Ca' Foscari University of Venice

Department of Economics

Working Paper

Agar Brugiavini
Petru Crudu

The Role of Disability Insurance on the Labour Market Trajectories of Europeans

ISSN: 1827-3580
No.20/WP/2023

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

The Role of Disability Insurance on the Labour Market Trajectories of Europeans

Agar Brugiavini
Ca' Foscari University of Venice;Institute for Fiscal Studies

Petru Crudu
Ca' Foscari University of Venice

Abstract
This work documents the role played by disability insurance, typically part of a wider public pension provision package, on the labour market trajectories and retirement decisions. We will first employ a machine learning approach to estimate a Transition Probability Model able to uncover the most likely labour market histories and then evaluate the effects of policy reforms, including reforms to the eligibility for disability insurance benefits. The main contribution is the introduction of disability insurance programs within a framework, which models the entire life course of older Europeans. This requires the detailed administrative eligibility criteria prevailing in each of the 11 countries from 1970 to 2017. Results show that the disability route and early retirement are substitutes. In addition, tightening eligibility rules of disability programs crowd out disabled workers, whose reductions in working capacities are correctly assessed, towards other compensatory schemes (e.g., unemployment benefits or early pension) in which working is not expected. On the contrary, individuals with over-assessed reductions in working capacities are the most reactive to disability policy restrictions. In conclusion, efficient disability assessment procedures are crucial for incentivising labour market participation without hurting individuals most in need.

Keywords
Retirement, Disability, Path Dependence, Simulation

JEL Codes
J14, J26, I38, H55

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

This Working Paper is published under the auspices of the Department of Economics of the Ca' Foscari University of Venice. Opinions expressed herein are those of the authors and not those of the Department. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional character.

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

The Role of Disability Insurance on the Labour Market Trajectories of Europeans

Agar Brugiavini* and Petru Crudu†

Abstract

This work documents the role played by disability insurance, typically part of a wider public pension provision package, on the labour market trajectories and retirement decisions. We will first employ a machine learning approach to estimate a Transition Probability Model able to uncover the most likely labour market histories and then evaluate the effects of policy reforms, including reforms to the eligibility for disability insurance benefits. The main contribution is the introduction of disability insurance programs within a framework, which models the entire life course of older Europeans. This requires the detailed administrative eligibility criteria prevailing in each of the 11 countries from 1970 to 2017. Results show that the disability route and early retirement are substitutes. In addition, tightening eligibility rules of disability programs crowd out disabled workers, whose reductions in working capacities are correctly assessed, towards other compensatory schemes (e.g., unemployment benefits or early pension) in which working is not expected. On the contrary, individuals with over-assessed reductions in working capacities are the most reactive to disability policy restrictions. In conclusion, efficient disability assessment procedures are crucial for incentivising labour market participation without hurting individuals most in need.

Keywords: Retirement, Disability, Path Dependence, Simulation

JEL Codes: J14, J26, I38, H55

*Ca' Foscari University of Venice, Institute For Fiscal Studies. Email: brugiavi@unive.it

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

1 Introduction

Disability insurance plays a fundamental role in protecting individuals hit by extreme health shocks. These shocks cause a reduction in the range of possible daily activities which could be performed, including working. Owing to the "low probability-high severity" of extreme health shocks, the provision of private insurance against disability is doomed to fail. Hence, in all developed countries, the responsibility of supporting individuals with health impairments is on government insurance programs. It is worth mentioning that the provision of a disability insurance program creates an insurance-incentive trade-off (Low and Pistaferri (2015)). If on the one hand government insurance programs act as a safety net for health-impaired individuals with income losses, on the other hand, individuals with high leisure preferences or hit by a negative productivity shock could exploit the disability benefit to exit the labour market and/or as an option for early retirement. Interestingly, despite the significant interconnections within the social protection system, the Disability Insurance scheme is seldom considered during pension reforms. Furthermore, while unemployment expenditure is typically a more frequently discussed subject, spending on disability-related benefits consistently surpasses that of unemployment insurance. In fact, disability expenditures consistently account for a substantial share of the GDP, never falling below 2%, and reaching peaks of 6% in "Nordic" countries. In this regard, the projected increase in the EU's old-age dependency ratio, soaring from 25% in 2000 to nearly 50% by 2050 emphasises the ahead demographic challenges. The reduction from four working individuals to just two supporting each elderly person poses a challenge to the sustainability of the public pension provision package. Given the ageing population, it is essential to meticulously fine-tune the complete social provision framework to ensure the continued provision of adequate protection to those who need it[1].

In this paper, we study the impact of disability insurance on labour market trajectories and retirement decisions of individuals across eleven European countries[2]. We add significant contributions to the existing literature. First of all, we gather the detailed administrative eligibility criteria for disability insurance programs, early retirement, and old-age pensions in each of the eleven European countries under analysis from 1970 to 2017. Secondly, we make use of a long retrospective panel of elderly Europeans from

[1]Section A provides a comprehensive discussion on the topic's relevance, delving into greater detail.

[2]The following 11 European countries are considered: Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland and Belgium.

the Survey of Health, Ageing and Retirement in Europe (SHARE Börsch-Supan et al. (2013)). One crucial advantage of employing survey data instead of administrative data in this context is its capability to encompass individuals with disabilities who may not seek any allowance or benefits. This valuable dataset offers comprehensive insights into individuals' entire life trajectories. Leveraging the path dependency of lifetime events, we explore retirement pathways, recognising that they are shaped not merely by personal choices but also by life opportunities and resources accumulated over an individual's life course (Schmälzle et al. (2019), Boeri and Brugiavini (2008)). Next, we introduce a novel health measure that evaluates the reduction in working capacity, capturing the most relevant aspect affecting labour supply decisions. Furthermore, this measure encompasses various intermediate values, allowing differentiation between different levels of limitations. Again, this approach permits a complete depiction of the disabled population in Western European countries throughout their entire life-cycle, which cannot be achieved using any administrative data.

To capture the complexity of the data and the path dependency we rely on a multiple-state Transition Probability Model (TPM) that estimates the probabilities of transiting between different states of the world, e.g., from work to retirement or from work to disability and then to retirement. In the second step, the estimated TPM is combined with Monte Carlo simulations to uncover the most likely labour market histories (Gritz and MaCurdy (1992), Brugiavini et al. (2020)). In addition to the previous literature, we estimate the TPM by making use of Extreme Gradient Boosting (XGBoost, Chen and Guestrin (2016)), a supervised learning methodology that allows generating several counterfactuals in a parsimonious and more accurate way.

Results confirm the substitutability hypothesis between the early pension and disability insurance programs (Behaghel et al. (2012)). In detail, the spillover is more relevant when disability insurance schemes are reformed. Moreover, simulations show that tightening eligibility conditions to access disability benefits may have detrimental effects on the extensive margin of labour supply. The reason why is that workers with disabilities are crowded out towards alternative compensating schemes (in which working activities are not expected). However, ignoring behavioural responses following any given reform is not realistic. Particularly, in a world in which the true underlying health conditions cannot be perfectly observed, a policy tightening might force individuals with over-assessed Reductions in Working Capacity to update their labour supply. The effect on the extensive margin of labour supply is positive only in the case in which the share of disabled individuals who are able (and want) to increase their working capacity is high enough.

This work has straightforward policy consequences. The impact of disability insurance policy reforms depends on the efficiency of the assessment procedures. Generalized DI policy restrictions may incentivize labour participation of individuals with over-assessed limitations at the cost of penalizing those people with correctly assessed constraints not able to access anymore the benefits. Hence, carefully designed disability schemes should focus more on the early processes to correctly assess and employ the working capacities of individuals with disabilities. Possible real-world (successful) reforms may involve increasing employer involvement (the Netherlands) or increasing focus on early intervention (Switzerland).

The chapter is organised as follows. In section 2 literature is outlined by focusing on retirement pathways and all relevant factors in retirement decisions. Next, section 3 presents the overview of European institutional design and the microdata. Then, section 4 delivers the data set with particular attention to the new health-disability measure of reduced working capacity. Section 5 explains the empirical strategy. Thereafter, section 6 outlines the results. Subsequently, section 7 extends the baseline model to include behavioural responses. In the end, section 8 concludes.

2 Literature review

Retirement constitutes a complex phenomenon that involves both the individual's socio-economic environment and the historical-institutional context. As individuals age, various alternative pathways to exit the labour market and retire, even before the statutory age, become available (for a complete overview of retirement decisions and labour supply see Blundell et al. (2016) or Coile and Milligan (2016)). Euwals et al. (2011) defines the different pathways to retirement as a combination of institutional arrangements to manage the transition process between the exit from work and entry into the pension system. The existing literature prominently emphasizes the significant role of social security financial incentives in shaping labour market participation and retirement decisions (among others, refer to Belloni and Alessie (2013)). However, while the financial aspect undoubtedly holds significance, it is not the sole determining factor. For instance, empirical evidence presented by Blundell et al. (2016) highlights that couples often tend to retire together. Furthermore, Browne et al. (2019) offers a systematic review of the impact of the workplace psycho-social environment on retirement decisions. Again, as people age, their health naturally declines, making it a crucial factor in retirement decisions (French and Jones (2017)). Deciding which retirement pathway to go through is not only a matter of choice but rather reflects the life chances and resources that a person has accumulated over her/his life course Schmälzle et al. (2019)). For instance, Boeri and Brugiavini (2008) document that women are less responsive to pension rules since the gaps in women's careers do not allow them to mature eligibility conditions.

Another noteworthy scenario pertains to extreme health shocks that can significantly diminish an individual's working/earning capacity, potentially leading them to turn to disability insurance schemes as an alternative early retirement pathway. Kerkhofs et al. (1999) were among the pioneering authors to express concerns regarding the utilisation of the disability insurance scheme as an alternative early retirement pathway in the Netherlands. A more recent work of Koning and Lindeboom (2015) points out that no less than 12% of the entire Dutch population in the eighties and nineties perceived a Disability Insurance benefit. The reason why is to find in the leniency of legislation requirements that created incentives to use Disability Insurance to exit from the labour market and to wait until age conditions allow access to pension benefits. In the early 2000s, the Dutch out-of-control disability program pushed towards reforms that reduced employers' incentives to move workers to disability and tightened eligibility criteria for individuals. The necessity to keep under control disability expenditures pushed the Dutch Ministry of Social Affairs to develop a dynamic micro-simulation model which models transitions in disability which is fully integrated into the budget process (Van Sonsbeek and Alblas (2012)). This is relevant since the forecasting models which are proposed make use of a similar framework to our model with the difference that the supervised learning methodology which we are proposing is generally more suitable when dealing with prediction tasks. Following that, Ben Jelloul et al. (2023) presents another related simulation approach. Specifically, they introduce a microsimulation designed to model transitions between disability states and aim to simulate various scenarios depicting the evolution of the French elderly disabled population.

Concerns about the misuse of Disability Insurance as an alternative route for early retirement have spread in different countries. For instance, in Canada, Gruber (2000) finds a sizable labour supply response to a Disability Insurance policy change. In Finland, Kyyrä (2015) investigates the effects of disability pension eligibility tightening on labour supply. The results show that the policy increased the average age at which workers leave employment by almost 4 months. Another interesting institutional design regarding disability to investigate is Austria. For instance, Staubli (2011) documents that a policy change that tightened eligibility criteria significantly reduced disability enrollment. In addition, the policy had a relevant spillover effect on unemployment and sickness programs which experienced relevant growth. This supports two relevant thoughts. Firstly, when a policy is designed, all possible alternatives are important. Secondly, disability programs provide a safety net for individuals who hardly would move back into employment. The rigidity in the labour supply of disabled individuals can be found also in Switzerland. For instance, Deuchert and Eugster (2019) point out that a tightening in disability benefits does not influence the employment and income of targeted disabled individuals due to caseworkers which readjusted the degree of reduced earning capacity. Another interesting country to consider is Germany. In particular, Hanel (2012) analyses how a 2001 policy change in disability pension benefits affected early retirement through disability. The results suggest that benefit levels have no effect on labour supply. Another valuable contribution to mention is Börsch-Supan (2007) which shows that differences in disability benefits uptakes between a bunch of European countries depend mainly on institutional factors, i.e., eligibility conditions and benefits generosity. Again, in front of demographic imbalance and concerns about the future sustainability of pension systems, most of the previously mentioned studies focus on the disincentive effects of Disability Insurance. Nevertheless, Low and Pistaferri (2015) recall that the problem should be seen as a (dis)incentive-insurance trade-off. By making use of a life-cycle framework they show that in the US welfare would increase in case of less stringent eligibility conditions or more generous disability benefits, despite worse work incentives.

Finally, this paper exploits supervised learning, a well-known technique part of the machine learning literature, within a social science framework. In detail, supervised learning is a widely used methodology for predictive tasks in general, however, its applicability in social science is largely debated due to limited interpretability and availability of statistical inference procedures. Still, whenever predictability is the main concern, supervised learning provides substantial improvements over the baseline procedures with a little additional cost regarding theoretical sophistication. In detail, this paper aims at predicting transitions between different states of the world. The baseline approach to deal with this goal is to set up (multinomial) logit regressions and eventually to manually fine-tune non-linearities and high-order interactions to improve accuracy (Brugiavini et al. (2020)). However, these procedures may be time-consuming and require customized tasks, e.g., smoothing splines. The way in which this task could be overcome is by employing off-the-shelf supervised learning algorithms that are already exploited in other branches of literature. For instance, Albanesi and Vamossy (2019), show that machine learning techniques are performing significantly better in predicting consumer default than standard credit scoring models, even though the same data is used. Bargagli-Stoffi et al. (2021) show how supervised learning tools effectively predict relevant steps of future firm dynamics. Again, Gu et al. (2020) document that measuring asset risk premiums with machine learning tools may even double the performance of leading regression-based strategies from the literature. In addition, recent developments (SHAP values Lundberg and Lee (2017)) allow the supervised learning framework to provide more interpretable results.

3 Data

We rely on two sources of data. Firstly, we gather all the detailed administrative eligibility criteria for access to disability insurance benefits, early and old-age pension schemes for the eleven countries from 1970 onward. Secondly, we gather detailed microdata containing long retrospective panels, providing information for each year of individuals' lives.

3.1 Administrative eligibility criteria

In this section, we briefly outline the main characteristics of institutional settings related to disability insurance, early, and old-age pensions. The description aims to provide a general overview focusing on the characteristics which are valid for the majority of individuals in each country[3].

While analysing the pension systems of the countries under study, we observe various differences, but some intriguing points emerge. Firstly, all of them operate as pay-as-you-go systems. Secondly, a majority of these countries predominantly depend on the first pillar. Despite significant variations, each country offers early retirement pathways. Another common characteristic between countries is that disability benefits recipients are automatically moved into the old-age scheme in almost all countries under analysis. In the appendix, section C offers a comprehensive discussion for each country, directing specific external references for further information. It is noteworthy to recall that, in the last years, countries started to tighten eligibility rules with the idea to extend working

[3] It is important to highlight that it is difficult to fully document each institutional setting due to small and numerous differences in cohorts, work category-task, gender, and so on.

life and postpone retirement. Table 1 presents the comprehensive minimum eligibility conditions for accessing old age and early pension benefits in each country. Notably, there has been a general trend across almost all countries to raise the statutory retirement ages and increase the minimum contributory years needed to qualify for early retirement schemes. As of 2017, the majority of countries (with some exceptions) set the minimum age for accessing old age pensions at no less than 65 (60) years for men (women) and no less than 60 years for early retirement.

Next, defining administrative eligibility criteria for disability in real-world scenarios is a complex task. The roots of disability assessment protocols can be traced back to ancient civilisations such as the Sumerians, Greeks, and Romans, who used evaluation tools to determine the economic value of health-impaired slaves (Dal Pozzo et al. (2002)). In the 17th century, François Barème, a French mathematician, introduced a table of ordered percentage values for bodily damage (Baremas), initially utilised for war pensions and later extended to impaired workers' compensations in the last century. In modern times, the scope of disability assessment has expanded to encompass not only evident physical injuries but also less visible impairments like backaches or mental health conditions such as depression. Moreover, disability is no longer limited to health-related dimensions; it is now associated with factors like labour market conditions. Table 2 presents the eligibility conditions for accessing disability insurance benefits[4]. This table offers a general overview applicable to the majority of individuals in each country. However, additional important features are available in the references provided in section C. The primary conditions for disability insurance eligibility revolve around the reduction in working capacity, minimum contributory years, and age requirements. Notably, compared to early and old age retirement schemes, disability insurance has undergone fewer reforms. A common characteristic is the requirement for fewer contribution years. The extensive time span covered by this information enables us to accurately attribute detailed eligibility conditions for each age of the sample's life.

[4]The paper by Low and Pistaferri (2020) highlights additional dimensions of the disability insurance policy that are not addressed in this study, with the main focus being on eligibility.

	Old-age				Early					
	Men		Women		Men			Women		
Country	Period	Age	Period	Age	Period	Age	N_W	Period	Age	N_W
Austria	1961-2017	65	1961-2017	60	1961-2000	60	35	1961-2000	55	35
					2001-2003	61	35	2001-2003	56	35
					2004-2007	62	37.5	2004-2007	57	37.5
					2008-2011	63	37.5	2008-2011	58	37.5
					2012-2015	64	37.5	2012-2015	59	37.5
					2015-2017	65	37.5	2015-2017	60	37.5
Belgium	1961-2017	65	1961-1999	61	1961-1998	60	20	1961-1998	60	20
			2000-2002	62	1999	60	24	1999	60	24
			2003-2005	63	2000	60	26	2000	60	26
			2006-2008	64	2001	60	28	2001	60	26
			2008-2017	65	2002	60	30	2002	60	28
					2003	60	32	2003	60	30
					2004	60	34	2004	60	32
					2005-2012	60	35	2005-2012	60	35
					2013	60	38	2013	60	38
					2014	61	39	2014	61	39
					2015-2016	62	40	2013	62	40
					2017	62	41	2017	62	41
Denmark	1961-2003	67	1961-2003	67	1979	60	5	1979	60	5
			2004-2017	65	1980-1991	60	10	1980-1991	60	10
					1992-2004	60	20	1992-2004	60	20
France	1961-1994	65	1961-1994	65	1972-1994	60	15	1972-1994	60	15
	1994-2013	60	1994-2013	60	1995-2007	-	-	1995-2007	-	-
	2013-2017	62	2013-2017	62	2008-2011	56	32	2008-2011	56	32
					2012-2016	56	40	2012-2016	56	40
					2017	57	40	2017	57	40
Germany	1961-2016	65	1961-2016	65	1970-1971	60	35	1970-2017	60	35
	2017	66	2017	66	1972-2017	63	35			
Greece	1961-1974	65	1961-1974	60	1970-2007	60	15	1970-2007	55	15
	1975-1998	62	1975-1998	57	2008-2017	62	15	2008-2011	57	15
	1999-2013	65	1999-2013	60				2012-2013	58	15
	2014-2017	67	2014	65				2014-2017	62	15
			2015-2017	67						
Italy	1961-1993	60	1961-1993	55	1961-1995	45	35	1961-1995	45	35
	1994	61	1994	56	1996-1997	52	35	1996-1997	52	35
	1995	61.5	1995	56.5	1998	54	35	1998	54	35
	1996	62	1996	57	1999-2000	55	35	1999-2000	55	35
	1997	63	1997	58	2001	56	35	2001	56	35
	1998	63.5	1998	58.5	2002-2007	57	35	2002-2007	57	35
	1999	64	1999	59	2008-2009	58	35	2009-2009	58	35
	2000-2011	65	2000-2011	60	2010	59	36	2009	59	36
	2012	66	2012	66	2011	60	36	2010	60	36
	2013-2015	66.25	2013-2015	66.25	2012-2017	62	42	2012-2017	62	41
	2015-2017	66.5	2015-2017	66.5						
Netherlands	1961-2011	65	1961-2011	65	1972-1995	60	10	1972-1995	60	10
					1996-2008	62	35	1996-2008	62	35
Spain	1961-2017	65	1961-2017	65	1972-1982	64	15	1972-1982	64	15
					1983-1993	60	15	1983-1993	60	15
					1994-2013	61	30	1994-2013	61	30
					2014-2017	61	33	2014-2017	61	33
Sweden	1961-1994	67	1961-1994	67	1963-1969	63	15	1963-1969	63	15
	1995-2011	65	1995-2011	65	1970-1997	60	15	1970-1997	60	15
	2012-2017	67	2012-2017	67	1998-2017	61	15	1998-2017	61	15
Switzerland	1961-2017	65	1961-1975	63	1972-2006	62	15	1972-2004	60	15
								2005-2006	61	15
			1976-2003	62	2007-2017	63	15	2007-2017	62	15
			2004	63						
			2005-2017	64						

Table 1: Eligibility rules for old-age and early pension schemes 1961-2017. Source: MISSOC and the country-specific sources in section C.

Country	Period	% of Reduction earning/working capacity	Contribution years	Age requirements
Austria	- 2017	50	5 in the last 10	
Germany	- 2017	20	Min 5 , 3 in the last 5	
Sweden*	- 2017	25	Min 8 if age <= 46 Min 7 if 47 <= age <= 49 Min 6 if 50 <= age <= 52 Min 5 if age <= 53	age <= 64
The Netherlands	- 2005	15		age <= 65
	2006 - 2017	35		age <= 65
Spain	- 2017	66	$\frac{1}{3}$ of years between 16 and disability	age <= 30
	- 2017	66	$\frac{1}{4}$ of years between 20 and disability (min 5)	age >= 31
Italy	- 2017	66	Min 5, 3 in the last 5	18 <= age <= 65
France	- 2017	66	1	age <= 60
Denmark	- 1966	66		18 <= age <= 60
	1967 - 1983	50		18 <= age <= 60
	1984 - 2002	50		60 <= age <= 66
	1984 - 2002	66		18 <= age <= 60
	2003 - 2017	75		age > 40
	2014 - 2017	60		age >= ret age - 5
Greece	- 2017	50	1 every 3, after age 21	18 <= age <= 65
Switzerland	- 2017	40	3	
Belgium	- 2017	66	1	age <= 65

Table 2: Minimum eligibility criteria for disability benefits eligibility. *Sweden: 3 years of residence. Source: MISSOC and the country-specific sources in C.

3.2 SHARE Data

The "Survey of Health, Ageing, and Retirement in Europe" (SHARE, Börsch-Supan et al. (2013)) serves as a valuable source of microdata, comprising the most extensive European social science panel. This comprehensive survey offers internationally comparable longitudinal data, facilitating in-depth analyses of the public health and socio-economic living conditions of individuals across Europe. Initiated in 2004, SHARE conducts biannual surveys, focusing on a nationally representative sample of European individuals aged 50 and above.

SHARE employs both regular and retrospective interviews. The regular waves focus on gathering current data about the socio-economic aspects of the participants' lives, offering a comprehensive perspective on the elderly European population. Additionally, waves 3 and 7, known as SHARELIFE interviews, collect life-history information by asking respondents to recall conditions from significant points in their past. This retrospective approach attempts to emulate longitudinal data that would track individuals throughout their entire lives. However, due to substantial time and cost constraints, achieving such complete longitudinal data is clearly impractical[5]. The SHARELIFE data is then reorganised as a long retrospective panel in which each respondent contributes as many observations as there are years from age of birth to age in the interview year (see Antonova et al. (2014) and Brugiavini et al. (2019)). Being able to exploit such a wide source of information allows for capturing all the lifetime relevant events which may influence retirement decisions later in life.

Due to the challenges associated with accurately measuring disability and the constraints of other available options, many datasets are constrained to utilise only a limited set of health measures, thus capturing only a partial aspect of health. As Gustman and Steinmeier (2018) highlighted, extensive surveys incorporate both objective and subjective health measures, significantly enhancing the opportunities for investigating the relationship between health and human behaviour. Utilising retrospective data allows for the construction of a health-labour measure that spans an individual's entire life, revealing long-term dependencies emphasised in the cumulative-(dis)advantage theory.

The final sample contains 36,770 individuals who worked for at least one year and have been interviewed in Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland or Belgium. Since we are able to observe each year of life of the individuals, we can rely on 2,022,440 observations[6]. The detailed sample selection procedure is explained in section B.

4 Data modelling

In this section, we present an overview of the various states that individuals can transition through, i.e., Schooling, Work, Other, Disability Work, Disability Other, and Retirement. A significant portion of the discussion focuses on modelling the disability state, where we carefully examine and construct different degrees of Reduced Working Capacity (RWC).

[5]Aside from SHARE, a comparable retrospective approach is adopted in the United States through the "Health and Retirement Survey" (HRS) and in the United Kingdom through the "English Longitudinal Study on Ageing" (ELSA).

[6]Only observations within the age range of 10 to 70 are retained.

Furthermore, we ascertain the eligibility variables for individuals at each age by utilising individual information in conjunction with administrative eligibility requirements. Using SHARE survey data, we can provide a complete picture of the disabled population that goes beyond what can be observed solely with administrative data.

4.1 States of the world

It is worth mentioning that, as pointed out by Brugiavini et al. (2019) and Antonova et al. (2014), the SHARE Job Episode Panel allows for the precise identification of detailed and mutually exclusive labour market states for each year of age among the respondents.

To explore the influence of disability on the labour market and retirement decisions, we model transitions between the following mutually exclusive states of the world: Schooling (S), Work (W), Other (O), Disability Work (DW), Disability Other (DO) and Retirement (R). In this model, individuals start with an exogenous schooling path (at least until age 10). The initial modelling assumption asserts that once an individual exits Schooling, there are no possible transitions back into Schooling. The state "Work" contains both employed or self-employed individuals, including those working for the family businesses. "Other" is the residual state which contains mainly individuals who declare themselves "Out of the labour force" or "Unemployed". These two states are merged since otherwise it would be difficult to model the very few transitions from and to unemployment[7]. To determine the disability and labour market conditions we leverage a set of different survey questions to construct mutually exclusive states (see subsection 4.2). Finally, "Retirement" is an absorbing state which follows Bonsang and Klein (2012), i.e., the individual is definitively out of the labour force with the intention of staying out permanently[8]. In addition, since retirement in my model is considered an absorbing state, to access it, the individuals should be at least 45 years old. One limitation of our approach is that data allows us to consider only the extensive margin.

[7]The reason why is due to the fact that data provides only long-term unemployment, i.e., unemployment spells lasting more than 6 months.

[8]Hence, by following this definition, the reception of an early/old-age pension is not a necessary condition to be classified as retired since what really matters is the perception of individuals. In such a way, a disabled individual who receives Disability Insurance benefits with the idea of never coming back to work is considered retired.

4.2 Disability state

We define disability status with a particular focus on how health impacts individuals' employability. This emphasis arises because a mere health measure does not fully encapsulate the ability to participate in gainful employment. In more formal terms, our disability state is defined as "an impairment (such as a chronic medical condition or injury) that prevents someone from engaging in gainful employment" (Merriam-Webster (2022)).

It is crucial to emphasise that the disability state centres around a health-labour measure of reduced working capacity and does not directly align with the receipt of Disability Insurance benefits. In particular, we utilise our data to initially identify the disability state and subsequently ascertain the extent of limitations.

Specifically, we use the following SHARE questions:

- Q1) The respondent reported a situation in "Sick or disabled"
- Q2) Have you ever received a physical injury that has led to any permanent handicap, disability or limitations in what you can do in daily life?
- Q3) Did you ever take a temporary leave of absence from a job for 6 months or more because of ill health or disability?
- Q4) Did you ever leave a job because of ill health or disability?
- Q5) How much did ill health or disability limit your ability to work
- Q6) After giving up this job, did you take up a job in which ill health or disability limited your ability to work to a lesser extent?

The retrospective nature of our data enables the distinction of the year in which individuals self-declare themselves in a "Sick or disabled" situation for each age of life. Nevertheless, due to the stigma associated with self-identifying as disabled and constraints posed by the survey structure[9], we have chosen to adopt a "broader" definition of disability. This broader definition incorporates other disability-related events to capture a more realistic representation of the prevalence of disabilities in the population. It is important to note that these questions do not solely pertain to health manifestations; instead, they encompass a blend of labour market conditions. Besides the explicit inquiry regarding

[9]The survey structure requires respondents to select only one state of the world for each year of life. However, since the list includes other states like working or housekeeping, there is a possibility of overlapping declarations with the category of "Sick or disabled." See section D for the complete list of states the survey structure proposes.

physical injuries resulting in limitations in daily life activities, we also utilise questions concerning job limitations to comprehensively assess the situation.

Table 3 displays the distribution of observations based on different states of the world. The majority of observations (over 65.51%) correspond to working states. About 2.58% of the observations represent individuals in working conditions with disabilities, while slightly over 1.55% are individuals who are out of the labour market due to severe health impairments. Regarding retirement observations, since retirement is an absorbing state, we retain only the first year. Preserving the first year of retirement is essential because it could be the year of eligibility and also serve as an early pathway to retirement, where individuals start receiving benefits.

JEP	Disability State	State of the World			Observations	Percent
Schooling	ANY	**Schooling**	(S)		319142	16.76
Work	Not Disabled	**Work**	(W)		1247811	65.51
Work	Disabled	**Disability Work**	(DW)		238613	12.53
Other	Not Disabled	**Other**	(O)		49079	2.58
Other	Disabled	**Disability Other**	(DO)		29737	1.55
Retired	ANY	**Retired**	(R)		20290	1.07
				Total	1904672	100

Table 3: Observations by states of the world. S stays for Schooling, W for Work, O for Other, DW for Disability Work, DO for Disability Other, and R for Retirement.

As depicted in Figure 1, disability is a relatively common occurrence throughout a person's lifetime. Indeed, the lowest proportion of individuals who have experienced at least one disability event is observed in Switzerland (15%), while this fraction is nearly one-third in the Netherlands. As highlighted by Kapteyn et al. (2007), the variation in cultural and institutional factors plays a role in how individuals from different countries respond to disability-related questions. To address this issue, the broader measure of disability will be integrated with the aforementioned eligibility conditions from administrative records. The hypothesis is that cultural and historical influences have shaped both administrative conditions and respondents' attitudes, and therefore, simulations on eligibility criteria should account for the variations in how different cultures approach answering.

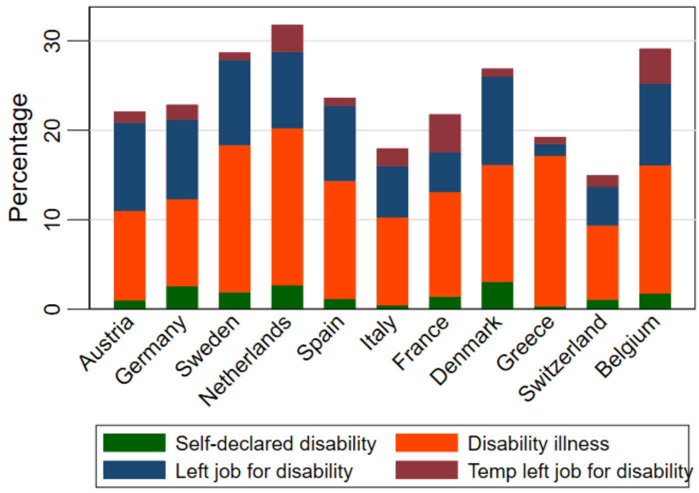

Figure 1: Fraction of individuals who experienced at least one disability event during their lifetime.

Next, in Figure 2, we validate our disability measure by comparing it with data on Disability Insurance benefits recipients from WHO (2021) and OECD (2021). As mentioned earlier, it is evident that the strict disability state underestimates the true value of disabled individuals. As a matter of fact, considering that not all disabled individuals can attain the benefit, the bias is likely to be substantial. Taking into account the broader disability measure, we notice that it consistently yields higher levels compared to the data of Disability Insurance (DI) recipients. Nonetheless, this is a reassuring finding, as individuals with disabilities may not always be eligible for the DI benefits due to various reasons, including a lower disability degree with the minimum required, income exceeding the maximum threshold in a means-tested program, or the availability of alternative support programs.

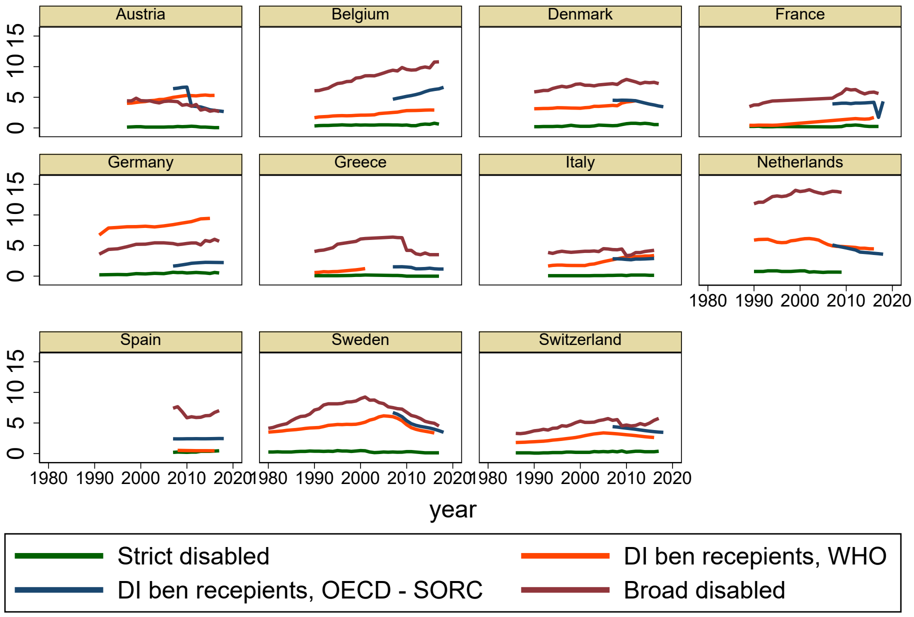

Figure 2: Comparison between "strict" and "broad" disability states and disability benefits recipients using World Health Organization and OECD data.

In Section D.1, we further validate our states by comparing them with additional information collected from previous regular interviews.

4.3 Reduced working capacity

We further leverage our data to develop a new measure of reduced working capacity[10] (RWC). Specifically, we utilise the questions in section 4.2 to determine the degree of disability, enabling a more in-depth examination compared to a simple disability state indicator.

Our approach is guided by the following rules:

- 100% - If Q1 reports "Sick or disabled"
- 100% - If Q3 reports temporary leave of job for disability
- If Q4 reports leave of job for disability. Q6 reports that the new job does not increase working capacity. Q5 assumes
  1. 25% - Very little
  2. 50% - Somewhat

[10]While "earning" and "working" capacity may have slightly different interpretations across different countries, they both encompass a similar aspect: the extent to which an individual has decreased their potential ability to work. Consequently, we employ the term "working" in our approach.

3. 75% - Very much
4. 100% - Could not do job anymore

- If Q4 reports leave of job for disability. Q6 reports that the new job does increase working capacity. Q5 assumes
  1. 0% - Very little
  2. 25% - Somewhat
  3. 50% - Very much
  4. 75% - Could not do job anymore
- If Q4 reports leave of last job for disability. Q5 assumes
  1. 25% - Very little
  2. 50% - Somewhat
  3. 75% - Very much
  4. 100% - Could not do job anymore
- For the disability illnesses in Q2, define a % following the procedure presented below
  1. Restrict the sample only to those observations with both disability illnesses and other disability events.
  2. Use the obtained restricted sample to regress the percentage of the reduced working capacity on the dummies of disability-health events.
  3. Use the coefficients in table 4 to predict the % of reduced working capacity.
  4. The % of reduced working capacity is assigned according to which of the 4 above-mentioned thresholds are overcome (e.g., a predicted value of 77 is assigned to the 75% of reduced working capacity).
- In case of overlaps, take the smallest one

Table 4 presents the results of the procedure to define the percentages of reduced working capacity for those who declare disability illnesses. Coefficients can be roughly interpreted as the average reduction of working capacity given any of the disability illnesses reported. For example, a reported heart attack in men causes on average a 41% reduction in working capacity[11]. Most relevant diseases which cause disability such as back pain, heart attack, stroke, cancer, psychiatric and so on are well captured by the estimation. A perfect estimation would produce coefficients in the 0-100 range. The only

[11]The reduced working capacity computed with generic coefficients or by using gender-specific ones does not change the final outcomes.

negative signs (in column (1)), which are not desirable, are headaches and allergies which are likely to be part of comorbidity and in any case, these diseases alone do not assign relevant disability recognition.

The whole procedure outlined above spans percentages of reduced working capacity as in table 5. Several points deserve particular attention. As expected, the percentages of reduced working capacity have different distributions according to which disability status is considered. Working disabled show a more uniform distribution in working capacity reduction, whereas those not in the labour market show a more skewed (towards more extreme cases) reduction of working capacity. Few individuals experience a disability event which does not assure at least 25% of reduced working capacity. In that case, reduced working capacity is set to 0 even though it is in the 0-25% range. Moreover, a similar skewness can be seen for retired individuals, i.e., on the one hand, worse health conditions "force" individuals to retire, on the other hand, older retired individuals have worse health. One concern is that the overall reduction of working capacity is right skewed, whereas it can be naturally expected a large fraction of small health impairments and a small fraction of large impairments. One explanation could be the "justification bias" Butler et al. (1987), however, an alternative explanation may be that very low impairments are not reported until a given threshold is overcome. In any case, since the identification of disability and its severity are important steps, additional robustness checks have been performed in the appendix.

	(1)	(2)	(3)
	All sample	Males	Females
Back Pain	42.05***	41.34***	42.05***
	(0.835)	(1.262)	(1.094)
Arthritis	26.22***	24.51***	31.06***
	(1.005)	(1.701)	(1.228)
Osteoporosis	3.011	-0.672	6.051**
	(1.675)	(3.515)	(1.908)
Heart Attack	41.25***	44.10***	26.26***
	(1.330)	(1.521)	(2.663)
Other Heart	19.58***	35.67***	-14.48***
	(1.356)	(1.633)	(2.339)
Diabetes	26.81***	34.48***	18.78***
	(1.372)	(1.915)	(1.918)
Stroke	38.05***	40.50***	33.09***
	(1.538)	(1.903)	(2.529)
Asthma	27.59***	32.31***	27.74***
	(1.910)	(2.994)	(2.462)
Respiratory	25.64***	25.01***	21.60***
	(1.418)	(1.893)	(2.099)
Headaches	-8.055***	-14.39***	-0.209
	(1.586)	(2.850)	(1.924)
Cancer	56.50***	56.96***	56.82***
	(1.271)	(2.094)	(1.567)
Psychiatric	44.88***	42.89***	47.67***
	(1.108)	(1.915)	(1.346)
Multiple Sclerosis Fatigue	17.28***	29.41***	11.00***
	(1.381)	(2.216)	(1.737)
Gynaecological	9.594***	-23.17	16.41***
	(1.937)	(17.79)	(1.989)
Eye	8.772***	9.994***	8.439***
	(1.535)	(2.317)	(2.010)
Infectious	34.00***	40.65***	31.67***
	(2.572)	(3.874)	(3.427)
Allergies	-18.47***	-17.20***	-19.28***
	(2.110)	(3.592)	(2.629)
Other	61.95***	64.49***	59.85***
	(0.799)	(1.091)	(1.133)
Observations	16414	7750	8664
R^2	0.698	0.714	0.706

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Linear regression of percentage of the reduced working capacity on disability-health dummies. In column (2) only males whereas in column (3) only females. Tuberculosis and leukaemia are reported rarely (less than 1%) and so they are aggregated to Other.

RWC	S	W	O	DW	DO	R	Total
0	318177	1247811	238613	3944	1014	17105	1789894
	(99.7)	(100)	(100)	(8)	(3.4)	(84.3)	(95.8)
25	276	0	0	10510	4962	711	16459
	(0.1)	(0)	(0)	(21.4)	(16.7)	(3.5)	(0.9)
50	304	0	0	12622	5332	761	19019
	(0.1)	(0)	(0)	(25.7)	(17.9)	(3.8)	(1)
75	175	0	0	8282	4298	527	13282
	(0.1)	(0)	(0)	(16.9)	(14.5)	(2.6)	(0.7)
100	210	0	0	13721	14131	1186	29248
	(0.1)	(0)	(0)	(28)	(47.5)	(5.8)	(1.6)
All	319142	1247811	238613	49079	29737	20290	1904672

Table 5: Reduced Working Capacity observations according to the states of the world. S stays for Schooling, W for Work, O for Other, DW for Disability Work, DO for Disability Other, and R for Retirement. Values show observations and relative percentages in brackets. For retired individuals is considered only the first observation.

Figure 3 illustrates the significant presence of country heterogeneity. First of all, "northern" countries are characterised by larger percentages of observations in disability states. The observed variations can be attributed to the differing cultural attitudes and variations in legislation leniency across countries. It is possible that the fraction of individuals in very bad health situations might be over-reported. Along with the above comments on the expected distribution of disabled, the 100% of individuals with reduced working capacity seems to contain also a relevant fraction of individuals in the 75% level. Although this raises some concerns, this limitation does not affect eligibility for disability benefits and in any case, those in the 75% category are severely impaired individuals which would be hardly distinguishable from the 100% group. Although this disability measure has as an ideal goal to represent the "real state of the world" regarding disability, as in the real world, it is likely to contain also false positives, i.e., relatively healthy individuals are able to achieve benefits even if not completely fulfilling the necessary conditions (Low and Pistaferri (2020)). On the contrary, since no direct cost or medical proof is required to be assessed with an RWC, it is quite likely that the share of false negatives (i.e., individuals with severe medical conditions which are not reported) is negligible. This knowledge about screening stringency would be further exploited in policy simulations.

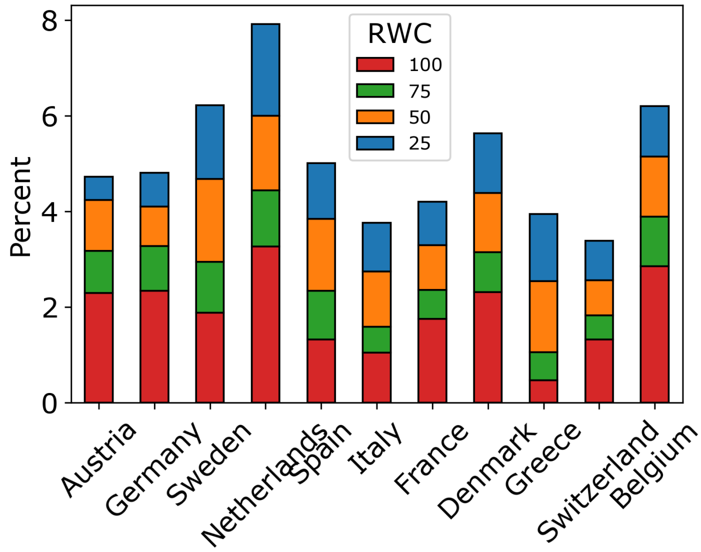

Figure 3: Percentage of observations of reduced working capacity (RWC) in each country.

4.4 Disability Insurance Benefits Eligibility

Up until now, we have established the different disability states and the extent of limitations using the RWC levels as our defining criteria. Our attention now shifts to our primary variable of interest, which is the eligibility for disability insurance benefits. To establish this, we utilize the comprehensive labour market history of each individual and country-year-specific administrative criteria. Specifically, we consider the following requirement for each individual at any given age:

1. Fraction of reduction of working capacity
2. Year of contributions
3. Age requirements
4. Additional requirements

A dichotomous variable is generated for each age of all individuals. This variable assumes value one if and only if the above conditions fulfil administrative conditions in the individual's country in any given year reported in table 2.

	S	W	O	DW	DO	R	Total
Not Eligible	319142	1247811	238613	19788	11567	18516	1855437
(% Not Eligible)	(100)	(100)	(100)	(40.3)	(38.9)	(91.3)	(97.4)
Eligible	0	0	0	29291	18170	1774	49235
(% Eligible)	(0)	(0)	(0)	(59.7)	(61.1)	(8.7)	(2.6)
Total	319142	1247811	238613	49079	29737	20290	1904672

Table 6: Disability benefit eligibility across the available states. S stays for Schooling, W for Work, O for Other, DW for Disability Work, DO for Disability Other, and R for Retirement. Values show observations and relative percentages in brackets. For retired individuals is considered only the first observation. Schooling, Work, and Other observations are not compatible with eligibility for disability benefits.

Overall, 49235 observations (around 2.6% of the overall sample) fulfil administrative requirements for disability insurance eligibility. As expected most disability eligibility is found in disability states. Any individual could have more than one observation which is eligible. One point to raise is the fact that this measure is not completely informative for retirement since RWC focuses on working capacity. Again, we recall that for retirement observations, table 6 keeps only the first year. Overall, women are slightly more eligible (2.7% of overall observations with respect to 2.4% of men observations). This is coherent with the tendency to report worse health status in the sample and quite plausibly due to the fact that men have more early retirement rights due to contributory histories.

In figure 4, the shaded areas present the fractions of observations in each country in our sample that are eligible for disability insurance benefits. By looking at country heterogeneity, it is possible to order the potential accessibility to the disability benefit across the sample. The simple arithmetic average of the percentages of observations eligible for all countries is 2.93. Only four countries, for which literature already pointed out the leniency in allowing disability benefits, are above the mean: the Netherlands[12], Sweden, Belgium and Germany. Mediterranean countries, such as Italy and Greece, have the lowest percentages, intuitively because more generous alternatives for early exits from the labour market have been allowed. Although this measure has the limitation to show only the potentiality to use this route and not its effective use, it is still able

[12]It is worthy to recall that for Netherlands data is available only in wave 3, i.e., only before 2009. Literature shows that recent trends in disability benefit access are declining.

to illustrate in which countries the disability route has been exploited as an early exit from the labour market. Moreover, the percentages of observations across different RWC levels offer additional evidence supporting the positive relationship between a country's tendency to report disability and the administrative eligibility criteria.

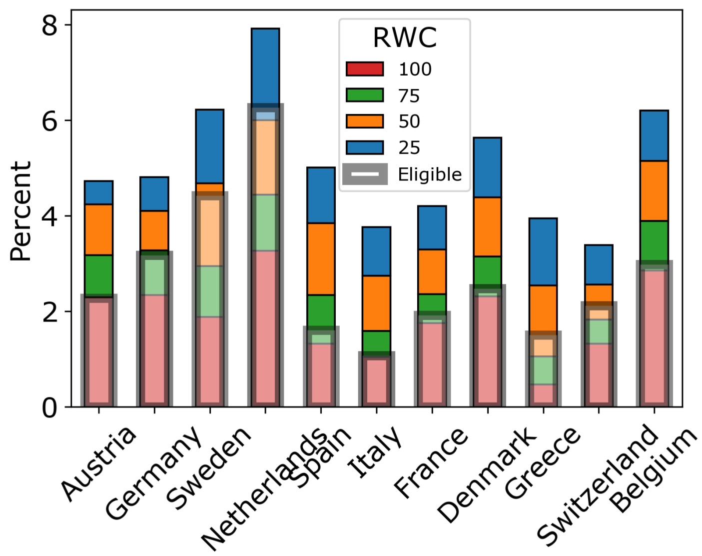

Figure 4: The shaded areas represent the proportion of observations eligible for disability benefits in each country. The colored stacked bars depict the distribution of observations based on the RWC levels.

5 Empirical strategy

It is worth recalling that the goal of this paper is to simulate pension policy reforms, with particular attention to eligibility to early pathways to retirement, i.e., early pension and disability insurance benefits. Since both processes of accumulation of contribution years and of health deterioration are highly path-dependent, the simulation strategy is set up in order to exploit the richness of the available microdata.

5.1 Transition Probability Model

The entire simulation framework is based on Gritz and MaCurdy (1992) and Brugiavini et al. (2020). In detail, we simulate each age of life of individuals starting from the year in which school ends until the individual is censored, retires or eventually becomes too old (70 years old). The model estimates three modular building blocks which provide the short-term relationships and then describes the likelihood of any given sequence of labour market episodes. In detail, the first module models the exit transition from school at the exogenous school-leaving age ($S \to \{W, O, DW, DO\}$). The second module estimates the discrete-time duration models for the length of spells in Work, Other, Disability Work, Disability Other ($\{W, O, DW, DO\}$). The third and last module models the exit transition across the labour market states and eventually retirement ($\{W, O, DW, DO, R\}$). After estimating the preceding building blocks, we conduct a sequential stochastic simulation to reveal the most likely labour market histories. The time metric we consider is the age at various stages of their life course.

To estimate the modular building blocks of the Transition Probability Matrix (TPM), we employ the eXtreme Gradient Boosting (XGBoost) estimator Chen and Guestrin (2016). XGBoost belongs to the supervised learning literature and is a prominent machine learning technique extensively used for predictive tasks. Specifically, XGBoost has demonstrated its effectiveness as a leading model for analysing standard socio-economic data, as supported by previous works such as Albanesi and Vamossy (2019), Bargagli-Stoffi et al. (2021), and Gu et al. (2020)[13]. In the appendix section E, we present more details about supervised learning and XGBoost. XGBoost, along with other supervised learning estimators, is a promising choice for this task due to several compelling reasons. Previous literature has traditionally estimated these blocks using parametric approaches, such as logit and multinomial logit models. However, more sophisticated supervised learning estimators are expected to outperform these standard methods. Additionally, to capture nonlinearities in the data, previous studies have relied on finely-tuned linear or cubic splines within the parametric framework. Despite their effectiveness, these models often lack interpretability due to their restrictions and nonlinear nature, making it challenging to interpret the estimated coefficients. In contrast, the growing demand for explainability in machine learning literature has led to recent developments in the literature on explainability. In this regard, we leverage the Shapley values Lundberg and Lee (2017) to offer insights into the section on explainability in the appendix (section G).

[13] At this stage it is worthy to mention Wolpert and Macready (1997) which states the "No free lunch theorem in optimisation". The theorem states that all optimisation algorithms perform equally well when their performance is averaged across all possible problems. This is to say that no methodology systematically outperforms the other alternatives in all the possible problems and so each tool must be used for the right problem.

5.1.1 Transition from school

In the initial stage of the model, each individual exogenously decides how much to remain in full-time education[14]. Afterwards, according to the individual's socio-economic characteristics, the individual transits into one of the mutually exclusive states which are available at younger ages, i.e., Working (W), Other (O), Disability Working (DW), Disability Other (DO).

The transition estimation probabilities are estimated via XGBoost:

$$Pr(S \rightarrow j|X_s) = \pi^{SCH}_{S,j}(X_s) = \pi^{SCH}_{S,j}, \tag{1}$$

where $\pi^{SCH}_{S,j}(X_s)$ is the optimal learner, i.e., an ensemble of shallow trees that matches as much as possible[15] the expected probability distribution and the "real" sample distribution. S indexes the origin state, i.e., School, whereas $j \in \{W, O, DW, DO\}$ indexes the admissible destination states. X_s is the vector of socio-economic explanatory variables at school-leaving age. In detail, X_s includes age, gender, marital status (indicated by a dummy variable), children born in the last three years, number of children, education level (categorised as low, middle, or high according to the ISCED classification), 7 cohort dummies representing different birth cohorts, country of residence, and the reduced working capacity measure. These variables are exogenously provided and are not influenced by the simulation exercise.

Figure 5 shows that the (in-sample[16]) transition probabilities estimated via XGBoost perfectly overlap the observed transition probabilities.

[14]Hence, the duration of schooling is taken as exogenous. Since the focus is posed on retirement decisions, once the predicted transitions out of school are satisfactory, endogeneising schooling decisions is not the main goal of this work.

[15]The matching between the estimated and sample probability distribution does not rule out the regularization part of the algorithm which assures generality by penalizing the complexity of the model.

[16]Out-of-sample transition probability provide almost identical results. Moreover, section 5.3 provides a more extended out-of-sample exercise.

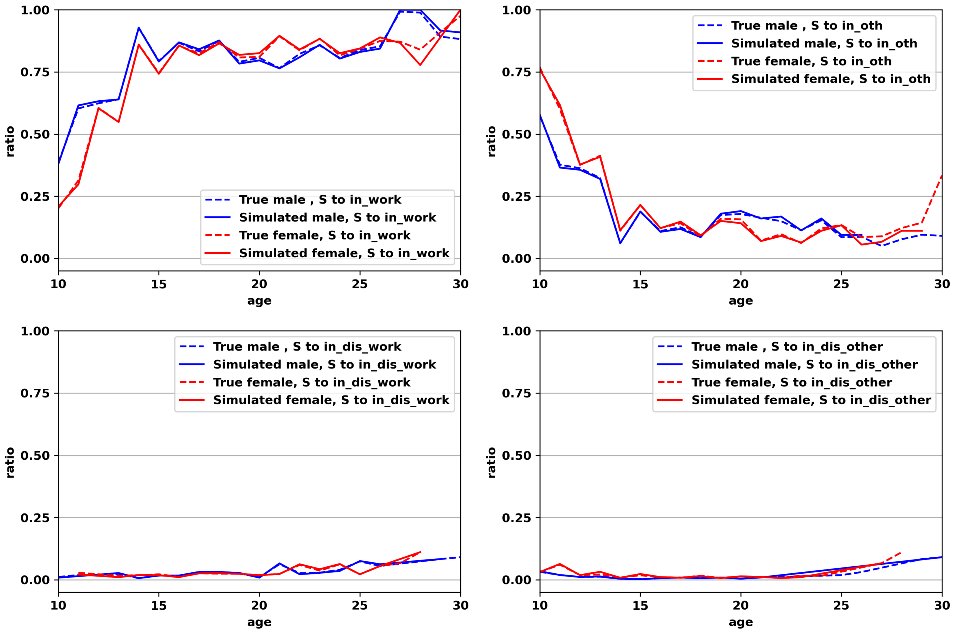

Figure 5: Observed and estimated age profiles of the initial transition probabilities.

5.1.2 Discrete time duration distributions

Once the individual leaves school, he/she moves into one of the admissible states. If for instance the individual moves into the Working spell, given socio-economic characteristics, labour market history and "luck"[17], he/she might continue working until the pension is awarded, or alternatively, have short spells duration that not allows maturing pension's eligibility conditions. This step is modelled by a set of discrete-time duration models for the number of years individuals spend in the proposed labour market states, conditioning on being in the same not absorbing state in the previous year.

Given a generic spell in state $j \in \{W, O, DW, DO\}$, let T_j be a discrete random variable for the duration of the spell in state j. T_j takes positive integer value τ with probability mass function

$$f_j(\tau) = Pr(T_j = \tau) = S_j(\tau - 1)\lambda_j(\tau), \quad (2)$$

where $S_j(\tau - 1) = Pr(T_j > \tau - 1) = \prod_{t=1}^{\tau-1}[1 - \lambda_j(t)]$ is the discrete-time survivor function whereas $\lambda_j(t) = Pr(T_j = t | T_j \geq t)$ is the discrete-time hazard function. The focus is on the "instantaneous survival function", i.e., the complement of the hazard $\pi_j(t) = 1 - \lambda_j(t) = Pr(T_j \geq t + 1 | T_j \geq t)$. The reason why is that $\pi_j(t)$ perfectly

[17] Due to the stochastic nature of the model, individuals with identical characteristics could experience entirely different life paths based on their initial transitions.

suits the problem at hand, i.e., it represents the conditional probability of continuing the current spell in state j, given that the individual has persisted in state j until time $t-1$.

Any discrete-time duration analysis is equivalent to a panel data approach for binary responses (Allison (1982) and Jenkins et al. (1995)). To model the instantaneous survival probability, a binary outcome variable y_{jt} is defined. It takes value 1 if the current spell in state j is still in progress at time t and 0 otherwise. To better explain the proposed framework, let's present some examples. For instance, an uncensored spell lasting $\tau > 1$ years yields a sequence $(y_{j,0} = 1, y_{j,1} = 1, \ldots, y_{j,\tau-1} = 1, y_{j,\tau} = 0)$. On the contrary, if the spell is censored at time t then the sequence becomes $(y_{j,0} = 1, y_{j,1} = 1, \ldots, y_{j,t-1} = 1, y_{j,t} = 1)$. Finally, a one-year spell yields a short sequence $(y_{j,0} = 1, y_{j,1} = 0)$.

As in the previous modular building block, the duration distribution is estimated via XGBoost:

$$Pr(T_j \geq t+1 | T_j \geq t) = \pi_j^{DUR}(X, H) = \pi_j^{DUR} \tag{3}$$

where π_j^{DUR} is the XGBoost learner. $j \in \{W, O, DW, DO\}$, X is the vector of socio-demographic characteristics, whereas H_t represents the vector of the history variables that are updated sequentially during the stochastic simulations. These variables capture the individual's past experiences in different states, including counting how many years the individuals spent in Work, Other, Disability Other, and Disability Work states. It also includes the duration of the current spell (in equation 3 it corresponds to t). Additionally, there are three dummy variables indicating whether the individual has spent at least one of the last three years in Other, Disability Other, and Disability Work states. Furthermore, considering the specific legislation of each country, there are three dummy variables representing eligibility for Disability Benefits, Early Pension, and Old-age Pension. Specifically, if the age and contributory requirements are met (see table 1) individuals become eligible for Early or Old-age pensions. Similarly, if the age, contributory requirements, and RWC met the conditions in table 2, individuals become eligible for Disability Benefits.

Again, evidence in figure 6 shows that the (in-sample) duration probabilities that are estimated via XGBoost follow almost the same exact pattern as the observed transition probabilities.

Figure 6: Observed and estimated age profiles of the instantaneous survival probabilities.

In this regard, it is worth mentioning that the proposed discrete-time duration analysis assumes that the lengths of subsequent spells are independent after conditioning on the explanatory variables X, H. It may seem a quite restrictive assumption that may rule out unobserved heterogeneity which leads to correlated durations of the spells, however, Heckman (1991) points out that the ability to distinguish between unobserved heterogeneity and duration dependence in single-spells duration models is based on maintaining explicit assumptions about how observables and unobservables interact (further discussion in Lancaster (1990) and Van den Berg (2001)). The key identification issue is that unobserved heterogeneity is indistinguishable from simply allowing for more flexible forms of duration dependence. Hence, instead of modelling unobserved heterogeneity, the problem is faced by proposing a flexible form of duration. Alternatively, unobserved heterogeneity could be identified by the information from repeated spells data on the same state (Honoré (1993) or Van den Berg (2001)). Nevertheless, in our data set only a small subset of individuals report multiple spells in the same state.

5.1.3 Exit transitions from current spell

Once individuals leave their current state, different alternative routes become available.

The probability of transiting from state j into $k, (k \neq j)$, conditioned on concluding a spell in state j, can be generally written as

$$Pr(j \to k|X_j, H_j) = \pi^{TR}_{j,k}(X_j, H_j) = \pi^{TR}_{j,k}, \tag{4}$$

where $\pi^{TR}_{j,k}$ is the XGBoost learner. j is the origin state and k the destination state whereas X_j and H_j are respectively the socio-demographic characteristic and the history variables. The available alternatives $k \in K$ change with age. In fact, in our model, it is assumed that retirement becomes available only for individuals aged at least 45.

Again, evidence in figure 7 shows that the (in-sample) estimation of the exit transition probabilities is correctly emulating the observed transition probabilities.

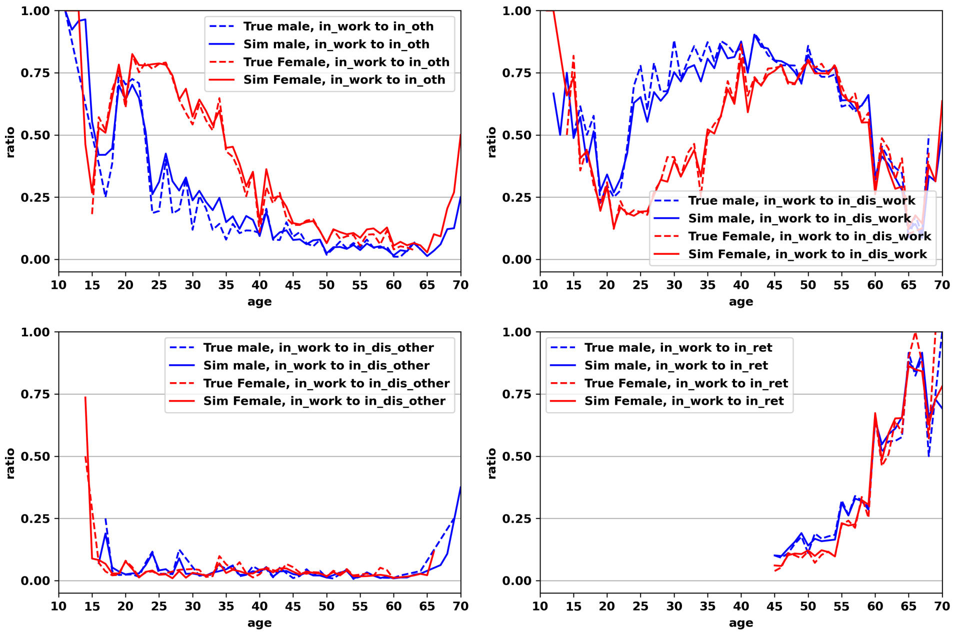

Figure 7: Observed and estimated age profiles of the exit transition probabilities.

5.2 Monte Carlo Simulations

Next, following Brugiavini et al. (2020), we build a Monte Carlo algorithm that uses the estimated probability distributions to predict the dynamic sequence of the labour market path of each individual.

Each individual's lifetime sequence is limited to a maximum length determined by their age at the year of the interview (or censored at age 70 if the interview took place afterwards). Then, we set up the exogenous schooling path for each individual. Finally, we employ the three steps algorithms described below to simulate the entire labour market career of each individual. This involves comparing independent draws from the uniform distribution with the estimated probabilities relevant to each age state. It is relevant to point out that socio-demographic characteristics are kept constant[18] for the whole path, whereas history variables[19] are updated at each point in time according to the simulated history.

The Monte Carlo prediction algorithm is implemented in Python. Each individual is indexed with i and aged with a.

1. Predict the initial transition probability from school, i.e., $S \to j \in \{W, O, DW, DO\}$ at the exogenously observed school-leaving age a_S.
   (a) Given the socio-demographic characteristics at school-leaving age X_{ia_S} and the supervised learner $\pi^{SCH}_{S,j}$ trained in the first building module, predict the probabilities of all admissible transitions from school, i.e., $\hat{\pi}_{i,S,j} = \pi^{DUR}_{Sj}(X_{ia_S})$, $j \in \{W, O, DW, DO\}$.
   (b) Get a pseudo-random draw ν_i from a uniform distribution.
   (c) Predict transition from school as

$$STATE_{i,a_S+1} = \begin{cases} W & if \;\; \nu_i \leq \hat{\pi}_{i,S,W} \\ O & if \;\; \hat{\pi}_{i,S,W} \leq \nu_i \leq \hat{\pi}_{i,S,W} + \hat{\pi}_{i,S,O} \\ DW & if \;\; \hat{\pi}_{i,S,W} + \hat{\pi}_{i,S,O} \leq \nu_i \leq \hat{\pi}_{i,S,W} + \hat{\pi}_{i,S,O} + \hat{\pi}_{i,S,DW} \\ DO & otherwise \end{cases}$$

2. Predict the duration of the current spell $j \in \{W, O, DW, DO\}$ in age a.
   (a) Given the vector of socio-demographic X_a and historical variables H_a at age a, the learner π^{DUR}_j is used to predict the instantaneous survival probability $\hat{\pi}_{i,a,j} = \pi^{DUR}_j(X_{i,a}, H_{i,a})$.
   (b) Get a pseudo-random draw ν_i from a uniform distribution.
   (c) Predict the $STATE_{i,a+1} = \begin{cases} j & if \;\; \nu_i \leq \hat{\pi}_{i,jt} \\ k \neq j, \text{Move to step 3} & otherwise \end{cases}$
   (d) If the individual's age reaches the maximum age limit[20], then take the next individual i and return to step 1.
   (e) Update history variables, including age, i.e., $a = a + 1$.

[18] Socio-demographic characteristics are constant explanatory variables. The only exception regards Reduced Working Capacity which exogenously updates according to age.

[19] So, eligibility conditions update given the simulated history variables and eventually reduced working capacity.

[20] The individual's age limit is the minimum between the age of the interview year and 70 years.

(f) Move to step 2a.

3. If step 2c determines the end of the spell in state j, predict the state k in which i transits at age a.

   (a) Determine the set of all the available destinations states
   $$K = \begin{cases} \{W, O, DW, DO\} & if \;\; a < 45 \\ \{W, O, DW, DO, R\} & otherwise \end{cases}$$
   (b) Given $X_{i,a}$, $H_{i,a}$ and the supervised learner $\pi^{TR}_{j,k}$ trained in the third building module, predict the probabilities of all admissible transitions from state j into $k \neq j$, i.e., $\hat{\pi}_{i,a,j,k} = \pi^{TR}_{j,k}(X_{i,a}, H_{i,a}), k \in K$.
   (c) Get a pseudo-random draw ν_i from a uniform distribution.
   (d) Predict the $STATE_{i,a+1} = k$ following the strategy in step 1c .
   To clarify, let for instance $j = W$ and $a >= 45$, then
   $$STATE_{i,a+1} = \begin{cases} O & if \;\; \nu_i \leq \hat{\pi}_{i,a,W,O} \\ DW & if \;\; \hat{\pi}_{i,a,W,O} \leq \nu_i < \hat{\pi}_{i,a,W,O} + \hat{\pi}_{i,a,W,DW} \\ DO & if \;\; \hat{\pi}_{i,a,W,O} + \hat{\pi}_{i,a,W,DW} \leq \nu_i < \hat{\pi}_{i,a,W,O} + \hat{\pi}_{i,a,W,DW} + \hat{\pi}_{i,a,W,DO} \\ R & otherwise \end{cases}$$
   (e) If $k = R$ or the individual's age reaches the maximum age limit, then take the next individual i and return to step 1.
   (f) Move to step 2c.

We perform the algorithm until the labour life history of all individuals is simulated.

5.3 Out-of-sample predictions

The quality of simulations hinges on the XGBoost learners' capability to capture complexities in labour market histories. We assess the predictive performance of the model, by conducting an out-of-sample cross-validation experiment following these steps:

1. Randomly split the sample by gender and country:
   - keep 75% of the sample as the training sample;
   - keep 25% of the sample as the test sample.
2. Re-estimate the building blocks of the TPM using only the training sample.
3. Predict labour market histories only for individuals in the test sample using the estimated learners from step 2.

4. Repeat previous steps 200 times to account for the randomness due to the stochastic split of training/test data and the stochastic nature of the Monte Carlo algorithm.
5. Evaluate the out-of-sample predictive accuracy by comparing the observed and predicted labour market histories in terms of
   - average number and duration of spells;
   - probabilities of exit transitions;
   - age-profiles of binary indicators for the labour market states.

Table 7 displays the predicted and observed average number of spells and their durations. Furthermore, it presents the predicted-to-observed ratio to facilitate performance comprehension. Ratios close to 1 indicate a good estimate of the predicted number and duration of spells. Ratios greater than one signify the model's overestimation of true histories, and vice versa. The fact that the ratios consistently remain close to 1 is reassuring, indicating that the Transition Probability Model accurately simulates the career pathways.

Table 8 provides a more detailed analysis of the model's ability to accurately simulate transitions between different states. The table includes the predicted-to-observed ratios of transition probabilities to better evaluate the simulation's effectiveness. Once again, the majority of the ratios are close to 1, indicating a good fit. However, there are a few exceptions involving rare transitions with small denominators, leading to seemingly large deviations even though they are small in reality. For example, the transition from state O to DW for women has a predicted probability of 0.09, compared to the observed probability of 0.10, resulting in a ratio of 0.89.

Gender		Predicted		Observed		Ratio	
		N spells	Mean duration	N spells	Mean duration	N spells	Mean duration
MEN	W	1,116.88	22.26	1,088.24	22.75	1.03	0.98
	O	291.79	4.16	276.68	4.08	1.05	1.02
	DW	749.05	5.92	734.95	6.21	1.02	0.95
	DO	276.31	8.91	261.54	8.65	1.06	1.03
	R	493.06	9.46	496.44	9.48	0.99	1.00
WOMEN	W	1,345.68	15.24	1,327.31	15.18	1.01	1.00
	O	649.54	7.29	617.95	8.08	1.05	0.90
	DW	777.88	5.95	778.45	5.86	1.00	1.02
	DO	433.04	9.55	413.98	10.13	1.05	0.94
	R	439.40	9.20	426.36	9.32	1.03	0.99

Table 7: Cross validation estimates of the average duration of spells. 200 out-of-sample predictions, Men-Women aged 10-70.

			Destination				
Gender	Origin	Uncen spells	W	O	DW	DO	R
MEN	Pred W	630.51	NaN	0.05	0.62	0.02	0.32
	Obs W	626.38	NaN	0.05	0.62	0.02	0.32
	Ratio	1.01	NaN	0.95	1.01	0.88	1.00
	Pred O	45.78	0.41	NaN	0.07	0.23	0.30
	Obs O	42.41	0.37	NaN	0.09	0.24	0.30
	Ratio	1.08	1.10	NaN	0.76	0.97	0.97
	Pred DW	457.46	0.26	0.02	NaN	0.33	0.40
	Obs DW	463.99	0.23	0.02	NaN	0.32	0.43
	Ratio	0.99	1.14	0.95	NaN	1.00	0.92
	Pred DO	122.79	0.04	0.06	0.16	NaN	0.74
	Obs DO	115.36	0.05	0.04	0.18	NaN	0.73
	Ratio	1.06	0.72	1.55	0.86	NaN	1.02
WOMEN	Pred W	568.66	NaN	0.10	0.63	0.02	0.25
	Obs W	559.51	NaN	0.09	0.64	0.02	0.25
	Ratio	1.02	NaN	1.13	0.97	0.87	1.03
	Pred O	133.01	0.37	NaN	0.09	0.34	0.21
	Obs O	124.94	0.37	NaN	0.10	0.31	0.22
	Ratio	1.06	1.00	NaN	0.89	1.10	0.92
	Pred DW	456.81	0.25	0.02	NaN	0.40	0.33
	Obs DW	459.72	0.24	0.03	NaN	0.39	0.34
	Ratio	0.99	1.06	0.87	NaN	1.01	0.96
	Pred DO	176.56	0.02	0.13	0.21	NaN	0.63
	Obs DO	175.72	0.03	0.09	0.24	NaN	0.64
	Ratio	1.00	0.71	1.37	0.90	NaN	1.00

Table 8: Cross-validation estimates of average transition probabilities. 200 out-of-sample replications, Men-Women aged 45-70. 75% of the sample was used as the training sample and the remaining 25% as the test sample.

Figure 8 offers a comprehensive overview of the out-of-sample simulation for binary indicators concerning the states of the world of the test sample. It depicts the proportion of individuals in each state by age. The dotted blue (red) lines represent the observed proportions for men (women), while the straight blue (red) lines indicate the simulated proportions for men (women). An important observation is that the observed and simulated proportions are nearly overlapping, demonstrating the model's ability to effectively capture the essential characteristics of career paths for both men and women. The only slight underestimation is observed in the proportion of individuals in retirement. However, this discrepancy occurs after the age of 65 when the number of observations becomes limited.

Additionally, figure 8 offers insights into the economically distinct pathways based on gender. It becomes evident that the proportion of men engaged in work is significantly larger compared to women. This disparity is particularly pronounced from the twenties to forties, suggesting that women might exit the labour market to prioritise childcare and household responsibilities. This phenomenon creates a difference in the proportion of women able to access retirement during older ages, mainly due to their fewer years of contributions to the pension system. Moreover, concerning disability states, women tend to have a higher proportion in the Disability Other category, indicating that they are not actively employed. However, regarding Disability Working status, there is no significant gender difference.

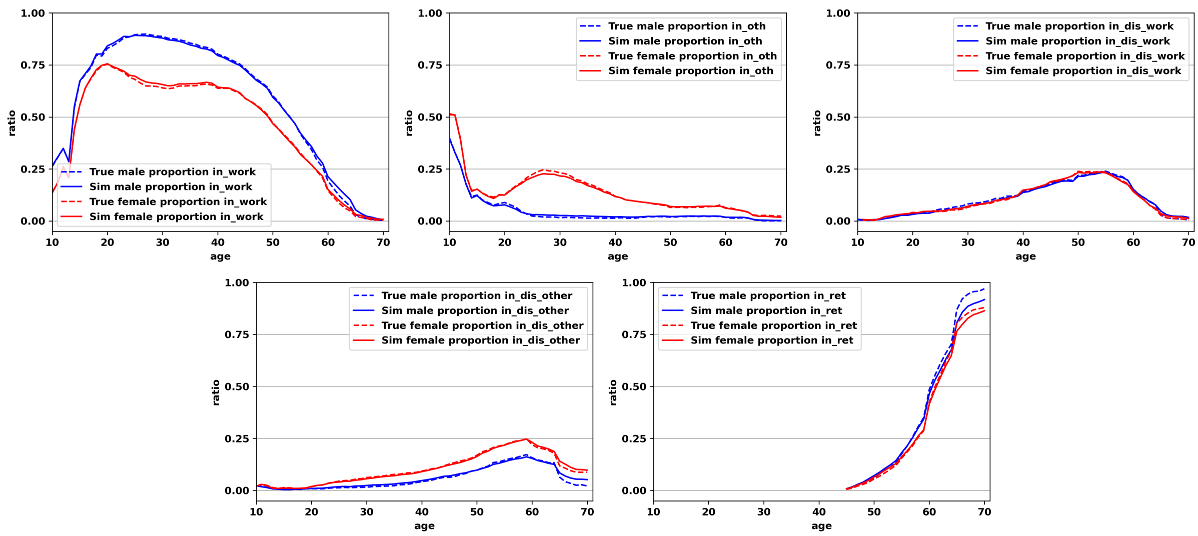

Figure 8: Observed and predicted age profiles of binary indicators for labour market states of men and women. 200 out-of-sample replications, Men-Women aged 10 - 70. 75% of the sample was used as the training sample and the remaining 25% as the test sample.

5.4 Simulation algorithm

In the previous section, we demonstrated that our TPM accurately captures career pathways. Now, we outline our approach to conducting policy reform simulations. Our primary focus is on eligibility conditions for both early pension and disability insurance benefits.

In detail, the following counterfactual scenarios are simulated:

1. Abolition of early pension
2. Abolition disability insurance benefits

Specifically, we perform twice the Monte Carlo algorithm in section 5.2 to evaluate the effects of policy simulations. First, we simulate the *status quo* scenario by using the pension/disability rules which are in place in each country each year. Then, in the second step, the *reform* scenario is simulated by changing the eligibility rules in each country year. For instance, the abolition of the disability route changes the way in which eligibility for disability pensions is set up by not allowing eligibility even if real legislation would allow it. The outcome of interest is the difference between the proportions of individuals in any given labour market state given the reform and the status quo scenarios.

Owing to the stochastic nature of the procedure, the results are proposed as an average of 30 independent replications of the Monte Carlo algorithm. In addition, it is proposed a 95% symmetric confidence band using the standard errors obtained from a stratified nonparametric bootstrap procedure with 200 replications. In detail, the bootstrap follows the following steps:

1. randomly sample with replacement individual labour market histories stratified by gender and country;
2. re-estimate the building blocks on the stratified sample from step 1;
3. run the Monte Carlo algorithm employing the building blocks estimated in step 2 to predict labour market histories under both the reform and status quo;
4. repeat the previous steps 200 times;
5. computes standard errors as standard deviations over the 200 bootstrap replications of the difference between simulated outcomes under the reform and status quo scenario.

In section F of the appendix, we present additional robustness analyses to further validate our findings. First, we estimate the Transition Probability Model (TPM) using logistic regression as an alternative approach. We also consider milder policy reforms and explore alternative definitions of the RWC.

6 Results

We show the age profiles of the differences between proportions of individuals in a given labour market state which result from the difference between the reform (counterfactual) and the status quo (baseline) scenarios, respectively. The bold line represents the average difference between counterfactual and baseline based on 30 independent Monte Carlo simulations. Shaded areas represent the 95% confidence interval from the stratified non-parametric bootstrap procedure. Given our primary focus on the disabled population, we restrict the main analysis to only 5752 individuals who experience an extreme health shock that makes them eligible for DI benefits. This reduction in the sample size significantly shortens the simulation time without affecting the main results, as individuals who are always healthy would never be impacted by reforms in disability insurance programs.

6.1 Early pension eligibility abolition

In the first simulation, we investigate the effect of the early pension scheme on the labour market trajectories of disabled individuals. To achieve this, we implement a reform scenario where we modify the country's rules to eliminate the option of becoming eligible for an early pension. The results are shown in figure 9. As expected, eliminating access to the early pension route decreases the proportion of individuals in retirement within the 60-65 age range, consequently compelling them to spend more time in work or disability states. It is important to recall that disabled people who might have exploited disability insurance benefits to retire are not affected by this policy simulation. On the contrary, the effects in figure 9 show the alternative routes would have been chosen by those who would have exploited the early pension route to access retirement. Since after the reform, this route is no more accessible, these individuals remain in a disability state, which under some conditions, provides eligibility for disability insurance benefits. Although the model does not allow to explicitly determine who accessed the benefits, it is quite reasonable to believe that disability benefits uptake might increase. Hence, potential substitutability between early pension and disability insurance benefits should be considered whenever policy reforms are planned. However, it is noteworthy to point out that the effects are relatively small and not significant.

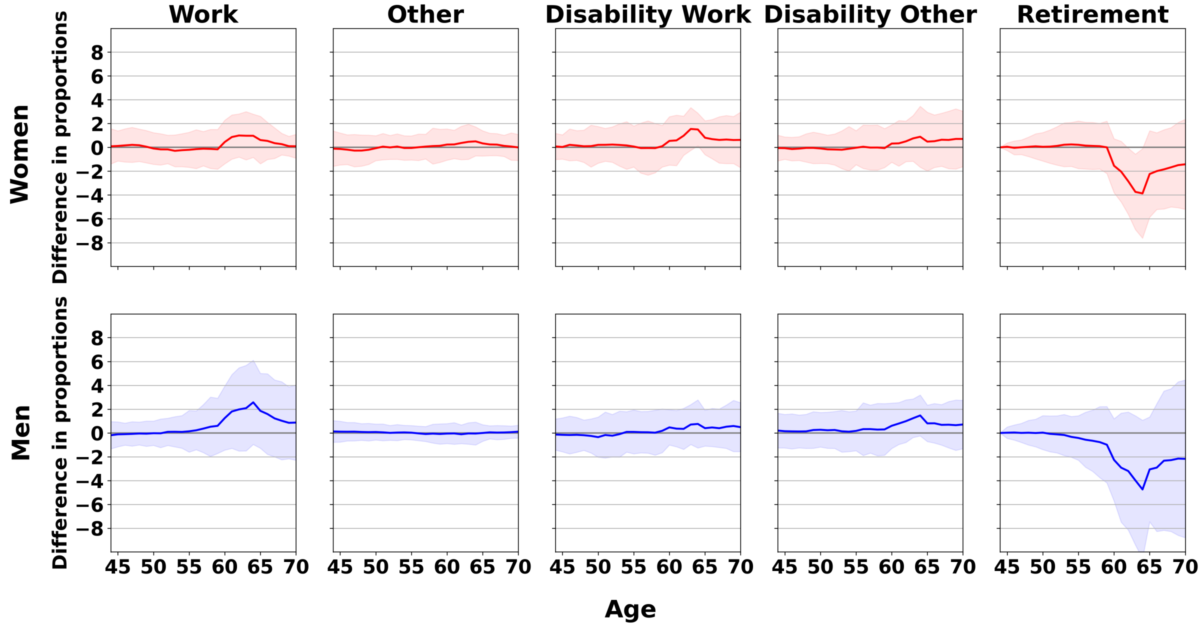

Figure 9: Age profiles of differences in proportions due to a restriction (abolition) in early pension eligibility on the restricted sample (more unhealthy). Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

6.2 Disability eligibility abolition

In the second policy reform, we modify the country's rules by abolishing the option to become eligible for a disability insurance benefit[21]. A priori, the abolition of disability benefits has two opposite driven forces on the extensive margin of labour market participation. On the one hand, disabled individuals who are not working and lose their benefits have an incentive to move back into the labour market to recover the income lost. On the other hand, those already in the labour market may search for alternative compensating schemes for which working activities might not be expected. Policy simulations in figure 10 show that those moving outside the labour market outweigh those who move inside. In fact, the reduction in the share of disabled workers is vastly compensated by alternative compensating schemes, i.e., unemployment benefits (Disability Other) and early pension (Retirement). An important aspect is that within these schemes is not usual and not expected to carry out work activities. Again, the results support the substitutability hypothesis between disability and alternative compensating schemes. As a result, policy restrictions on disability insurance benefits eligibility would hardly improve the

[21] It is important to stress that the policy is rather extreme only for illustrative purposes. Further evidence about less extreme reforms is presented in the robustness section F in the appendix.

sustainability of social provision schemes if the alternative routes are easily accessible. In addition, as it can be observed in figure 11, the policy effect on net employment[22] is negative.

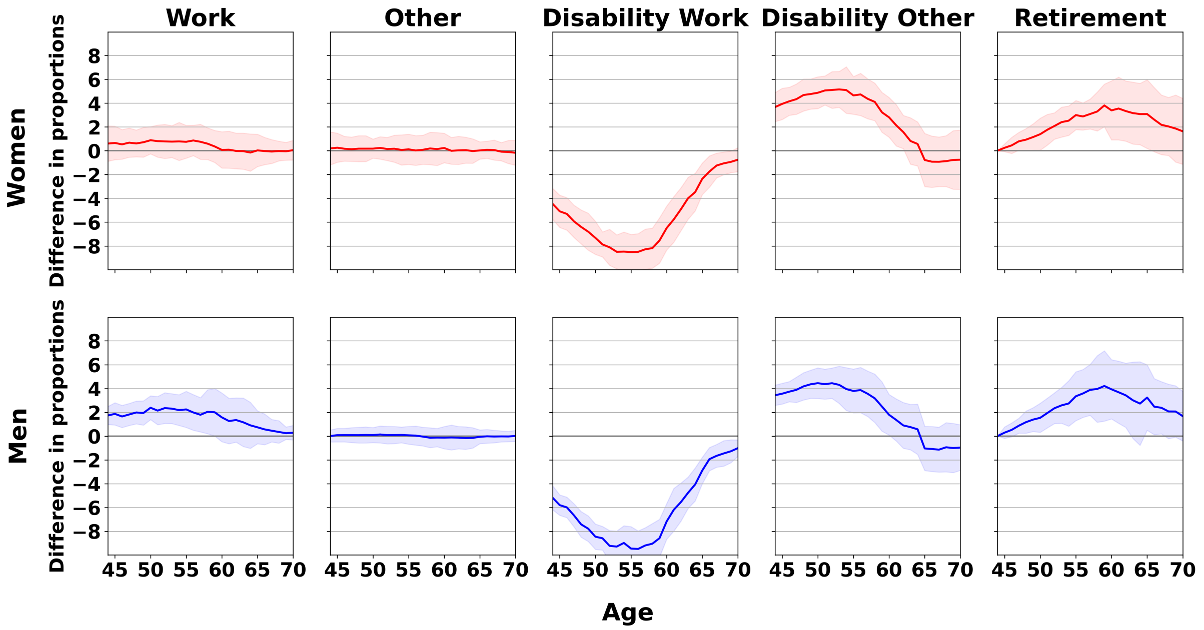

Figure 10: Age profiles of differences in proportions due to a restriction (abolition) in disability insurance benefit eligibility on the restricted (more unhealthy) sample. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

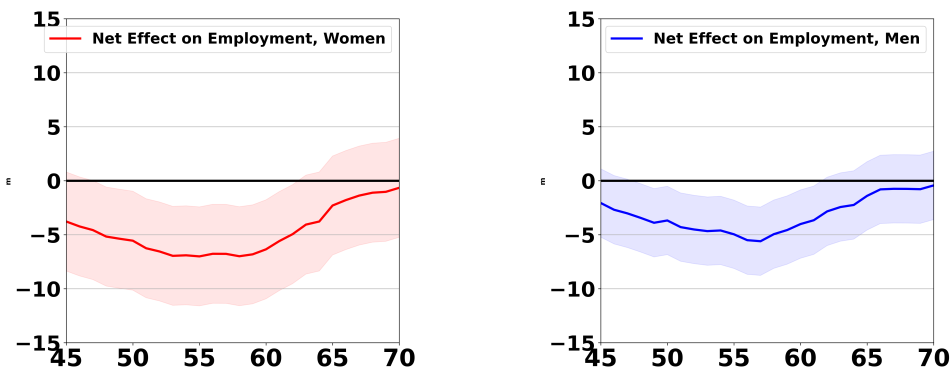

Figure 11: Age profile of net effect on employment of disability insurance benefit abolition. In red women, in blue men. Effect computed as $W + DW$ using the average of the 30 simulations.

[22]Net employment is computed as the sum of Work and Disability Work.

6.3 Policy effects discussion

Two main points deserve further critical discussion. First of all, it is noteworthy to point out the differences in magnitude between the two policy reforms. On the one hand, tightening the early pension pathway does not markedly influence disability pathways simply because healthy individuals cannot easily (and perhaps do not want to) fake a disability condition. Hence, their natural response to the policy is generally to continue labour market activities. On the other hand, the impact of the abolition of disability insurance benefits is much larger and as a consequence, significant spillovers towards alternative compensating schemes are observed. Secondly, at first glance, the results appear to be unexpectedly contrary to what the findings observed in a vast majority of the literature. It is worth mentioning that by exploiting the above simulation framework which allows for a high degree of control over the environment, it comes out that restricting access to disability insurance benefits does not necessarily guarantee a more sustainable social protection system nor an increase in labour market participation. This is important because if the positive effect on labour market participation is not directly caused by a pure restriction of the disability insurance, but instead the restriction affects other characteristics which in turn are the main causal drivers, understanding the underlying mechanisms may allow to better target the intervention. Hence, trying to understand which are the relevant aspects which guide the behaviour of individuals may help to better allocate the limited resources and avoid reforms which hit indiscriminately all.

7 Behavioural Response: Update in Working Capacity

In this section, we extend the analysis by introducing possible behavioural responses after policy reforms. It is worth mentioning that the results in the previous section rely on two implicit assumptions. The first one regards the fact that the alternative compensating schemes are fully available, whereas the second one refers to the fact that working capacity does not update to the new contingency. Hence, since after a policy reform it is not granted that the alternative compensating schemes would have been available, some individuals may be forced to reveal their maximal level of working capacity to recover the lost income. Evidence about under-exploitation of working capacity is well-documented[23]. However, determining how working capacity adapts to policy reforms is not a trivial exercise. To set it up in a credible way, we refer to the literature on screening errors during application processes for disability insurance benefits. In brief, screening errors can be of two types: under-assessment or over-assessment of Reduced Working Capacity. The literature (Nagi (1969); Low and Pistaferri (2015, 2019); Benitez-Silva et al. (2004)) extensively examines the two extreme cases[24], namely, the rejection into the program of individuals who are actually disabled (type I error) and acceptance of individuals who are not disabled (type II errors). Literature shows huge type I errors (from 37 to 54%) and relevant type II errors (from 18 to 28%). These errors are highly dependent on the stringency of the assessment procedures. In our framework, the underlying mechanism of the screening process is well known since the Reduced Working Capacity is determined according to survey answers. On the one hand, this means that reporting a disability condition has no cost and no medical proof is required. As a consequence, under-assessment of RWC in this setting is quite unlikely. On the other hand, the extremely lenient procedure might raise concerns about the over-assessment of RWC. This is important because it is quite natural to believe that the most reactive individuals in updating the working capacities following a policy reform are those whose Reduced Working Capacities are over-assessed during the screening process. Hence, we simulate the disability insurance benefit abolition by proposing different shares[25] of random individuals who have an over-assessment in RWC and update their working capacity after policy reforms[26] The net effects on employment are presented in figure 12.

[23]See for instance the literature investigating the responses of current disability insurance beneficiaries to incentives to return to work Von Wachter et al. (2011); Kostol and Mogstad (2014); Borghans et al. (2014). Their main conclusion is that some DI recipients may have significant capacity to work.

[24]The reason why most evidence is on the "extreme" cases of assessment is because the US disability insurance system relies on an acceptance/rejection basis, in contrast with the majority of European countries in which is measured the degree of disability.

[25]Determining the exact share of individuals who might update their working capacity is not trivial. For this reason, the share of type II errors from literature is taken as a reference point. A general consensus in the literature points out that around 20% of type II errors is a reasonable amount. However, it refers to the restricted self-selected sample of those who apply for the benefits. By considering a broader sample, this share can be reasonably larger. Simultaneously, it should be considered that not all over-assessed individuals might update their working capacities after policy reforms. These are the reasons why different shares of "reacting" individuals are proposed and the highest share in the type II errors consensus in the literature.

[26]In this setting the updating in RWC is rather extreme. The share of random individuals become fully capable to work in the reform scenario. Different specifications create intermediate results.

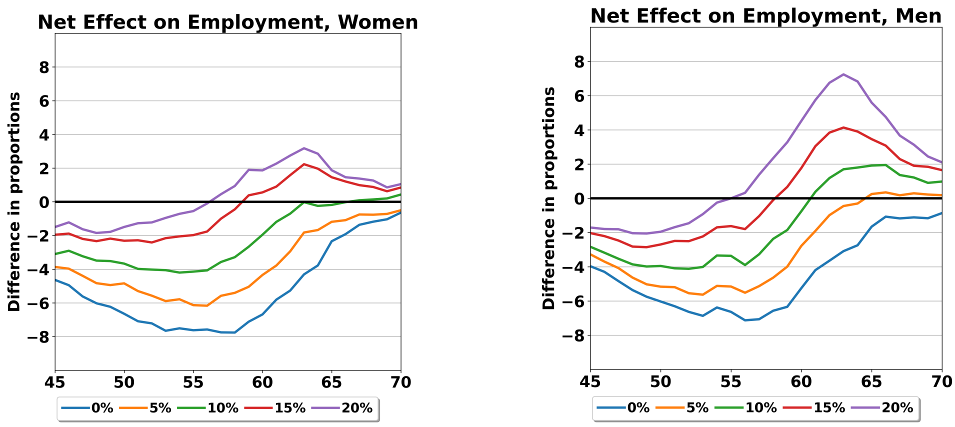

Figure 12: Age profile of average net effect on employment after disability benefit abolition given different shares of over-assessed RWC individuals. On the left women, on the right men's effects.

Figure 12 clearly shows that the abolition of disability insurance benefits affects net employment according to how many individuals have over-assessed RWC and can (and want to) contingently increase their working capacity. In detail, the net effect on employment becomes more positive for men when working capacities are updated. Additionally, the effects are notably more pronounced at older ages, which can be attributed to the increased likelihood of limitations occurring as individuals age. This stands in contrast to limitations that may arise at younger ages, which could be more severe and less likely to be recoverable. Furthermore, the elimination of the disability route positively impacts employment only when approximately 15% or more of individuals can enhance their working capacity.

This is crucial since the same policy intervention might have different outcomes according to the quality of the assessment of the RWC. Furthermore, this result sheds light on the heterogeneous effects of the DI policy restrictions. Individuals unable to update their working capacity would prefer alternative compensating schemes and in case these are not available, are doomed to drastically reduce their income. On the contrary, individuals able to update their working capacity, in front of a lack of valid alternative compensatory schemes, would be able (at least partially) to compensate for the lost income. Since the majority of DI beneficiaries likely have little or no updating ability, policy reforms might have unexpectedly harsh effects on the most vulnerable individuals.

7.1 Critical discussion and comments

This work explicitly put emphasis on the importance of being aware of the composition of DI beneficiaries. Specifically, if restrictions in generosity and eligibility of DI increase labour market participation through the exploitation of unused working capacities, knowing the composition of the target policy may help to reduce undesirable side effects. As a matter of fact, policy reform restrictions may harm also those who have no opportunity to expand labour market participation. Therefore, future research should delve deeper into how institutional settings influence the composition of DI recipients and explore strategies to encourage the full utilisation of individuals' working capacities from the outset of the disability event. For instance, Hullegie and Koning (2018) point out that the most effective DI reforms in the Netherlands were the ones in which the burden of sickness benefits and sickness monitoring obligations were passed to employers (early 2000s). The logic behind this intervention relies on the fact that employers can activate early reintegration activities and workplace accommodations for sick and disabled individuals. Then, another institutional aspect which receives little attention is the reassessment process. Low and Pistaferri (2020) remarks that in practice in the US and UK benefit award is substantially permanent since reassessment is infrequent and conducted in an inaccurate way[27]. More frequent and careful reassessment, of those individuals for which an improvement is feasible, may help to reveal individuals with residual working capacity without harming the entire population of DI beneficiaries. Then, another strand of literature (Nagi (1969); Low and Pistaferri (2015, 2019); Benitez-Silva et al. (2004)), shows that it is possible, at least to some extent, to induce behavioural responses aimed at increasing labour participation by providing monetary incentives to DI beneficiaries. Next, one general aspect which is crucial is the timing of the intervention. For instance, Hemmings and Prinz (2020) claims that the Swiss DI reform (2003-2016) that focused on early recognition of problems and early intervention was fundamental to reducing disability benefit requests. Early intervention is important because long-term inactivity encourages the depreciation of human capital and simultaneously the mindset of individuals accommodates inactivity as the new status quo. For this purpose, early interventions involving vocational training might be helpful to secure job retention or to support job search. As a matter of fact, de Jong (2011) underlines the positive association between employment rates and vocational rehabilitation spending. However, we should not ne-

[27] The same is generally true also for European countries by observing the low share of outflows beneficiaries from DI schemes (OECD).

glect the fact that also the demand side plays a critical role. For this purpose, a valuable mechanism to induce employers to retain or hire people with disability are mandatory employment quotas. In addition, to ensure equal treatment of people with disability, anti-discrimination legislation should be enacted. Both tools need to be accurately designed to be adequately enforced. All previous suggestions are in line with OECD (2010) report. Nevertheless, enforcing a more efficient assessment procedure which tries to encourage the exploitation of any residual working capacity is not free of drawbacks. Each of the above-mentioned proposals comes at a cost, both from a budgetary and organisational point of view. In addition, another hidden drawback could arise (e.g., increasing employer duties may discourage them to employ sick or disabled people). Hence, future research should investigate how to optimise the implementation of policy tools aimed at encouraging labour market participation by trying to avoid undesired policy reforms which affect indiscriminately all the disabled population.

8 Conclusion

In this work, we examine the impact of disability insurance on the labour market trajectories of individuals across eleven European countries. The most innovative part of the work relates to the fact that the analysis is performed on the whole life course of each individual. This requires detailed labour market and health information for each age of life of each individual and complete disability eligibility criteria for each country in each year of the analysis. To comply with this requirement, different data sources have been collected. First, for each country, we collected detailed administrative requirements regarding eligibility for disability insurance and pension schemes. Afterwards, recognising the limitations of administrative data in providing a comprehensive view of the entire disabled population, we collect microdata from SHARELIFE, a European representative survey of older individuals containing retrospective information which captures the most relevant labour and health aspects. The microdata allows us to measure the degree of disability for each individual for each age of life. This granular data is used to estimate a Transition Probability Model employing machine learning techniques. Afterwards, following Brugiavini et al. (2020), we set up a Monte Carlo algorithm to uncover the most likely labour market histories. This simulation framework provides a high degree of control over each aspect of the environment and allows the investigation of the underlying drivers of individual behaviour. Policy reform simulations show that early retirement and disability insurance schemes are potential substitutes. However, if on the one hand, restricting early retirement routes has no or little effect on disability routes, on the other hand, restricting disability insurance accessibility has a significant impact on alternative compensating schemes. In addition, it is noteworthy to report that purely restricting disability insurance accessibility does not increase labour market participation nor incentives sustainability of social security provision. This result is relevant since highlights that policy restrictions may act as triggers for underlying mechanisms which modify the behaviour of individuals. Simulations show that one possible channel through which individuals change behaviour is the update in the working capacity for those individuals who experienced an over-assessed degree of disability. In particular, our simulations point out that the policy effect on employment depends on the share of individuals who are able (and want) to increase their working capacity.

Finally, two main policy implications are raised. Since the effects of disability insurance reforms are highly dependent on the composition of the population, the first implication is that it is essential to be aware of the efficiency of the screening procedures. The second policy implication is that, if the update in working capacity is one of the main drivers which increase labour supply contingently to a DI policy restriction, being able to directly focus on interventions aimed at exploiting the residual working capacity may avoid proposing policy reforms which affect indiscriminately all the population. Reforms in the Netherlands which increased employer involvement or in Switzerland which focused on early intervention are perfect examples of how policy reforms could target simultaneously sustainability of social security provision and maintain the living standards of disabled individuals.

References

S. Albanesi and D. F. Vamossy. Predicting consumer default: A deep learning approach. Technical report, National Bureau of Economic Research, 2019.

P. D. Allison. Discrete-time methods for the analysis of event histories. *Sociological methodology*, 13:61–98, 1982.

L. Antonova, L. Aranda, G. Pasini, and E. Trevisan. Migration, family history and pension: the second release of the share job episodes panel. *MEA, Max Planck Institute for Social Law and Social Policy, SHARE Working Paper*, (18-2014), 2014.

F. J. Bargagli-Stoffi, J. Niederreiter, and M. Riccaboni. Supervised learning for the prediction of firm dynamics. In *Data Science for Economics and Finance*, pages 19–41. Springer, Cham, 2021.

L. Behaghel, D. Blanchet, T. Debrand, and M. Roger. *8. Disability and Social Security Reforms: The French Case.* University of Chicago Press, 2012.

M. Belloni and R. Alessie. Retirement choices in italy: What an option value model tells us. *Oxford Bulletin of Economics and Statistics*, 75(4):499–527, 2013.

M. Ben Jelloul, A. Bozio, E. Perdrix, A. Rain, and L. Toulemon. Dynamic of the disablement process in aging populations. 538(1):13–31, 2023.

H. Benitez-Silva, M. Buchinsky, and J. P. Rust. How large are the classification errors in the social security disability award process?, 2004.

P. Bingley, N. D. Gupta, and P. J. Pedersen. *5. Disability Programs, Health, and Retirement in Denmark since 1960.* University of Chicago Press, 2012.

R. Blundell, E. French, and G. Tetlow. Retirement incentives and labor supply. In *Handbook of the economics of population aging*, volume 1, pages 457–566. Elsevier, 2016.

T. Boeri and A. Brugiavini. Pension reforms and women retirement plans. *Journal of Population Ageing*, 1(1):7–30, 2008.

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

L. Borghans, A. C. Gielen, and E. F. Luttmer. Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American economic Journal: economic policy*, 6(4):34–70, 2014.

A. Börsch-Supan and H. Jürges. *7. Disability, Pension Reform, and Early Retirement in Germany*. University of Chicago Press, 2012.

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.

A. H. Börsch-Supan. Work disability, health, and incentive effects. 2007.

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

L. Breiman, J. Friedman, C. Stone, and R. Olshen. *Classification and Regression Trees*. Taylor & Francis, 1984. ISBN 9780412048418. URL https://books.google.it/books?id=JwQx-WOmSyQC.

P. Browne, E. Carr, M. Fleischmann, B. Xue, and S. A. Stansfeld. The relationship between workplace psychosocial environment and retirement intentions and actual retirement: a systematic review. *European journal of ageing*, 16(1):73–82, 2019.

A. Brugiavini and F. Peracchi. 4. health status, welfare programs participation, and labor force activity in italy. In *Social security programs and retirement around the world*, pages 175–216. University of Chicago Press, 2012.

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: Munich Center for the Economics of Aging (MEA)*, 36, 2019.

A. Brugiavini, G. De Luca, T. E. MaCurdy, and G. Weber. The effects of social policies on the working careers of europeans. Technical report, IFS Working Papers, 2020.

J. S. Butler, R. V. Burkhauser, J. M. Mitchell, and T. P. Pincus. Measurement error in self-reported health variables. *The Review of Economics and Statistics*, pages 644–650, 1987.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

C. C. Coile and K. Milligan. *Social security programs and retirement around the world: disability insurance programs and retirement*. University of Chicago Press, 2016.

C. Dal Pozzo, H. Haines, Y. Laroche, F. Fratello, and C. Scorretti. Assessing disability in europe–similarities and differences. *Council of Europe, Strasbourg*, 2002.

P. R. de Jong. Sickness, disability and work: Breaking the barriers—a synthesis of findings across oecd countries–oecd, 2011.

E. Deuchert and B. Eugster. Income and substitution effects of a disability insurance reform. *Journal of Public Economics*, 170:1–14, 2019.

R. Euwals, A. Van Vuren, and D. J. van Vuuren. The decline of early retirement pathways in the netherlands: An empirical analysis for the health care sector. 2011.

E. French and J. B. Jones. Health, health insurance, and retirement: a survey. *Annual Review of Economics*, 9:383–409, 2017.

P. García-Gómez, S. Jiménez-Martín, and J. V. Castelló. 3. health, disability, and pathways into retirement in spain. In *Social security programs and retirement around the world*, pages 127–174. University of Chicago Press, 2012.

R. M. Gritz and T. E. MaCurdy. *Participation in low-wage labor markets by young men*. US Department of Labor, Bureau of Labor Statistics, 1992.

J. Gruber. Disability insurance benefits and labor supply. *Journal of political economy*, 108(6):1162–1183, 2000.

S. Gu, B. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

A. L. Gustman and T. L. Steinmeier. *The Role of Health in Retirement*. Emerald Publishing Limited, 2018.

B. Hanel. The effect of disability pension incentives on early retirement decisions. *Labour Economics*, 19(4):595–607, 2012.

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. Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review*, 81(2):75–79, 1991.

P. Hemmings and C. Prinz. Sickness and disability systems: comparing outcomes and policies in norway with those in sweden, the netherlands and switzerland. 2020.

B. E. Honoré. Identification results for duration models with multiple spells. *The Review of Economic Studies*, 60(1):241–246, 1993.

P. Hullegie and P. Koning. How disability insurance reforms change the consequences of health shocks on income and employment. *Journal of Health Economics*, 62:134–146, 2018.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

S. P. Jenkins et al. Easy estimation methods for discrete-time duration models. *Oxford bulletin of economics and statistics*, 57(1):129–138, 1995.

L. Jönsson, M. Palme, and I. Svensson. *2. Disability Insurance, Population Health, and Employment in Sweden*. University of Chicago Press, 2012.

A. Jousten and M. Lefebvre. Retirement incentives in belgium: estimations and simulations using share data. *De Economist*, 161(3):253–276, 2013.

A. Kapteyn, J. P. Smith, and A. Van Soest. Vignettes and self-reports of work disability in the united states and the netherlands. *American Economic Review*, 97(1):461–473, 2007.

M. Kerkhofs, M. Lindeboom, and J. Theeuwes. Retirement, financial incentives and health. *Labour Economics*, 6(2):203–227, 1999.

P. Koning and M. Lindeboom. The rise and fall of disability insurance enrollment in the netherlands. *Journal of Economic Perspectives*, 29(2):151–72, 2015.

A. R. Kostol and M. Mogstad. How financial incentives induce disability insurance recipients to return to work. *American Economic Review*, 104(2):624–55, 2014.

T. Kyyrä. Early retirement policy in the presence of competing exit pathways: Evidence from pension reforms in finland. *Economica*, 82(325):46–78, 2015.

T. Lancaster. *The econometric analysis of transition data*. Number 17. Cambridge university press, 1990.

H. Low and L. Pistaferri. Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, 105(10):2986–3029, 2015.

H. Low and L. Pistaferri. Disability insurance: Error rates and gender differences. Technical report, National Bureau of Economic Research, 2019.

H. Low and L. Pistaferri. Disability insurance: Theoretical trade-offs and empirical evidence. *Fiscal Studies*, 41(1):129–164, 2020.

S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Merriam-Webster. Merriam-webster definition of disability 2a. *https://www.merriam-webster.com/dictionary/disability*, 2022.

S. Z. Nagi. Disability and rehabilitation: Legal, clinical, and self-concepts and measurement. 1969.

F. Nogueira. Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–. URL https://github.com/fmfn/BayesianOptimization.

OECD. *Sickness, Disability and Work: Breaking the Barriers*. 2010. doi: https://doi.org/https://doi.org/10.1787/9789264088856-en. URL https://www.oecd-ilibrary.org/content/publication/9789264088856-en.

OECD. Social benefit recipients, 2021. URL https://www.oecd.org/els/soc/recipients-socr-hf.htm.

p. OECD, year=2019. *Pensions at a Glance 2019: OECD and G20 Indicators*.

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.

M. Schmälzle, M. Wetzel, and O. Huxhold. Pathways to retirement: Are they related to patterns of short-and long-term subjective well-being? *Social science research*, 77:214–229, 2019.

S. Staubli. The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics*, 95(9-10):1223–1235, 2011.

G. J. Van den Berg. Duration models: specification, identification and multiple durations. In *Handbook of econometrics*, volume 5, pages 3381–3460. Elsevier, 2001.

J.-M. Van Sonsbeek and R. Alblas. Disability benefit microsimulation models in the netherlands. *Economic Modelling*, 29(3):700–715, 2012.

T. Von Wachter, J. Song, and J. Manchester. Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *American economic review*, 101(7):3308–29, 2011.

E. H. I. G. WHO. Number of people receiving social/disability benefits, 2021. URL https://gateway.euro.who.int/en/indicators/hfa_415-2721-number-of-people-receiving-socialdisability-benefits/.

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

Appendix

A Relevance of the topic

Preserving the sustainability of the public pension provision package in an increasingly elderly population by providing adequate protection to individuals who most need it, requires an accurate fine-tuning of the design of the social provision framework.

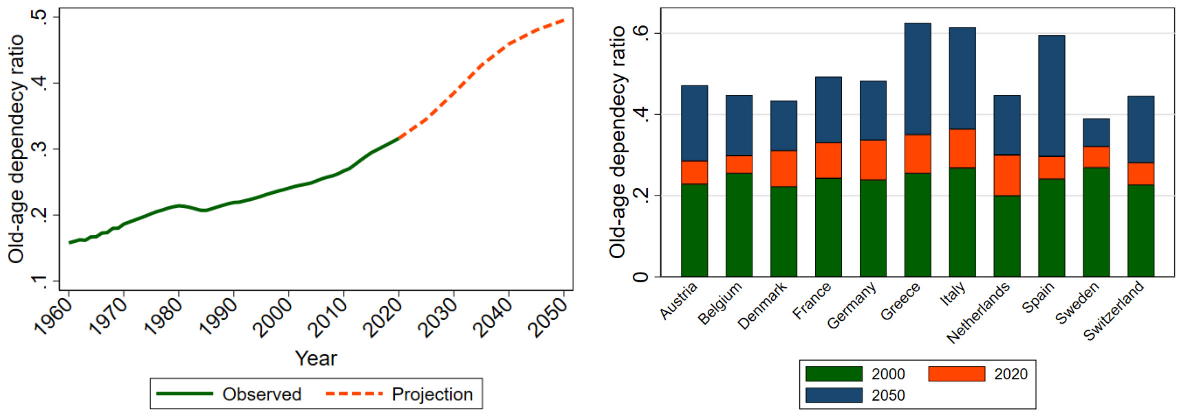

Figure 13: On the LEFT, the trajectory of old-age dependency ratio (mean of 11 European countries). On the RIGHT, the old-age dependency ratio by country in 2000, 2020 and 2050 projections (Eurostat data, 2019, 2021). The dependency ratio is computed as the ratio between the 65+ and 15-64 population.

As it can be noticed in figure 13, the old-age dependency ratio (the ratio between the number of individuals older than 64 and the number of individuals aged 15-64) shows an unequivocal increasing pattern (EUROSTAT data). In 2000, the EU's old-age dependency ratio was around 25%, whereas in 2020 the ratio jumped to almost 32%. By 2050, projections show that the EU's old-age dependency ratio will skyrocket to roughly 50%. To sum up, if in early 2000 for each old individual (65+) there were four individuals of working ages (15-64), in 2020 for each old individual only three potentially working individuals were available. In 2050 is expected that each old individual will depend only on two potentially workable individuals. In addition, as it is clear from figure 13 on the right, country heterogeneity shows that "Mediterranean" countries are those who are more at risk. Of course, this dependency ratio measure does not take into account whether a person is economically active at any given age. For instance, individuals who have long-lasting schooling carries, disabled who are forced to withdraw from the labour market or early retired individuals are not actively supporting the social provision framework. Similarly, individuals who postpone their retirement beyond age 64 are not to be considered dependants. Recent reforms, which are mostly targeting the ”relatively” younger cohorts, aim to postpone exit from the labour market by increasing statutory ages for old-age retirement and reducing the general availability of early exit routes[28]. Hence, the recent legislative restrictions on ”traditional” retirement routes, raise questions about the future take-up of alternative compensating schemes, e.g., disability or unemployment benefits.

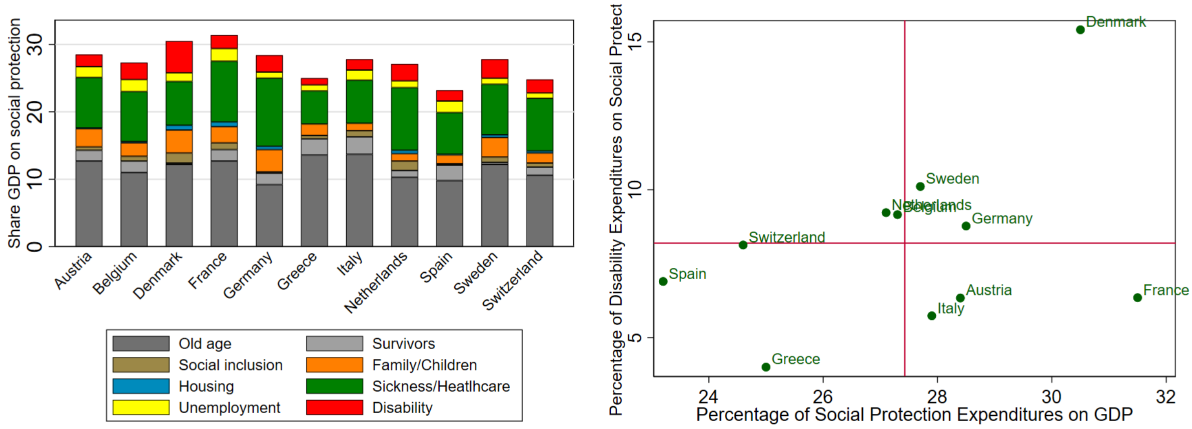

Figure 14: On the LEFT, Social Protection Expenditures as a share of GDP by function in each country in 2018 (EUROSTAT, ESSPROS). On the RIGHT, Disability Expenditures on Social Protection Expenditures and Social Protection Expenditures on GDP, 2018 EUROSTAT.

European countries are characterised by welfare states which provide protection against several risks which each individual may face during their lifetime. One way to understand the magnitude and the most relevant spheres of intervention in each country is to analyse the shares of GDP spent on different social protection functions (figure 14 shows 2018[29] Eurostat data). The overall share of expenditures on GDP is quite comparable, and although notable differences can be observed, on average almost 28% of GDP is absorbed by Social Protection Expenditures. By analysing the composition of the social expenditures, country heterogeneity can be observed. It is worth noticing that more than half of social protection expenditures in Greece and Italy is absorbed by old age and survivor pensions, leaving very few resources for alternative protection functions. Moreover, also the share of expenditures for disability varies across countries (see figure 14 on the right). One pos-

[28]Hence, it is expected that also the rough definition of dependency ratio will change by considering old individuals above the 65 years of age threshold.

[29]2018, is the last year available which provides reliable information for all countries. It is noteworthy to point out that the amount and composition of expenditures varied during the last decades. Hence, this figure documents only the recent position of countries.

sible interpretation of this heterogeneity can be found in the fact that similar needs and risks have been satisfied and protected by providing different institutional frameworks. For instance, in some countries, individuals hit by an extreme health shock are protected by Disability Insurance, in other countries, are favoured early retirement routes, or again, Unemployment Insurance may be preferred. This is important since policy reforms may have unintended spillover effects on alternative compensating schemes or individuals' behaviour. In this regard, policy simulations will be performed to test the substitutability hypothesis between early pension routes and Disability Insurance benefits.

Another reason why Disability Insurance is important is that disability-related expenditures play a not negligible role in public pension provision across Europe. In figure 14, the 2018 expenditure decomposition is presented. Figure 15, OECD (2021) data[30] displays the historical pattern of public spending as a share of GDP on disability and unemployment by country. Although cultural and historical differences shaped institutional heterogeneity in social security provision, two patterns are quite common. The former one is that disability-related spending tends to be systematically larger than on unemployment insurance[31]. The latter one is that the extent of expenditures on disability is a not negligible share of GDP. As a matter of fact, most countries rarely present shares of GDP spent on disability lower than 2%, with peaks reaching 6% for "Nordic" countries.

[30]OECD data are used since a more long-term period can be taken into account.

[31]For this reason, plus the fact that unemployment covers also a vast portion of healthy individuals in search of a job, Disability Insurance is considered the main alternative route for early retirement.

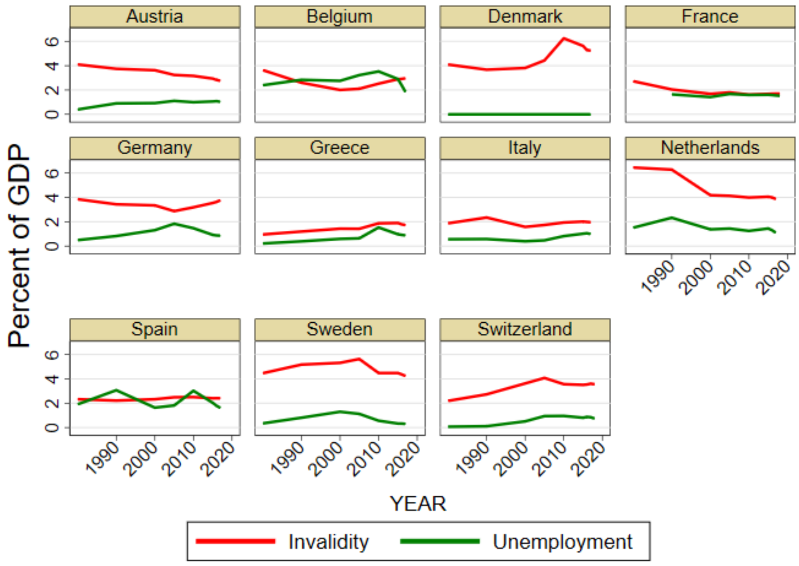

Figure 15: OECD data on Invalidity related (OECD (2021), Public spending on incapacity (indicator). doi: 10.1787/f35b71ed-en) and unemployment spending as a share of GDP in European countries. OECD data is used since a longer-term horizon is available. Data for unemployment in Denmark is not available.

B Sample selection

The starting Job Episodes Panel database originally collects information about 29 countries, 91,743 unique individuals and 6,169,132 observations. However, since Eastern countries had very different institutional frameworks during the Communist ages, the focus is posed on the following 11 Western countries [32]: Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland and Belgium. By keeping the above-mentioned countries, the sample is represented by 48,419 unique individuals and 3,258,255 observations[33]. In the next step, in accordance with Brugiavini et al. (2020), sample restrictions are performed by dropping sequentially the following individuals:

- 2,757 individuals with less than 50 and more than 79 years in 2006[34] (kept those born between 1927 and 1956) for those interviewed in wave 3
- 3,210 individuals with less than 50 and more than 79 years in 2015[35] (kept those born between 1936 and 1965) for those interviewed in wave 7;

[32] Ireland is removed due to a small sample and concerns about sample design.)

[33] At this stage a small typo correction is done for those cases of obvious errors, e.g., those individuals always reported in retirement and so on.

[34] 2006 is the reference year regarding wave 2.

[35] For 107 individuals who were interviewed for the first time in wave 7, the reference year is 2017.

- 1,462 individuals with back-transition into schooling
- 349 individuals who exit from school after 30 years old
- 584 individuals who enter into retirement before age 45 [36]
- 4 individuals who exit from retirement

Individuals who are older than 79 years are dropped for two main reasons. The former one involves considerations about likely coverage errors due to cross-country differences in sample procedures. The latter deals with differences in health/mortality between the disabled and not individuals. In fact, individuals who survive longer are likely to represent the most healthy part of the population and so keeping them would generate an under-representation of disabled individuals. Moreover, the above restrictions implicitly outline restrictions in the proposed model. Individuals are not allowed to remain in full-time education after 30 years old and are not allowed to back-transit into full education once they finish the first cycle of full-time education. In addition, those retiring very early are removed because represent an extremely rare sub-sample or errors in self-reports. Finally, retirement is assumed an absorbing state so once an individual transits into retirement cannot move out.

After this selection, the data set contains 40,053 individuals. Next, in order to clean the data from individuals who never changed their situation (e.g., homemakers who remain always out of labour and eventually perceive a survivor pension) 3,283 individuals are dropped.

Since the research question is focused on old-age behaviour, it is assumed that all individuals attended school at least until age 9 (1,019 individuals affected). To sum up, the cleaned dataset contains 36,770 unique individuals and 2,412,894 observations.

In addition, since only 127 individuals (out of 20,290 individuals in the sample who retired) report a retirement age above 70 and ages before age 10 are of no interest, the time span of interest is 10-70. In detail, it is imposed a right-censored age limit equal to the minimum between the years of age at the interview date (2008 for wave 3 and 2017 for wave 7) and the upper bound of 70 years old.

The final working sample contains 36,770 individuals (who worked for at least one year) and 2,022,440 observations.

[36] They represent a very special category of individuals who are not linked to the eligibility rules (both retirement and DI) and may be difficult to capture in simulations (strongly imbalanced).

Sample	Unique individuals	Observations	Average observations per id	% individuals worked at least one year
JEP	91,743	6,169,132	67.24	91.9
JEP, West	48,419	3,258,255	67.29	90.9
JEP, West + restr	40,168	2,642,678	65.79	93
JEP, West + restr + worked	36,770	2,412,894	65.62	100
JEP, West + restr + worked + cens	36,770	2,022,440	55.00	100

Table 9: Unique individuals and observations for each proposed sample. "JEP" contains the original SHARE data. "JEP, West" contains the original data for 11 above-mentioned western countries. "JEP, West + restr" contains data about 11 western countries by restricting the sample following the above rules. "JEP, West + restr + worked" contains the 11 countries' restricted sample with individuals which worked at least 6 months. Finally, "JEP, West + restr + worked + cens" apply the censoring to keep only the age span of interest.

Preliminary descriptive evidence

It is useful to start by looking at the general and preliminary evidence about the disability state within the data set.

Sample	Unique individuals	Individuals disability (Situation == 5)	% individuals disability (Situation == 5)
JEP, worked	84,329	1,210	1.43
JEP, West + worked	44,023	677	1.53
JEP, West + worked + restr + cens	36,770	576	1.57

Table 10: **Unique individuals** and percentages of unique individuals which reported at least one spell in disability (Situation == 5. "Sick or disabled") for each of the above-mentioned samples conditioning on keeping only individuals which worked at least for one year.

Sample	Observations	Observations in disability (Situation == 5)	% of observations in disability (Situation == 5)
JEP, worked	5,642,081	8,473	0.15
JEP, West + worked	2,947,581	4,643	0.16
JEP, West + worked + restr + cens	2,022,440	4,099	0.20

Table 11: **Observations** and percentages of observations reported in disability (Situation == 5. "Sick or disabled") for each of the above-mentioned samples conditioning on keeping individuals who worked at least for one year.

576 individuals out of 36,770 (1.57%) reported at least one year of disability. Overall, 4,099 observations out of 2,022,440 (0.20%) are reported in disability. Since these self-reports seem to be particularly low, a deeper analysis is performed.

C Country institutional details

Austria

The Austrian pension system relies mainly on the public pension system (first pillar) which is by far the primary source of income for retirees and it is based on a pay-as-you-go (PAYG) scheme. The statutory retirement age is 65 for men and 60 for women[37]. The public pension scheme allows for early retirement, disability insurance benefits, and survivors' pensions. Nowadays, (2021 data) the minimum age to access the early retirement corridor is 57 years for women with at least 42 years df contributions and 62 for men with no less than 40 years od contributions.

In case of earning reduction due to health impairment, three alternative compensating schemes are offered: disability insurance, sickness insurance, and unemployment insurance. A detailed institutional setting of Austria can be found in Staubli (2011). Since the early 1980s, roughly 8% of working-age individuals perceive a disability insurance benefit. Hence, disability benefits can be arguably hypothesised as an early pathway to retirement. In brief, the Austrian disability scheme defines eligibility criteria by looking at age and occupation. For blue-collar workers below age 55 (57 after 1996) it is required a 50% reduction in earning capacity in any reasonable occupation the individual is able to carry out. Whereas, for all white-collar workers and blue-collar workers above 55 (57 after 1996) the reference for an occupational group is more limited to a similar occupation. As a result, for older workers is easier to access the benefits. In addition, since women had the opportunity to access early routes at the age of 55, men tended to access disability routes more frequently. Another important aspect of Disability Benefits is that the program is a means-tested program (in 2010 the maximum threshold allowed to earn was 360 euros per month). As a result, only 15% of DI recipients worked.

Alternative income replacement schemes are the sickness insurance system and the unemployment benefits. Both replacement rates are comparable with disability benefits, i.e., around 55%.

[37] Women's statutory retirement age will be gradually raised to 65 by 2033.

In 2014 a new law reformed disability insurance with the aim to re-integrate health-impaired people into the labour market.

Belgium

The Belgian pension system runs on a pure PAYG basis. In 2017, the old-age retirement age was set up at 65 years with a replacement rate hovering between 60 and 75%. In addition, the early route could be accessed at 62 years and a half if no less than 41 years of contributions have been completed.

Jousten and Lefebvre (2013) point out that along with the "traditional" disability insurance, in the last decades, Belgium was characterized by several alternative compensating schemes for health-impaired individuals, i.e., Unemployment Insurance, Old-age Unemployment Insurance, Conventional Early Retirement, Special Early Retirement and Retirement Pre-pension. The disability benefit eligibility conditions to fulfil are the loss for at least 12 months of 66% of earning capacity in the usual job and at least half of a year of contributions. The system is financed as a pay-as-you-go. Replacement rates depend on the household composition ranging from 40 to 65%. In addition, the return to work is conditioned on medical approval and the program is means-tested since a cap on benefits is applied for high levels of income[38]. Unemployment Insurance requires active contributions periods to access and assures a 60% of replacement rate. Old-Age Unemployment Insurance is more generous and does not require active job searching. All the other early retirement schemes are(were) mainly created with the aim of turnover old individuals into unemployed ones.

Denmark

The public pension scheme in Denmark is universal and covers the entire population. Eligibility for pension is based on residency in Denmark and does not depend on the payment of contributions. Specifically, the pension scheme is public and includes a basic pension. In addition, there is a mandatory occupational pension scheme known which is funded through lump-sum contributions. After the 1999 reform which reduced the minimum age for the old-age pension from 67 to 65, recently, it was gradually raised to 69 for all those born after 1967. Early retirement can be accessed at a maximum of three

[38] https://www.inami.fgov.be/fr/themes/incapacite-travail/montants/salaries-chomeurs/Pages/indemnite-diminution-activite-autorisee-salarie.aspx

years before the old age threshold and it is conditioned on contribution years and year of residence plus nationality considerations[39].

Bingley et al. (2012) point out that in the period 1970-2008, at least eight different programs for early retirement have been available in Denmark (although not all during the same years). The most important program which aimed to protect health-impaired individuals is the Social Disability Pension (SPD). SPD was financed by general tax revenues and eligibility depended both on medical and social criteria. In 1984 smaller programs have been merged with SPD and three levels of disability have been proposed. A 2003 reform tightened eligibility criteria to access disability benefits. Furthermore, in 2014 a new disability pathway was opened for workers with five years less than the old-age requirement and reduced working capacity due to work-related ill health. One specificity in Denmark's legislation is that local authorities determine the entitlement to Disability Pension.

Historically, the most important alternative early retirement scheme was the Post Employment Wage (PEW). It started in 1979 and targeted individuals aged 60-66 with enough contributions to unemployment schemes. The rationale behind this was to open an early retirement route for hard physical workers without having to fulfil the formal medical criteria of disability benefits. Simultaneously, an additional goal was to reduce younger unemployed workers. Hence, it was a means-tested program which allowed only 200 employable hours per year. Another program, called Transitional Benefits Program, was present from 1992 to 1996 and targeted 50-59 unemployed workers. Moreover, in 1998, the Flex Job Program was introduced with the aim to foster work opportunities for disabled people.

France

The French retirement system is based on a pay-as-you-go basis. The old-age full-rate entitlement is available at ages 65-67 according to the claimant's year of birth and circumstances. It is still possible to retire at a full rate at age 62 in case of a very long contributory career. The system proposes early retirement routes based on long-term careers, arduous work and disability. The disability route into retirement is available at 55-59 years with at least 50% of permanent earning reduction due to health impair-

[39] All details can be viewed at https://www.borger.dk/

ments[40].

Behaghel et al. (2012) points out that the relatively low ages required for "normal" retirement in the French system, avoided mass access towards the disability insurance route. In addition, other alternative compensatory schemes were "traditionally" employed to transit towards retirement, i.e., Unemployment Insurance and other preretirement schemes.

Germany

The main pillar of German old-age security is general pension insurance. The statutory old age pension is at age 66 which will be gradually raised to 67, with some exceptions for long-term insured. It is possible to access an Old-age Pension for Severely Disabled Persons at age 60 or more with at least 50% reduction of working capacity and no less than 35 years of contributions[41].

To access invalidity benefits, at least three years of contributions in the last five years are required, with a minimum of five years. Benefits depend on the degree of disability and are capped in case of high income[42]. Börsch-Supan and Jürges (2012) point out that in 2010 almost 7% of 50-64 Germans received Disability Insurance benefits. One alternative retirement route is accessible through the Unemployment Insurance channel. It provided replacement rates hovering between 65% and no job-search activities were requested in case of old age. Moreover, Börsch-Supan and Jürges (2012) show that the disability route was an important route to exit the labour force. Reforms gradually decreased its importance being substituted by the Unemployment route for men and by special retirement routes for women.

Greece

The Greek pension system is a public scheme which provides a non-earnings related part based on years of residency (basic part) and a contributory part depending on the years of insurance. The statutory retirement is 67 with at least 15 years of contributions or 62 in the case of 40 years of contributions. Early retirement is possible from age 62 with a benefit reduction. According to OECD, in 2018 the Greek public pension spending in the

[40] Further details on the Social Security System can be found on https://www.cleiss.fr/docs/regimes/regime$_{f}rance/an_{3}.html$

[41] https://www.deutsche-rentenversicherung.de/DRV/EN/Leistungen/leistungen.html

[42] https://www.deutsche-rentenversicherung.de/DRV/DE/Muttertexte/04$_{l}eistungen/01_{r}ente/erwerbsminderung.htm$

percentage of GDP was 16.9, one of the highest in the world. The leniency in normal and early retirement routes does not raise particular concerns about disability route usage.

Italy

The Italian pension system is based on a compulsory pay-as-you-go basis which is gradually moving towards a notional defined contribution scheme (NDC)[43]. The statutory retirement age for all is set to 67 and a minimum of 20 years of contributions are required. In 2022 it is expected an early retirement route at the age of 64 with no less than 38 years of contributions. Earlier options are available for women (Opzione Donna). Some early route is available at age 63 in case of at least 30 years of contributions and being disabled, unemployed or a caregiver (Ape Sociale).

Brugiavini and Peracchi (2012) document that the availability of routes towards early retirement decreased the importance played by Disability Insurance benefits. As for the Greek case, a large fraction of GDP (more than 15%) is spent on the public pension system. The Disability Insurance system provides two kinds of benefits: a contributory and a welfare benefit. On the one hand, the contributory benefit is non-means-tested and depends on contributory history. On the other hand, the welfare benefit is a means-tested program which provides income replacement for health-impaired individuals in need. In 2017, 1.2 million benefits have been provided in the contributory scheme and 4.3 million in the welfare scheme.

Netherlands

In 2022, the Netherlands' statutory retirement age is 66 and 7 months which will gradually move to 67 in 2024. The entitlement and the benefit of full old-age retirement depend on the years lived or worked in the Netherlands. There is no general early retirement channel. In addition, the leniency in Disability Insurance legislation led to a clear framework in which Disability Insurance acted as an early pathway to retirement, which only recently has been reformed. The overall evolution of the rise and fall of disability is well documented in Koning and Lindeboom (2015). The Dutch disability program covers all workers against health impairments and was particularly generous in timing and amount. If compared with Unemployment Insurance benefits, both replacement rates were around 70%. The main difference was that the disabled individual had no job search requirement. In addition, also employers had incentives in preferring disability channels instead

[43]The NDC will be fully effective for people who entered the labour market after 1995.

of unemployment for their employees. Thus, both employers and employees had a mutual interest in exploiting Disability Insurance as a substitute pathway into early retirement. As a result, from 1980 onward around 12% of those insured received a Disability Insurance benefit. Several reforms have been proposed: change in employer incentives (1996, 1998), tightening in screening procedures (2002), stricter eligibility criteria and more work incentives (2006). All these reforms were able to almost halve the benefit recipients.

Spain

The Spanish pension system relies mainly on the first pillar. The main component involves an earnings-related benefit, whereas a means-tested minimum pension is allowed for those who have not acquired enough pension contributions. Between 2013 and 2027, the statutory retirement age will increase from 65 to 67 years. Early retirement is possible 4 years before the statutory age in case of unemployment and long contribution history, with actuarial benefit reduction. In addition, are available further partial and flexible retirement schemes OECD.

Spain provides two types of safety nets for health-impaired individuals. The first one is non-means-tested and is based on the contributions to the social security system. The second one is managed at the regional level, is means-tested and does not consider contribution conditions. The non-contributory program is almost 4 times smaller with respect to the contributory one. Moreover, the benefits are relevantly smaller, i.e., the non-contributory assures 400 euros versus roughly 800 (García-Gómez et al. (2012)). Alternative compensating programs are sick leave (or temporary disability) and Unemployment Insurance. In addition, García-Gómez et al. (2012) found that Disability Insurance and old age pension were clearly substituted after the 1997-2002 government reforms.

Sweden

Sweden's retirement pension relies on a pay-as-you-go notional accounts system and a mandatory funded defined contribution pension. Moreover, the occupational pension plans with defined benefits and defined contribution elements have broad coverage (more than 90% of employees is covered) OECD. No formal statutory age is defined. The minimum age to access retirement was 61 in 2018 and will gradually increase to 64 by 2026. The benefits are automatically adjusted by actuarial reduction depending on the age of retirement. The "right" to work is allowed until age 68 and can be postponed only with the employer's agreement. One interesting aspect of the Swedish framework is that some occupational pensions are paid out automatically when the age of 65 is reached.

Sweden has always been particularly generous in allowing Disability Insurance benefits. For instance, Jönsson et al. (2012) point out that in 2009, 20% of men and 30% of women aged 60-64 received disability benefits. In fact, the main common early pathway to retirement is ascribable to Disability Insurance. The replacement rate was almost 60%.

Switzerland

The statutory retirement age in Switzerland is 65 for men and 64 for women. It is possible to retire 2 years earlier than statutory ages, but with a penalty on benefits[44]. Disability Insurance is a compulsory social insurance scheme for all persons living in Switzerland [45] and is quite generous (Deuchert and Eugster (2019)). A minimum of a permanent earning reduction of 40% is required to access the disability benefit. In case of a full benefit (at least 70% of earning reduction), the replacement rate hovers around 60%. Benefits are a stepwise function of the disability degree which is computed by caseworkers who evaluate it by looking at medical conditions and earning history. The benefit is provided according to a non-means-tested program (??), although the benefit is adjusted according to the earnings.

[44] In 2021, one year earlier with respect to statutory age corresponds to a 6.8% reduction, and to 13.6% if the retirement happens two years early.

[45] https://www.ch.ch/en/retirement/retirement-income/

D Survey structure

Code	Self-declared situation
1	Employee or self-employed
2	Unemployed and searching for a job
3	Unemployed but not searching for a job
4	Short-term job (less than 6 months)
5	Sick or disabled
6	Looking after home or family
7	Leisure, travelling or doing nothing
8	Retired from work
9	Training
10	In education
11	Military services, war prisoner or equivalent
12	Managing your assets
13	Voluntary or community work
14	Forced labour or in jail
15	Exiled or banished
16	Labor camp
17	Concentration camp
18	Other

Table 12: Alternative self-declared situations in SHARELIFE interviews.

D.1 Comparison with regular interviews

Evidence about the goodness of the "broader" definition of disability comes from the comparison with SHARE regular interviews, which, it is worthy to recall, rely on labour market questions asked years before the retrospective interview. From table 13 it is clear that the retrospective states are well resembling the information available from regular waves. Although there is no perfect match, likely due to recall bias, or simply ambiguous states[46], both DW and DO states seem to be in line with regular waves' information.

[46]E.g. an individual who is unemployed, employed and retired during the same year may answer in 3 different ways according to which is the state at the moment of the interview or which is the a posteriori perception.

DW captures mostly individuals who are actively participating in the labour market, whereas DO captures individuals out of the labour market and a part of those declaring unemployment (389 observations). Among those permanently sick or disabled in regular waves, almost 40% (727 observations) perceive themselves as retired, i.e., permanently outside the labour market.

	Employed or self-employed	Unemployed	Homemaker	Other	Permanently sick or disabled	Retired
Work	27611	852	576	282	221	520
Other	614	1166	3902	230	206	340
Disability Work	2127	132	62	46	229	58
Disability Other	110	389	550	76	1286	174
Retired	1308	418	678	303	727	24387

Table 13: Alternativa A: original ep005 states. Sample containing all observations of individuals starting from age 10 to age 70 or the age of the interview.

E Supervised Learning

In this section, we provide an overview of the supervised learning methods employed in our analysis. For more detailed information, we refer readers to the works of Albanesi and Vamossy (2019) or James et al. (2013). In supervised learning, a learner receives in pairs input/output data. If the output data is continuous, then the supervised learning problem at hand is a ”regression problem”. If the output data is categorical (as it is in our case), the problem becomes a ”classification problem”. The task of the learner is to find a function that maps the input data to the output values. In order to do so, supervised learning algorithms employ a subset of input/output data as a training sample, to learn to make predictions, and the remaining subset of input/output data referred to as the test sample is kept separate to test the predictions, i.e., to perform out-of-sample analysis (Bargagli-Stoffi et al. (2021)). It is noteworthy to mention that, the goal of supervised learning is to find a function that generalizes beyond the training data[47]. This is the well-known variance-bias trade-off in which an excessive fit between input/output in training data corresponds to in-sample low bias but with the risk of poor predictions in

[47]One straightforward strategy for the learner would be to memorize exactly the value of input/output pairs and so as to perfectly map input and output data in the training data set. However, in case of noise or completely new data, forecasts' quality is likely to fall.

case of never seen before data. In order to tackle this issue, the performance of learners is tested out of the sample. In addition, the more advanced algorithms are optimized by introducing "regularization terms" with the aim to reach a greater generalization of the learnt function.

From an operational point of view, learners "learn" by following different steps according to which algorithm is taken into account. Nevertheless, all supervised learning problems are set up following the same structure. First of all, according to which is the problem at hand, an objective function is chosen. Afterwards, the objective function is optimized following the technique which is more suitable for the task[48].

A generic objective function is represented as follows:

$$Objective\ function = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(f) \tag{5}$$

where i is the observation, $l()$ is the training loss function, y_i is the output variable for observation i whereas $\hat{y}_i$ is the predicted output. Ω is a generic regularization term which penalizes model complexity.

eXtreme Gradient Boosting (XGBoost)

The supervised learning framework which is proposed is based on eXtreme Gradient Boosting (XGBoost), a leading model for working with standard tabular data (Chen and Guestrin (2016)). In brief, XGBoost exploits decision tree models (outlined below).

Decision Tree Models

Tree-based models partition the data several times based on certain cutoff values in the input data. In different words, distinct partitions of the data set are created and each observation is part of one and only one partition. The literature proposes several decision tree models. The most well-known decision tree models are Classification and Regression Trees, i.e., CART (Breiman et al. (1984)), and Random Forest (Breiman (2001)). For illustrative reasons, the functioning of the Classification and Regression Trees (CART) is shown.

[48] For a good introduction to the topic the reader is invited to look after Hastie et al. (2009).

CART

Given a pair of input/output data x_i/y_i of an observation i, CART models the outcome y_i as

$$\hat{y_i} = \hat{f}(x_i) = \sum_{m=1}^{M} c_m I\{x_i \in R_m\}, \tag{6}$$

where each observation x_i belongs to exactly one subset R_m. The indicator function I is equal to one if x_i is in R_m and 0 otherwise. If x_i is in R_l, then the predicted outcome is $\hat{y} = c_l$, where c_l is the mean of all training observations y in R_l. In brief, tree-based models partition the data several times based on certain cutoff values in the input data. This procedure is performed by taking each variable in the data set and computing the cut-off point that minimizes the loss function of the distribution of y, which makes the two resulting subsets as different as possible. Whenever the procedure is performed on all variables, the algorithm split the data into two subsets according to which is the best candidate (the one which reduces the most loss). The algorithm is repeated until a stopping criterion is reached.

Like all methodologies, tree-based models present advantageous and disadvantageous characteristics. Among the advantages it is worth mentioning that tree-based models are invariant to monotonic transformations, are well suited to capture interactions between variables in the data and any single tree is highly interpretable.

Nevertheless, tree-based models present limitations too. The first one relates to the fact that tree algorithms partition data using step functions, which are intrinsically non-linear transformations, and as a consequence, they are poor at detecting linear relationships. Next, since small changes in the training data set could generate a completely different tree, they tend to be unstable. Finally, trees tend to overfit training data. Hence, in order to overcome these drawbacks, further developments were proposed. Among these developments, "Gradient Boosted Trees" plays an important role.

Gradient Boosted Trees

Gradient Boosting Trees is an ensemble learning methodology which is characterized by a combination of several over-simplified trees (called "weak learners") that combined form a "strong learner". Albeit each over-simplified tree ("weak learner") has little predictive power if considered alone, the boosting theory shows that an ensemble of these weak learners generates a "strong learner" which outperforms a single complex tree. All gradient-boosting tree algorithms (including XGBoost), are built in order to overcome the tendency of the basic tree-based models to overfit the training data.

In detail, Gradient Boosted Trees exploits an additive strategy, i.e., at each step m $(1 \leq m \leq M)$ an estimator function[49] f_i is computed on the residuals from the previous models' predictions.

For instance, in order to write the prediction value at step m as $\hat{y}_i^m$

$$
\begin{aligned}
\hat{y}_i^{(0)} &= 0 \\
\hat{y}_i^{(1)} &= \hat{y}_i^{(0)} + f_1(x_i) \\
\hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\
&\cdots \\
\hat{y}_i^{(m)} &= \sum_{j=1}^{m} f_k(x_i) = \hat{y}_i^{(m-1)} + f_m(x_i)
\end{aligned}
$$

The above equations show that the prediction after m steps, $\hat{y}_i^m$, is found using an additive strategy by keeping what was previously learned and by adding a new tree at a time. At each step, it is chosen the estimator function that optimizes the objective function. In the general case, a Taylor expansion of the loss function up to the second order is taken and the objective function is written only in terms of first and second derivatives. In such a way, every loss function, including the logistic and multinomial logistic which are exploited in the next steps, can be employed.

Loss function

So far, it has been discussed how XGBoost, and all the Gradient Boosted Trees, exploit an additive strategy to find the optimal function which maps input to output data. To better explain how at each step the optimal function is chosen, it is necessary to introduce in more detail the objective function which is optimised. The objective function which is proposed, from now on will be called Loss Function, is the standard choice whenever a (multi)classification problem is tackled, i.e., (multinomial) logistic regression loss[50]. In

[49] Which corresponds to a shallow tree, which can also be called "weak learner".

[50] Alternatively called (m)log loss or cross-entropy loss.

brief, the (m)log loss uses the output probabilities and measures the distance from the true values.

Binary classification

In the baseline problem in which only two states of the world are possible $y \in \{0, 1\}$, and the probability estimate is $p = Pr(y = 1)$, the log loss of each observation i is the negative log-likelihood of the classifier given the true state of the world:

$$L_{log}(y_i, p_i) = -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \tag{7}$$

For instance, let's assume that $y_i = 1$, then only the first term is relevant in determining the loss. Since $p_i \in [0, 1]$, if p_i is close to 0, then the loss explodes. On the contrary, if the algorithm is able to correctly assign a large value to p_i, then the loss tends to be 0. The negative sign is needed in order to change the sign into a positive and to be able to interpret the result as a loss.

Multiclass classification

However, not always the problem contains only two alternative states of the world. It is quite common to face a multiclass case, i.e., more than two mutually exclusive states of the world are possible. Given a general case with K states of the world, if $y_{i,k} = 1$, then observation i is in state k. Moreover, if $p_{i,k} = Pr(y_{i,k} = 1)$ is the probability estimate that observation i is in state k, then the multinomial log loss for the whole set is:

$$L(Y, P) = -\frac{1}{N} \sum_{i}^{N} \sum_{k}^{K} y_{i,k} log p_{i,k} \tag{8}$$

Hence, at each step of the procedure presented above, the algorithm behind XGBoost updates the probabilities in such a way the expected probability distribution matches as much as possible the "real" distribution. From a broad point of view, the main difference with the already proposed methods in literature is that instead of using just one estimation of a (multinomial) logistic model carefully set up in order to capture non-linearities, boosting models use an ensemble of "weak learners" which capture non-linearities by additively updating the functional form.

E.1 Hyper-parameters tuning

Existing literature demonstrates that random forest algorithms exhibit reduced sensitivity to hyperparameter tuning when the covariate space is relatively small (Probst et al. (2019)). However, despite this, we performed hyperparameter tuning using both a standard grid search and a Bayesian Optimization function for XGBoost Nogueira (2014–). We did not observe any enhancement in predictive performance. Therefore, to alleviate the computational burden, we opted to utilize the baseline hyperparameters.

F Robustness

F.1 Comparison with logistic regression

In this robustness subsection, we replicate the main analysis by employing standard logistic regression to estimate the TPM, followed by simulating the main policy reforms.

Table 14 displays the estimates of the average number of spells and duration of spells obtained with logistic regression using 200 out-of-sample predictions. Furthermore, a comparison is made with the main XGBOOST estimates. The ratios of the predicted and observed averages of the number and duration of spells are calculated. It is important to note that the optimal prediction would yield ratios closest to 1. The last column of table 14 presents a rough measure of the excess error in the logistic regression prediction with respect to the XGBOOST. Positive values indicate a larger deviation of logistic regression from the optimal target 1 with respect to XGBOOST, suggesting that logistic regression performs worse in terms of predicting both the number of spells and their duration (and viceversa). The sum of the excess error for both the number of spells and duration is 0.19. This discrepancy is primarily driven by the challenge of accurately predicting the number of spells and duration in the "Other" category for men. Next, table 15 presents the probabilities of exit transitions between ages 45 and 70 based on 200 out-of-sample replications. Similarly to the previous analysis, the excess error in predicting transitions is shown in table 16. The sum of the excess error is 0.63, indicating that logistic regression performs worse in predicting the transitions compared to XGBOOST. The out-of-sample stochastic simulations conducted using logistic regression estimation perform worse compared to the baseline XGBoost approach . Figure 16 (compared to Figure 8) demonstrates that logistic regression underestimates the proportion of working men and overestimates it for women. Additionally, the proportion of retired individuals is consistently underestimated for both men and women.

		Logistic: Predicted		Logistic: Observed		Ratio				Excess error	
						Logistic		XGBOOST			
Gender		N spells	Duration	N spells	Duration	N spells	Duration	N spells	duration	N spells	Mean duration
MEN	W	1.151.20	21.49	1.089.72	22.76	1.06	0.94	1.03	0.98	0.03	-0.04
	O	332.75	4.82	278.18	4.10	1.20	1.17	1.05	1.02	0.15	0.15
	DW	737.78	5.80	732.48	6.20	1.01	0.93	1.02	0.95	-0.01	-0.02
	DO	286.78	8.97	260.13	8.67	1.10	1.04	1.06	1.03	0.04	0.01
	R	459.63	9.21	496.94	9.44	0.92	0.98	0.99	1.00	0.07	-0.02
WOMEN	W	1.326.88	15.68	1.327.29	15.16	1.00	1.03	1.01	1.00	-0.01	0.03
	O	625.14	6.89	619.02	8.08	1.01	0.85	1.05	0.90	-0.04	0.05
	DW	774.80	6.14	780.12	5.86	0.99	1.05	1.00	1.02	0.01	0.03
	DO	424.66	9.63	414.70	10.13	1.02	0.95	1.05	0.94	-0.03	-0.01
	R	428.09	9.11	424.12	9.31	1.01	0.98	1.03	0.99	-0.02	0.01
									Sum	0.19	0.19

Table 14: Cross validation estimates of the average duration of spells obtained with Logistic regression and comparison with main estimates (XGBOOST). 200 out-of-sample predictions, Men-Women aged 10-70. The Excess error measures the difference in error prediction between Logistic regression and XG-BOOST (a positive value means a larger deviation of the Logistic regression with respect to XGBOOST from the optimal target).

Gender	Origin	Uncen spells	W	O	DW	DO	R
MEN	Pred W	602.60	NaN	0.04	0.63	0.02	0.31
	Obs W	627.63	NaN	0.05	0.62	0.02	0.32
	Ratio	0.96	NaN	0.82	1.02	1.04	0.99
	Pred O	53.06	0.40	NaN	0.09	0.25	0.26
	Obs O	43.58	0.37	NaN	0.09	0.24	0.30
	Ratio	1.22	1.08	NaN	1.01	1.05	0.85
	Pred DW	444.24	0.26	0.02	NaN	0.34	0.38
	Obs DW	463.21	0.23	0.02	NaN	0.32	0.43
	Ratio	0.96	1.15	1.20	NaN	1.04	0.89
	Pred DO	120.17	0.04	0.08	0.20	NaN	0.68
	Obs DO	115.43	0.05	0.04	0.19	NaN	0.72
	Ratio	1.04	0.84	1.87	1.07	NaN	0.94
WOMEN	Pred W	569.59	NaN	0.10	0.60	0.03	0.27
	Obs W	558.97	NaN	0.09	0.65	0.03	0.24
	Ratio	1.02	NaN	1.20	0.92	1.04	1.13
	Pred O	130.47	0.32	NaN	0.13	0.34	0.21
	Obs O	124.45	0.37	NaN	0.10	0.31	0.22
	Ratio	1.05	0.85	NaN	1.33	1.12	0.95
	Pred DW	443.75	0.26	0.03	NaN	0.40	0.31
	Obs DW	460.61	0.24	0.03	NaN	0.39	0.34
	Ratio	0.96	1.09	1.13	NaN	1.01	0.91
	Pred DO	162.50	0.03	0.12	0.23	NaN	0.62
	Obs DO	175.97	0.03	0.10	0.24	NaN	0.63
	Ratio	0.92	1.10	1.28	0.95	NaN	0.97

Table 15: Cross-validation estimates of average transition probabilities estimated by Logistic regression. 200 out-of-sample replications, Men-Women aged 45-70. 75% of the sample was used as the training sample and the remaining 25% as the test sample.

Gender	Origin	W	O	DW	DO	R	Sum
MEN	Ratio W	NaN	0.82	1.02	1.04	0.99	
	Ratio W XGBOOST	NaN	0.95	1.01	0.88	1.00	
	Excess error W		0.13	0.01	-0.08	0.01	0.07
	Ratio O	1.08	NaN	1.01	1.05	0.85	
	Ratio O XGBOOST	1.10	NaN	0.76	0.97	0.97	
	Excess error O	-0.02		-0.23	0.02	0.12	-0.11
	Ratio DW	1.15	1.20	NaN	1.04	0.89	
	Ratio DW XGBOOST	1.14	0.95	NaN	1.00	0.92	
	Excess error DW	0.01	0.15		0.04	0.03	0.23
	Ratio DO	0.84	1.87	1.07	NaN	0.94	
	Ratio DO XGBOOST	0.72	1.55	0.86	NaN	1.02	
	Excess error DO	-0.12	0.32	-0.07		0.04	0.17
WOMEN	Ratio W	NaN	1.20	0.92	1.04	1.13	
	Ratio W XGBOOST	NaN	1.13	0.97	0.87	1.03	
	Excess error W		0.07	0.05	-0.09	0.10	0.13
	Ratio O	0.85	NaN	1.33	1.12	0.95	
	Ratio O XGBOOST	1.00	NaN	0.89	1.10	0.92	
	Excess error O	0.15		0.22	0.02	-0.03	0.36
	Ratio DW	1.09	1.13	NaN	1.01	0.91	
	Ratio DW XGBOOST	1.06	0.87	NaN	1.01	0.96	
	Excess error DW	0.03	0.00		0.00	0.05	0.08
	Ratio DO	1.10	1.28	0.95	NaN	0.97	
	Ratio DO XGBOOST	0.71	1.37	0.90	NaN	1.00	
	Excess error DO	-0.19	-0.09	-0.05		0.03	-0.30
						Sum	0.63

Table 16: The Excess error measures the difference in error prediction between Logistic regression and XGBOOST (a positive value means a larger deviation of the Logistic regression with respect to XGBOOST from the true value)

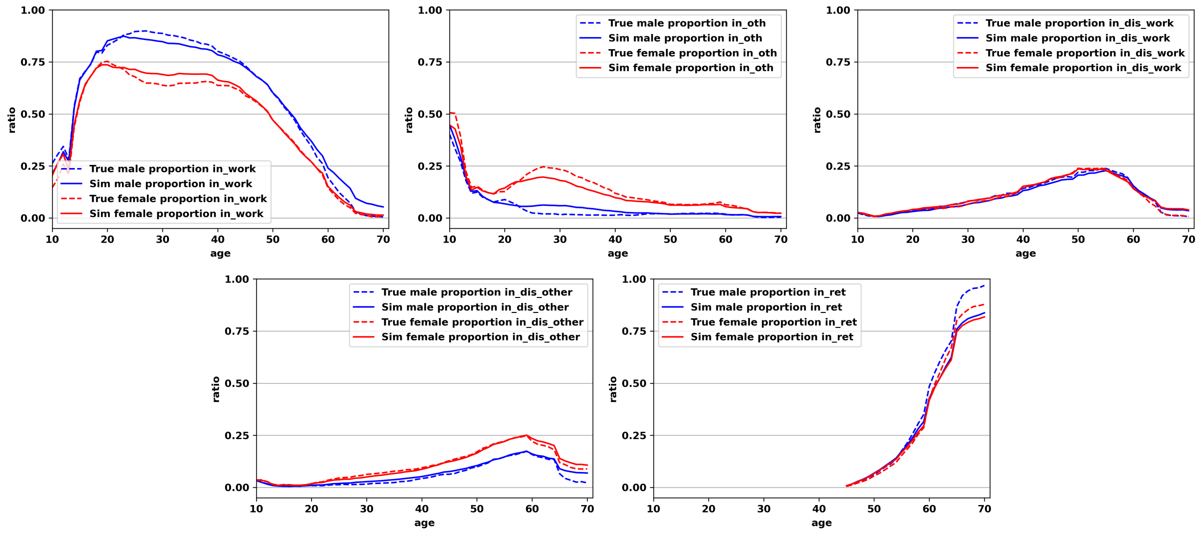

Figure 16: Observed and predicted age-profiles of binary indicators for labour market states of men and women using logistic regression for estimating building blocks. 200 out-of-sample replications, Men-Women aged 10 - 70. 75% of the sample used as training sample and the remaining 25% as test sample.

Despite the inferior predictive performance of logistic regression, it is still worth exploring if the simulations would yield significantly different results. Figures 17 and 18 illustrate the impact of abolishing the disability and early retirement schemes, respectively. The results remain robust compared to the main analysis, although the effects appear to be less pronounced if compared with the baseline figures 9 and 10.

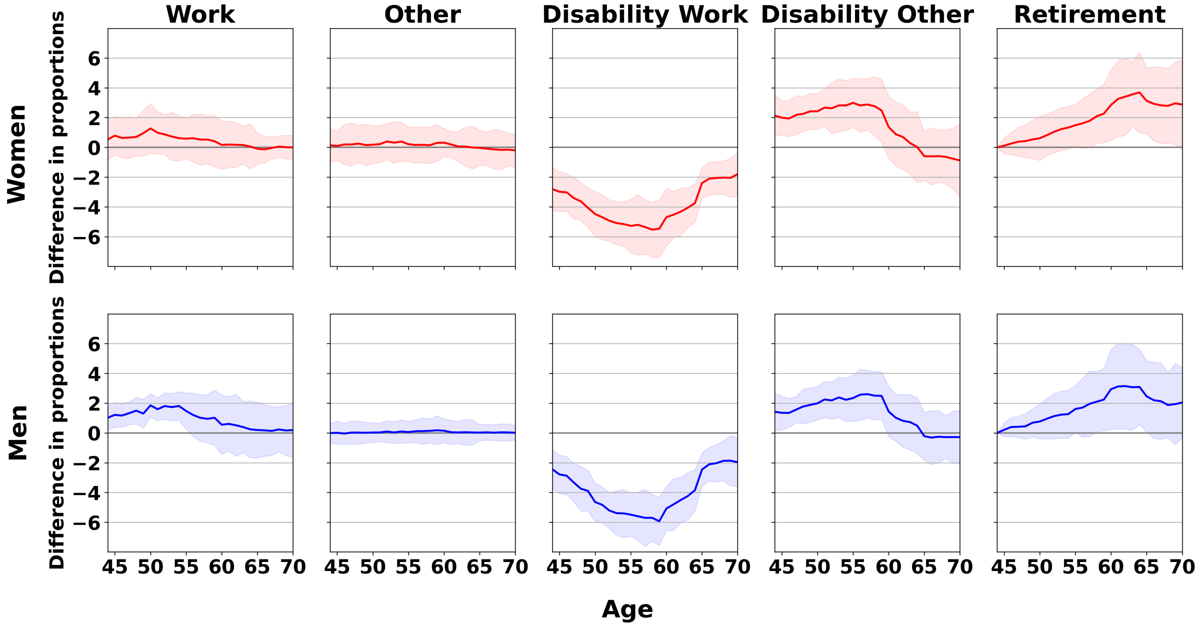

Figure 17: Effect of the abolition of Disability Insurance benefit by employing logistic regression in estimating building blocks. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

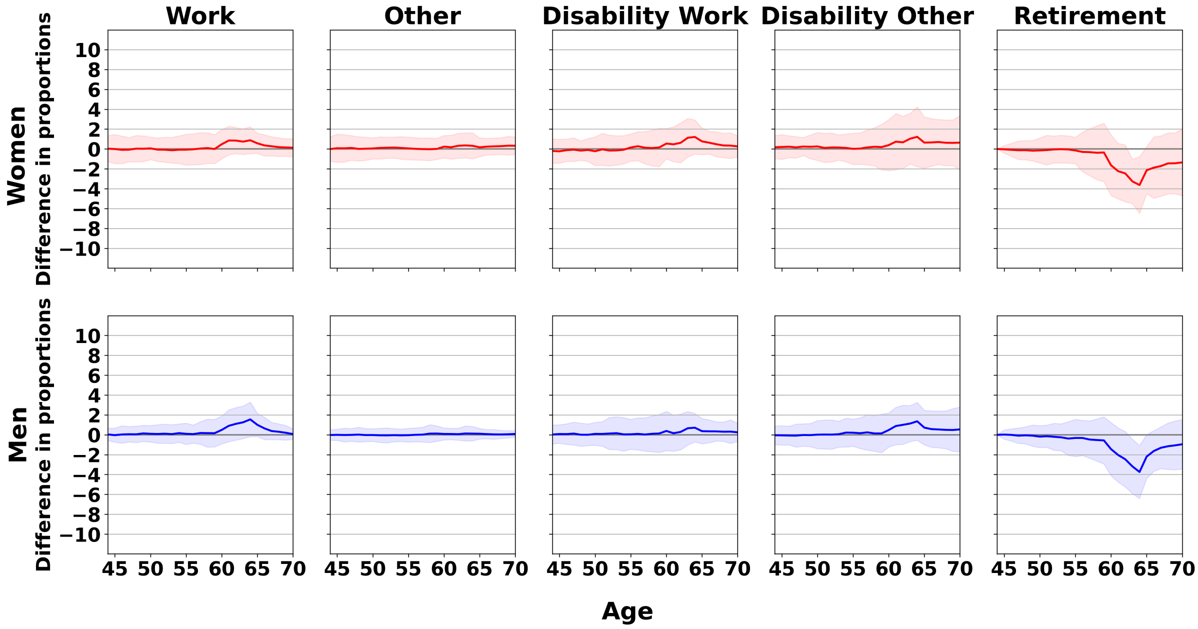

Figure 18: Effect of the abolition of Early Pension by employing logistic regression in estimating building blocks. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

F.2 Milder policy reforms

To approximate a milder policy reform instead of completely abolishing disability insurance, we simulate a moderate restriction on the eligibility requirements. In Figure 19, we present the results of a simulation where we increase the minimum required RWC by 10 percentage points (e.g., from 66% to 76%). This adjustment reflects a more realistic approach to limit access to disability benefits. Similarly, instead of completely closing the early retirement route, we simulate the impact of a one-year delay in the early retirement age. Figure 20 illustrates the outcomes of this milder policy change.

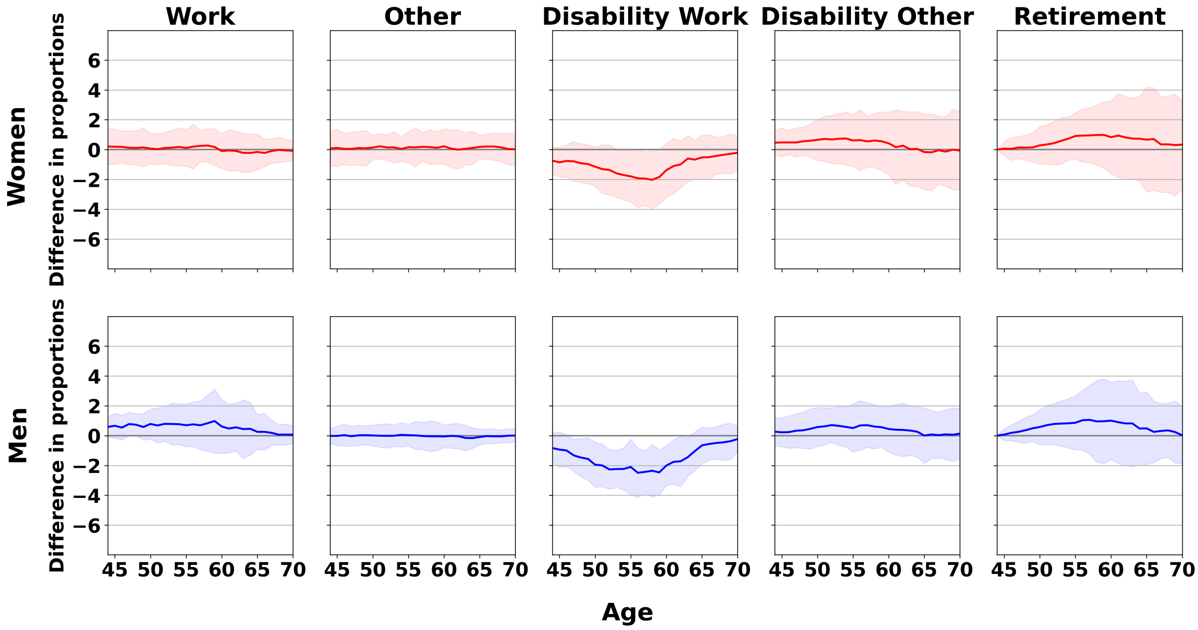

Figure 19: Effect of 10% increase in legal RWC requirements to be eligible for Disability Insurance benefits. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

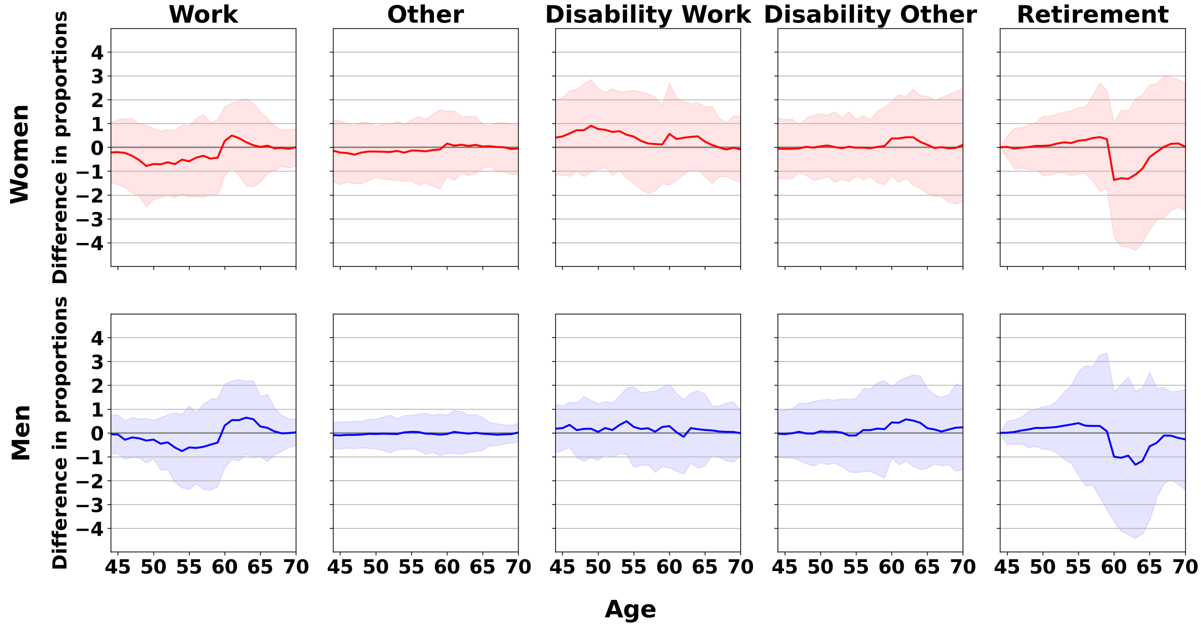

Figure 20: Effect of one-year delay of requirements for early pension eligibility. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

F.2.1 Alternative definition of RWC

Another concern might regard the way in which the RWC was defined for each individual. In this regard, several adjustments of the baseline RWC have been performed and results remained robust. In figure 21 are shown the results of a more "conservative" assignment of RWC.

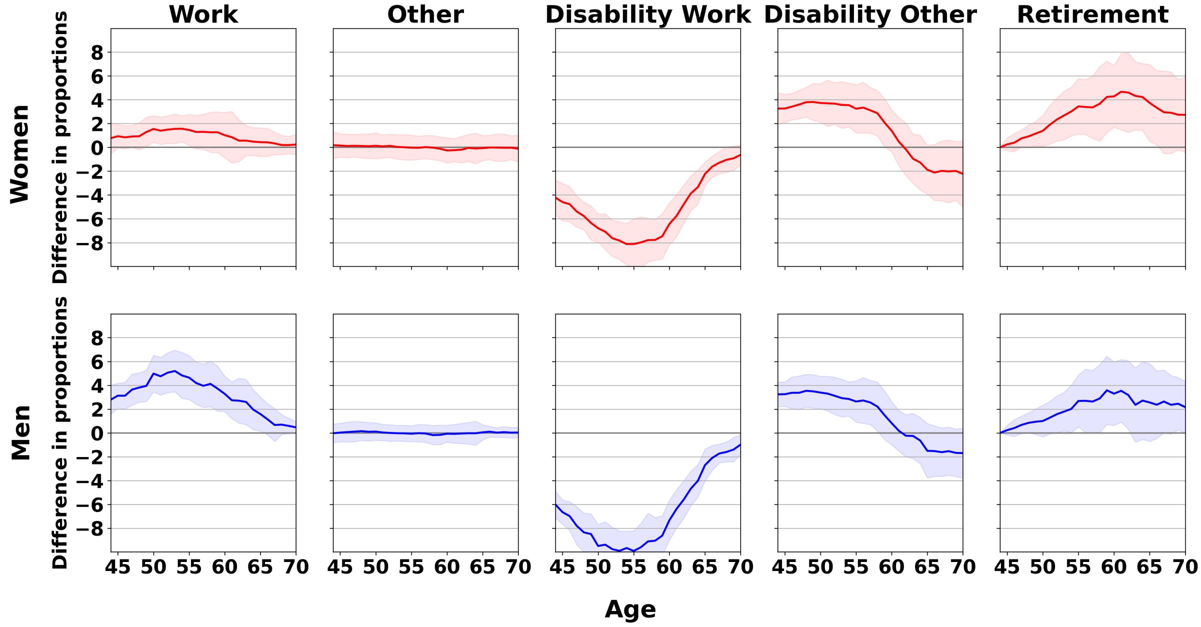

Figure 21: Effect of Disability Insurance benefit abolition employing a more conservative approach to measure RWC, i.e., 100% of RWC randomly assigned to 75% to end up with a more "uniform" distribution. Women in red, men in blue. Mean results after 30 simulations, 95% confidence interval from a stratified nonparametric bootstrap procedure.

G SHAP for Explainable Machine Learning

The increasing demand for explainability in machine learning algorithms has driven the development of new approaches. SHapley Additive exPlanations (SHAP Lundberg and Lee (2017)) is a technique that explains XGBoost model predictions. It assigns individual feature contributions to each prediction, showing how much each feature affects the model's output for a particular data point. By adhering to cooperative game theory principles, SHAP values ensure a fair distribution of the model's prediction among the features based on their significance. This approach provides valuable insights into the model's decision-making process and the impact of various features on the predicted outcomes.

Figure 22 illustrates the contributions to each prediction for the Instantaneous Survival probabilities, ordered from the most important. Notably, gender emerges as the primary explanatory variable for remaining in the Working state. Additionally, having a new child significantly increases the probability of transitioning to the Other state. Concerning disability states, it is observed that a larger RWC (100 or 75%) is associated with a higher probability of staying in Disability Other, whereas for Disability Work, lower

RWC values play a crucial role. Moreover, the years spent in the different states of the world consistently exhibit importance, underscoring the significance of path dependency.

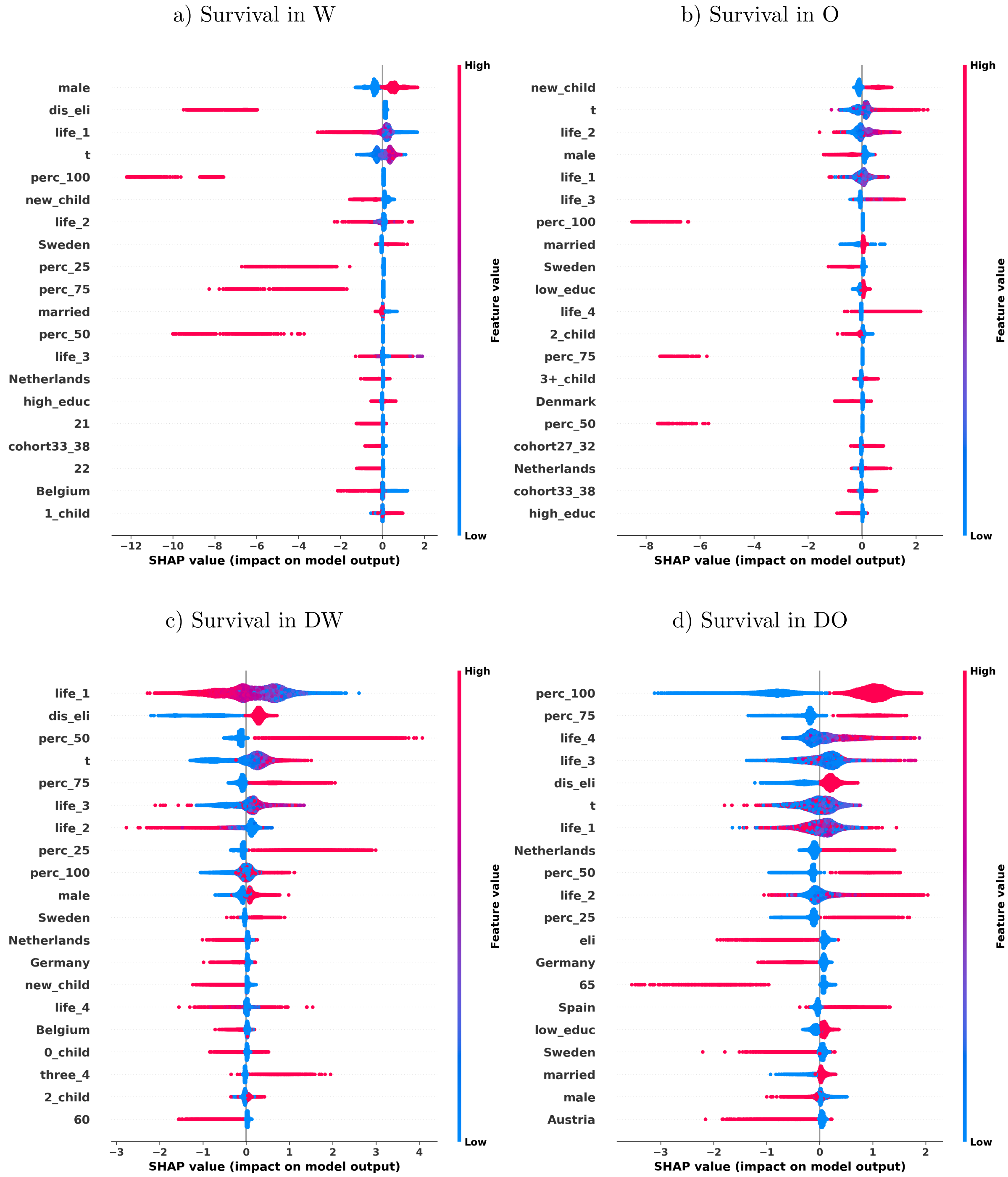

Figure 22: SHAP (SHapley Additive exPlanations, (Lundberg and Lee (2017) for Instantaneous Survival Probability in Work. SHAP is a game theoretic approach to explain the output of any machine learning model.

H Data

Variable	Full sample		Restricted sample	
	mean	std	mean	std
age	36.97	16.78	36.82	16.68
male	0.50	0.50	0.49	0.50
married	0.62	0.49	0.61	0.49
new child in last 3 years	0.10	0.30	0.10	0.30
low education	0.41	0.49	0.44	0.50
middle education	0.36	0.48	0.37	0.48
high education	0.23	0.42	0.19	0.39
cohort 27-32	0.07	0.25	0.06	0.24
cohort 33-38	0.13	0.34	0.12	0.33
cohort 39-44	0.21	0.41	0.21	0.40
cohort 45-50	0.24	0.43	0.25	0.43
cohort 51-56	0.21	0.41	0.22	0.42
cohort 57-62	0.12	0.32	0.12	0.32
cohort 63-68	0.02	0.14	0.02	0.14
Austria	0.08	0.28	0.08	0.27
Germany	0.11	0.31	0.14	0.35
Sweden	0.09	0.28	0.11	0.32
Netherlands	0.05	0.22	0.09	0.29
Spain	0.11	0.31	0.09	0.28
Italy	0.11	0.31	0.07	0.25
France	0.10	0.30	0.10	0.30
Denmark	0.08	0.27	0.08	0.27
Greece	0.07	0.26	0.03	0.16
Switzerland	0.07	0.25	0.05	0.21
Belgium	0.13	0.34	0.17	0.38
0 child	0.13	0.34	0.14	0.34
1 child	0.17	0.37	0.17	0.38
2 child	0.40	0.49	0.36	0.48
3 child	0.18	0.39	0.20	0.40
3+ child	0.09	0.29	0.11	0.31
old-age eligible	0.06	0.23	0.05	0.22
early eligible	0.04	0.19	0.03	0.18
disability eligible	0.03	0.16	0.17	0.37
25% REC	0.01	0.09	0.02	0.14
50% REC	0.01	0.10	0.03	0.17
75% REC	0.01	0.09	0.05	0.22
100% REC	0.02	0.15	0.14	0.34
three Other	0.14	0.34	0.09	0.29
three Disability Work	0.03	0.17	0.13	0.34
three Disability Other	0.02	0.12	0.08	0.27
life in Work	16.11	14.07	16.50	13.77
life in Other	3.07	6.98	2.17	5.16
life in Disability Work	0.43	2.64	1.89	5.29
life in Disability Other	0.21	1.85	1.04	4.10
t (years in spell)	12.90	12.39	10.16	10.24

Table 17: Descriptive of main variables. The restricted sample contains only those individuals experienced at least one disability event in their whole life. It is used for simulating the effects of disability insurance policies which by definition are not affecting the healthy population.