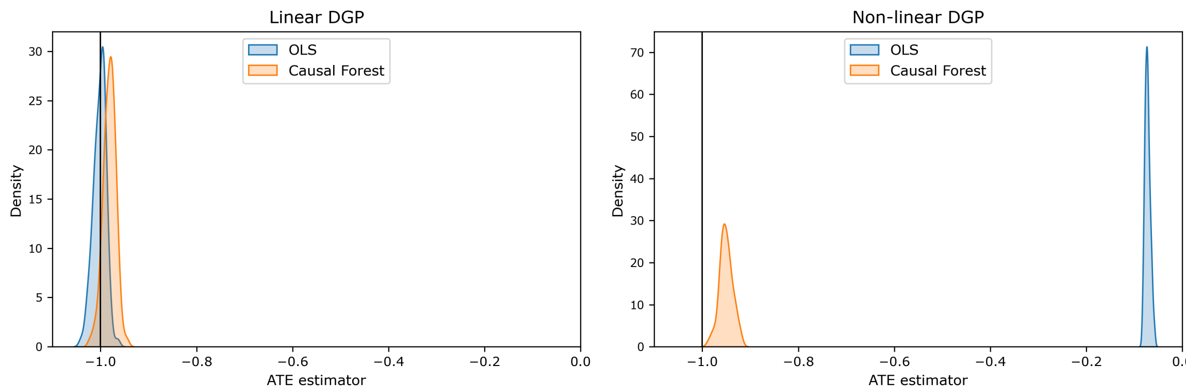


Figure 30: On the left, the distribution of point estimates of OLS and Causal Forest after 100 simulations using the linear DGP. On the right, the distribution of point estimates of OLS and Causal Forest after 100 simulations using the non-linear DGP. The true value for β is -1.

E.2.2 Finite sample analysis

Since the (Group) Average Partial Effects estimator relies on the asymptotic theory, I investigate the sensitivity of the main results to different finite sample sizes. Again, I start with the 5223 observations available for "Hourly earnings" and I draw different subsamples. Figure 31 shows the average of 10 simulations of the point estimates. It can be seen that the estimates are stable when the sample size reaches 1000 observations. This is reassuring since all estimations, including the heterogeneity analysis, involve a larger set of observations. Again, figure 31 (on the right) shows the gains of statistical testing power of larger sample sizes. For instance, for the case of educational GAPE, middle-educated individuals rely on 2299 observations versus 1186 low-educated. Hence, the lower statistical significance of the detrimental point estimates of low-educated individuals is also explained by standard errors which are approximately 50% larger if compared with the middle-educated which can rely on almost a double sample size.

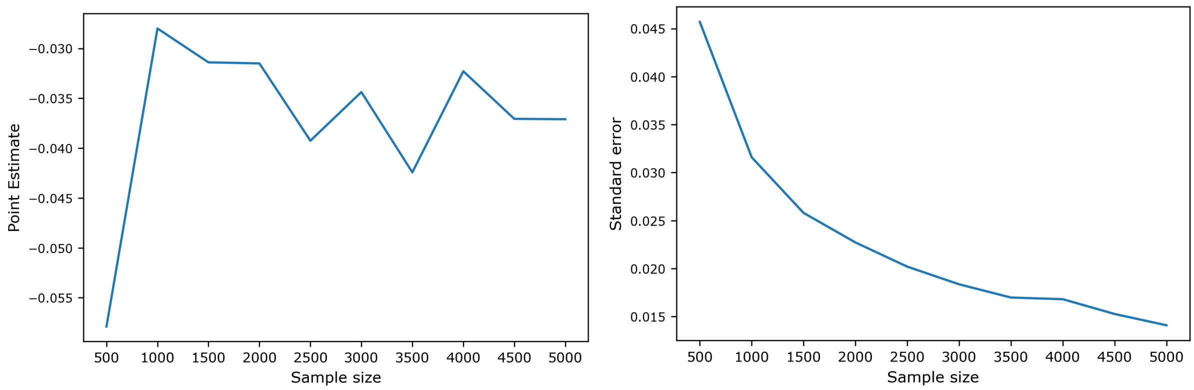


Figure 31: Finite sample point estimate stability and standard error gains from the sample size. The figures display the average of 10 simulations for each sample size.

However, in the previous analysis, I do not know the "true" underlying parameter. Hence, I repeat the analysis by employing the data-generating processes in equations 18 and 19 for which the "true" value of the parameter of interest is known, i.e., $\beta = -1$. As expected, figure 32 shows that OLS converges more rapidly if the underlying DGP is linear. On the contrary, if the DGP is non-linear and the OLS model is not correctly specified, the sample size does not reduce the bias. Again, Causal Forest requires a larger sample size to converge. However, after 2000-3000 observations, the estimator provides estimates economically equivalent with respect to the true value.

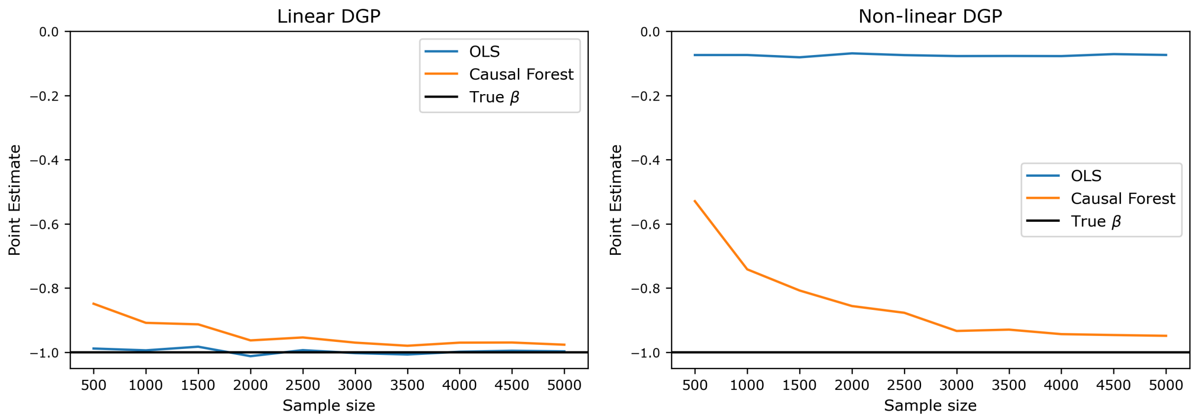


Figure 32: Finite sample properties of Causal Forest and OLS estimates using linear and non-linear data-generating processes.

E.2.3 Heterogeneity simulations

Now, I further expand the simulation exercise to investigate how well the GRF detects heterogeneity. In detail, I expand the data-generating process in equation 19 by introducing heterogeneous effects according to education.

$$Y_i = \beta_{education} T_i + X_1 + 2X_1^2 + X_2 - 2X_1X_2 + \epsilon_i$$
$$\beta_{education} = \begin{cases} -6 & if\ education < High\ school \\ -3 & if\ education = High\ school \\ 0 & if\ education = University \end{cases} \tag{20}$$

Figure 33 shows again that in the case of a non-linear data-generating process, OLS estimates are biased whereas the Causal Forest's CAPEs are centred around the true underlying parameters represented by the vertical dotted lines.

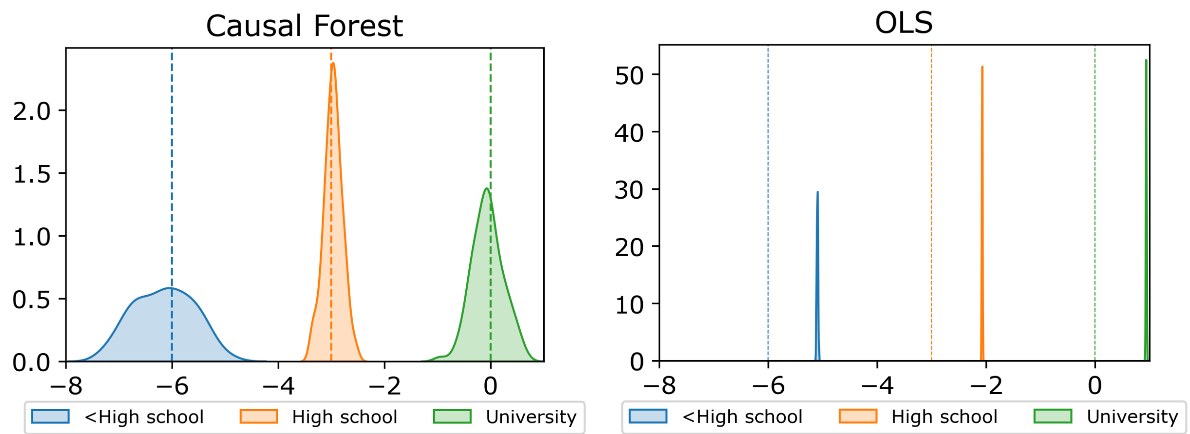


Figure 33: Density distribution of the GAPE according to education for 100 simulations in case of heterogeneity.

Next, to control if the estimation procedure detects any heterogeneity when no heterogeneity is expected, I estimate the educational GAPEs using the DGP in equation 19. Figure 34 depicts that no relevant heterogeneity is detected when no heterogeneity is present. The OLS bias due to the functional form misspecification is as in subsection E.2.1.

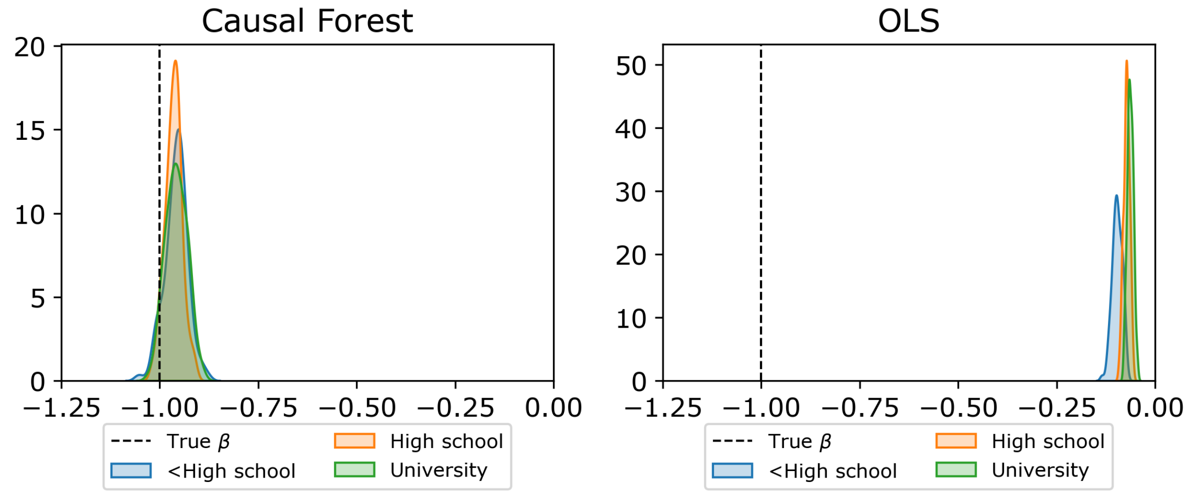


Figure 34: Density distribution of the GAPE according to education for 100 simulations in case of no heterogeneity.