Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Essays on the Self-Employed in the Netherlands and Europe

Elisabeth Beusch

PhD 12/2020-004

Essays on the Self-Employed in the Netherlands and Europe

ELISABETH BEUSCH

TILBURG UNIVERSITY

Essays on the Self-Employed in the Netherlands and Europe

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Portrettenzaal van de Universiteit op vrijdag 18 december 2020 om 13:30 uur door

Elisabeth Beusch

geboren op 11 november 1987 te Zürich, Zwitserland.

PROMOTIECOMMISSIE:

PROMOTORES:	Prof. dr. A.H.O. van Soest
	Prof. dr. T.E. Nijman
OVERIGE LEDEN:	Prof. dr. M.G. Knoef
	Prof. dr. R.J.M. Alessie
	Prof. dr. C. van Ewijk
	Dr. M. Mastrogiacomo

This thesis was funded by Instituut Gak through Netspar.

To Irene, the original.

Acknowledgements

Everyone who knows me will certainly agree that writing has not always been a forte of mine. Yet, the struggle behind some of the pages hereafter is paling compared to writing these Acknowledgments. After all, these pages will almost surely be the most read part of this thesis, and, more importantly, I doubt that I will ever be able to sufficiently express my gratitude to the many extraordinary individuals who have contributed to this thesis in their own way. The incredibly journey, both personally and academically, that led me to this thesis would not have been possible without the help, support and encouragement by all the people I met along the way.

First and foremost I was extremely lucky to have Arthur as my supervisor. I approached you during the second year of the Research Master because I had heard that you had a project in need of a PhD student and you gave me the chance to be part of it. I have learnt a lot from you not only academically but also personally. I very much appreciate the freedom and trust you gave me that allowed me to borrow from a different field for my work. You have devoted uncountable hours to me as my advisor and showed infinite patience. Thank you for our weekly discussions and your nudges whenever I lost focus or got carried away in details. I will try to be less critical of my future work – luckily I can for now mostly be critical of others'.

Further, I also want to express my gratitude to Theo, who was my second supervisor. Thank you for your guidance that put me back on track when I fell behind and for your valuable comments and discussions.

In addition, I want to thank my doctoral committee – Marike Knoef, Rob Alessie, Casper van Ewijk, and Mauro Mastrogiacomo – for taking the time and effort to read the first draft of this thesis. Thank you for all your constructive comments and insightful remarks and thoughts during the pre-defense.

My deepest gratitude also goes to Instituut Gak without whose funding my research would have never happened. My thanks also go to Netspar for enabling this.

I also want to thank many people in administrative roles. Anja and Anja thanking for all the work and help for all EOR Department related things. Ank, Bibi, Cecile and Corine and the rest of the Graduate School thank you for being there at any time when us students need you. Korine, thank you letting us always rely on you to solve all IT and financial issues. Future generations of PhDs won't know what they are missing.

My thanks also go to Bart, Jaap, and Tobi for organising the Structural Econometrics Group and giving all PhDs a platform and opportunity to present their work there. I have learnt a lot during these meetings either by presenting my own work, listening to my peers, or by observing how the three of you approach difficult problems.

During my years in Tilburg I met many people to who have enriched my life and deserve some words of gratitude. First, all my officemates – Mitzi, Laura, Yi, and Rafael – that I had throughout the years. I enjoyed our time together and learned a lot from you during our coffee breaks talking

about various topics as well as during our many discussions about research and teaching. Among the many others at Tilburg University, I also want to thank (in no particular order) Renata, Irina, Carlos, Lei, Jelmer, Christos, Ittai, Emanuel, Victor, Santiago, Oliver, Dorothee, Myrthe, Richard, Deriya, Ana, Suraj, Lieske, Ming, Zihao, Maciej, Nick, Nickolas, Chen, Stefan, Jan, Andreas, and Mika.

A special mention has to be given to the following persons. Masha, thank you for always being there when I needed you. Ana, my sweet sugar pie, our relationship will forever remain a special one. I can only reciprocate your words and add: #soblessed. Yeqiu and Yan thank you for always feeding me. Clemens, Manuel, Lenka, Peter, Shan, Thijs, Ricardo, Hanan, Marie, and Madina – I never thought I'd spend evenings slaying goblins, other monsters, and the occasional dragon but oh what fun it has been and still is. Bas, Mario, Sebastian, and Tomas – I don't know if this PhD took longer or less long because of the passion for cycling you have instilled in me. Thank you for the many tours and believing in me at the beginning. You will always be welcome in the Swiss mountains. Andrin, Caroline, Christina, Elodie, Janine, Jessica, Loretta, and Wischiro – I hope we can finally meet more now that I'm back. My longstanding friends Corinne, Daniela, Irene, and Ulla – there will never be enough words for you.

The biggest thanks of it all go to my family. Mami und Papi, without you all of this would not have happened. You instilled a curiosity for the world in us and let us choose our own path. Papi, no matter what comes, you will always have the upper hand when it comes to anything that requires practical solutions. Mami, thank you for always being my biggest cheerleader and keeping me well fed no matter the distance. Irene, as I can now stop with calling you Frau Doktor Beusch, thank you for literally always being there for me.

And last but not least: Nick, thank you for your endless patience. The past years with you have been wonderful and I'm looking forward to many more. I know I have not always been the easiest to live with in the last year so thank you for putting up at times with my cranky self. I hope that we can finally go on many long rides again. You will be happy to read this: *Ik ben klaar.*

Elisabeth Beusch
Zürich, September 2020

Table of Contents

List of Tables

Table **Page**

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

As apparent from its title – Essays on the Self-Employed in the Netherlands and Europe – this doctoral thesis' focus lies on the self-employed. To be more precise, the three essays presented here study the dynamics of self-employment in the labour market and the resulting career trajectories over time.

The motivation for this thesis lies in the recent increase in the number of self-employed in the Netherlands and the resulting concerns of policy makers about the impact of this increase on the social security system. One particular concern are the pensions, as most self-employed, unlike the majority of employees, are excluded from participating in the second pillar of mandatory occupational pensions. Instead they are expected to make sufficient provisions for their old age voluntarily (the third pillar). As mandatory pension savings and career decisions are therefore directly linked with each other, it is important to understand which individuals are more likely to remain in self-employment – the subject of the first essay. Moreover, it is also important to understand how career trajectories are related with income, savings, and financial well-being – the subject of the second and third essay.

Chapter 2, entitled *A Dynamic Multinomial Model of Self-Employment in the Netherlands* (co-authored with Arthur van Soest), focuses on transitions into and out of self-employment, wage employment and non-participation in the labour market. We estimate a dynamic multinomial logit model using Dutch micro-panel data, the LISS panel, covering a period of ten years. Our model incorporates unobserved heterogeneity through random effects that are correlated across labour market states. Adding these to the dynamic model allows us to differentiate between the effect of what Heckman (1981b) calls true state dependence and spurious state dependence. True state dependence refers to the causal effect of being in a labour market state today on being in e.g the same labour market state next period. Spurious state dependence refers to non-causal reasons why individuals observed to be more likely in a specific labour market state are also more

likely to be in an other state. The distinction is important because of the differences in policy implications.

In addition to demographic characteristics, our model also incorporates personality traits. These allow us to not just simulate employment paths for benchmark individuals with different demographic characteristics but to also have a more nuanced picture by looking at individuals who are better or less well suited for entrepreneurship. These simulation results in turn help us to illustrate the limitations of the common assumption in wealth and pension income modelling, that individuals remain in their observed labour state until retirement. We show that while the probability to not remain in self-employment on a year-on-year basis is less than 10%, the probability to remain more than five out of ten years, not necessarily consecutively, in self-employment for the least-suited to entrepreneurship male benchmark individual may be even less than 50%. The simulations also indicate that persistence in self-employment is even lower for women. Hence we argue that simulations that project future pension incomes and pension adequacy should account for labour market dynamics, as the paths that "static" projections assume, with constant self-employment for those observed as self-employed, are not necessarily representative for many individuals.

Chapter 3, entitled *Labour Market Trajectories of the Self-Employed in the Netherlands* (co-authored with Arthur van Soest), takes the concerns raised in Chapter 2 to the data and looks at realised labour market trajectories between 1989 and 2017 of Dutch individuals born between 1936 and 1980. Using Dutch administrative data, we analyse the trajectories of more than 50,000 individuals including 13,000 with some self-employment experience. We find that a large share of the individuals that are at least one year self-employed do not remain self-employed for a long time. Overall, a quarter of the individuals spends no more than three years in self-employment. Less than half of all individuals with some self-employment experience spend more than 10 out of the maximum of 29 years as self-employed, and only a third spends more than half of these years as self-employed.

Using the technique of sequence analysis developed in the social sciences, we cluster the trajectories based on the patterns of self-employment, wage employment and non-employment that individuals with some self-employment experience display. We find seven different clusters with distinct life-cycle trajectories. We then study the characteristics of these clusters in terms of individuals' income, wealth and pension investments. We find that incorporated individuals, the so-called DGAs, are the "positive" outliers among the self-employed – they are financially better off than, e.g., individuals who spend their whole career as wage employees. At the other end we find three clusters (self-employed with low labour market attachment, those that spend a large part of their trajectories as benefit recipients, and employees with short self-employment spells) whose members do not spend a long time in self-employment. These individuals are found to have large gaps in their occupational pension accumulation and are not more likely than employees to save in the third pillar. Individuals that switch from employment to self-employment

late in their career are not found to perform much "better". That is, they do not seem to take measures to compensate for the loss of further contributions to their pension accumulation in the second pillar. Last, we find that, despite of frequent worry about the lack of pension preparedness of the groups with self-employment experience, individuals that spend most of their career in self-employment have also accumulated significantly more non-pension wealth than employees. We therefore suggest that policies that target the pension incomes of self-employed should differentiate between short- and long-term self-employed.

Last, Chapter 4, entitled *Self-Employment Careers and Financial Well-Being in Old Age in Europe*, uses information from retrospective interviews in the Survey of Health, Ageing and Retirement in Europe on individuals' careers to examine whether certain trajectories are correlated with more financial difficulties for individuals born between 1931 and 1955 once they are aged 60 and older. I use again sequence analysis to group the individual career trajectories, and I find different clusters of short- and long-term self-employed. Because the sample consists of different European countries with a variety of institutions that might impact self-employed individuals differently, I divide the countries into groups based upon their percentile ranking based on the World Bank's Rule of Law indicator in 1996, a time when the individuals in the sample were mostly active in the labour market. This indicator captures perceptions regarding the extent to which individuals and firms have confidence in and abide by the rules of their society.

The biggest difference I find across the country groups once I study different measures for financial well-being is that self-employed women in the highest ranking countries seem to be mostly second income earners. They have a higher household income on average and the majority of clusters of individuals with careers involving some self-employment do not have a lower pension income than women that were mostly employees. This stands in contrast to women in the lowest ranking countries that seem to have been self-employed out of necessity, and who report lower incomes. Overall though, the result that prevails when I also study indicators of financial well-being, is that individuals who were always self-employed are on average worse off financially than their peers who were employees. This suggests that the results from Chapter 3, where we found that long-term self-employed were better off financially than other self-employed, might have to be taken cautiously. Furthermore, in particular if we also consider that the current self-employed are often not accumulating as much capital in their companies as older generations did, these individuals' risk for financial difficulties once they are older may be even higher than what I observe for the older generations in the sample. Economic policy measures should address these shortcomings. This holds particularly in the light that several European countries still exempt most self-employed from enrolling in earnings related pensions that are mandatory for employees.

CHAPTER 2

A Dynamic Multinomial Model of Self-employment in the Netherlands

This chapter is based on the identically entitled working paper which is co-authored with Arthur van Soest

This paper presents a dynamic multinomial logit model to explain the transitions into and out of self-employment using Dutch micro-panel data, the LISS panel. Based on the estimates we simulate employment paths for benchmark individuals. These are used to illustrate the limitations of the common assumption in wealth and pension income modeling, that individuals remain in their observed labour state until retirement. In particular, we find that although one year transition probabilities out of self-employment are not more than 10%, the chances that individuals who are self-employed remain self-employed for the majority of the next ten years can be much smaller, and vary substantially with individual characteristics such as education level and personality.

2.1 Introduction

In recent years the number of self-employed in the Netherlands has grown substantially, leading to an increase of almost 30% in their share in the working population: from 12.8% in 2003 to 16.6% in 2017.[1] The main driver behind this growth have been the so called solo self-employed (SSE; in Dutch "zzp'ers"= zelfstandigen zonder personeel). As can be seen in Figure 2.1, the share of SSE in the working population increased from 8.1% in 2003 to 12.3% in 2015 and has remained

[1] All numbers are based on CBS Statline, *Arbeidsdeelname; kerncijfers*, downloaded on 4 May 2018.

rather stable since then. The share of other self-employed has, on the other hand, seen a slight decline since the financial crisis.

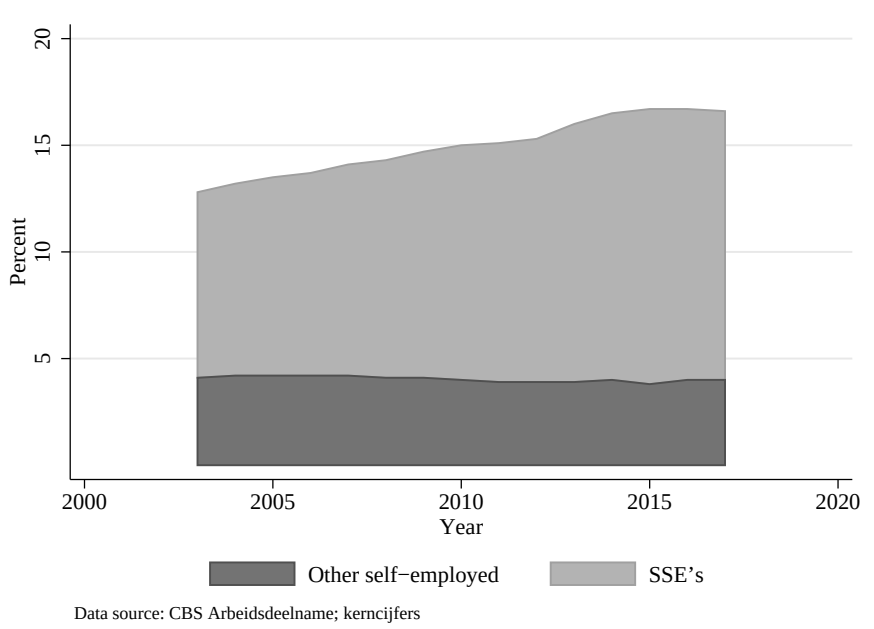


Figure 2.1: Cummulative share of self-employment in the working population

Because of the growing numbers, Dutch policy makers have become more interested in the effects the self-employed or SSE in particular may have on the labour market, social security, and government finances. Accordingly, several recent policy papers describe the trend in self-employment and the characteristics of the self-employed, or analyze their performance; see, e.g., Bosch et al. (2012), Bosch (2014) or Bolhaar et al. (2016). One key concern of the Dutch policy makers in relation to social security is adequacy of the pension savings of the self-employed; see, e.g., Mastrogiacomo and Alessie (2015) or Knoef et al. (2016). While the pay-as-you-go pension (the so called AOW, the first pillar of the Dutch pension-system) covers all individuals who have lived in the Netherlandsbetween ages 15 and 65, contributions to a fully funded pension plan (the second pillar) are, unlike for the large majority of employees, neither mandatory nor accessible for most of the self-employed.[2] Instead, the self-employed are expected to save themselves through (voluntary) savings (the third pillar). Such pension savings are tax-favoured for everyone with an "incomplete" second-pillar pension, in order to stimulate that individuals indeed save enough for

[2]Most second pillar pensions are built up via employer based pension plans or industry specific pension funds. In total about 90% of all employees are required to contribute to a pension plan (https://www.rijksoverheid.nl/onderwerpen/pensioen/opbouw-pensioenstelsel). Self-employed individuals in industries with industry specific pension funds (e.g., painters and doctors) are required to participate in a second pillar pension. The majority of the self-employed is not in such a sector.

their pension.[3] This then raises the question whether the self-employed save enough in the third pillar.

It turns out that the policy makers' concerns have some basis. Mastrogiacomo (2016) shows that while the self-employed have the same savings ambitions as the employed, they are not more likely to save in the third pillar. Only one third of the self-employed contributes to the third pillar, which indicates that the majority will fall short on their savings. In line with this finding, earlier studies by de Bresser and Knoef (2015), and Knoef et al. (2016) found that the self-employed are less likely to meet their retirement expenditure or saving goals. Zwinkels et al. (2017) focus on the solo self-employed and estimate that more than 40% of SSE households fall short on their savings if a target replacement rate of 70% of earnings is used.

One simplifying assumption made in the pension wealth projections by Zwinkels et al. (2017), but also by de Bresser and Knoef (2015) and Knoef et al. (2016), is that the observed individuals remain in the labour state in which they were at the point in time when the data were collected. To our knowledge this assumption ("static micro-simulation") is standard in the pension literature and its consequences have not been discussed so far. Still, given that the savings in the second (and third) pillar – a large share of most individuals' pension wealth – are linked directly to the individuals' labour state, it may be worthwhile to study the validity and consequences of this assumption. This paper therefore studies the dynamics in the Dutch labour market, considering self-employment as one of the labour market states. For instance, it asks how likely it is that somebody who is observed in self-employment will remain self-employed, depending on the individual's characteristics.

We will use data from the LISS (Longitudinal Internet Studies for the Social sciences) panel, a representative sample of adult individuals in the Netherlands administered by CentERdata (affiliated with Tilburg University). It is based upon a random sample of Dutch households drawn by Statistics Netherlands. Individuals of age 16 and older in the participating households are invited to answer survey questions on a monthly basis. The surveys cover domains such as work, education and income, but also a wide range of other topics, like health and personality, thus offering a rich set of information on which we can build our analysis. It also allows to distinguish between employees, SSE and other self-employed. Because sample size limitations, the main analysis is done without a distinction between different self-employment types, even though such a distinction might be desirable given the specific interest in the SSE in the Dutch policy debate.[4]

In addition to a set of personal and household characteristics also included in most of the studies cited above, we control for personality traits and a (lagged) health index. Recent work on the economic importance of personality traits (see e.g. Borghans et al., 2008) has shown that personality traits matter for different labour market outcomes and, particularly, the decision to become self-employed (see e.g. Beugelsdijk and Noorderhaven, 2005) This has also been found

[3]Recently, a specific pension fund for the self-employed has been opened but it should be considered as a third pillar annuity.

[4]As illustrated in Figure 2.1 most of the recent dynamics in self-employment seem to be driven by the SSE.

in sociological research on career counseling. For instance, Obschonka et al. (2013) construct an Entrepreneurship-Prone Big-Five Profile (EP) Distance measure and find that in the US, the EP distance's geographical distribution corresponds to observed entrepreneurial activity. We therefore also include the EP distance in our analysis. Good health has been identified as a factor that increases the probability to become self-employed for older workers in the US (Rietveld et al., 2015). We use the rich nature of the LISS data to construct a health index and study the role of health for self-employment transitions in the Netherlands.

We first model self-employment in a static multinomial choice panel data framework with unobserved heterogeneity. We then extend our model to include dynamics to demonstrate the importance of state dependence. We not only consider self-employment and wage employment, but also account for transitions into and out of paid work. Our dynamic multinomial logit model is similar to that of e.g. Gong et al. (2004), who model the choice between not working, informal work, and formal sector work in Mexico, or Buddelmeyer and Wooden (2011) who model dynamics between casual and other types of employment in Australia. Oguzoglu (2016) follows Gong et al. (2004) to model the influence of disability on employment decisions, and Zucchelli et al. (2012) consider self-employment as an alternative to part-time employment for the elderly under possible ill-health. Another case in point is Prowse (2012) who includes self-employment when modelling the labour participation of women. Finally, Been and Knoef (2017) also use a dynamic multinomial logit model to explain self-employment decisions in the Netherlands, focusing on workers of ages 50 and above and using administrative data. We consider all individuals of working age and use survey data, which has the advantage of providing rich background information such as personality or health indicators, as already emphasized above.

Our models incorporate unobserved heterogeneity, allowing for correlated random effects following Train (2009). Adding this to the dynamic model allows us to differentiate between what Heckman (1981b) calls spurious and true state dependence, which is important to understand the dynamics in the data. We solve the problem of initial conditions that arises in dynamic models following Wooldridge (2005) and Albarrán et al. (2019).

The paper continues as follows. Section 4.2 discusses the LISS panel and our sample selection process. Our model is presented in section 2.3 and the corresponding estimation results in section 2.4. Section 2.5 presents the simulation results based on the estimations. Section 4.5 concludes the paper.

2.2 Data

In this paper we make use of the LISS (Longitudinal Internet Studies for the Social sciences) panel. The LISS panel consists of monthly Internet surveys to a representative sample of households drawn from the Dutch population register.[5] Among the monthly surveys there are ten annual

[5]Households that do not own a computer or Internet connection are provided with such so that they can nevertheless participate.

or biennial longitudinal core studies. Additionally, individuals are asked to fill in a basic survey, the household box, about the most important general characteristics of their household and its members such as age, gender, education, marital status, as well as their primary occupation and gross income. Individuals, or the contact person of the household — if there is more than one member of the household participating in the LISS panel — are asked to fill in the household box at the beginning when joining the panel and then prompted every month before each survey to fill in changes if such have occurred.

2.2.1 Self-employment

Among all longitudinal surveys, there are three instances within the LISS panel through which we can identify self-employed individuals. First, information about an individual's labour market status is stored in the household box. Instead, self-employed individuals can be identified using either the *Work and Schooling* or the *Economic Situation: Income* core study. We will base our analysis on the income study for two main reasons. First, the income study allows us to identify solo-self-employed (SSE) while the work and schooling study does not allow for a distinction between SSE and other self-employed. Second, the income study based sample suffers less from selection or attrition bias than the work and schooling study.

The income survey has a different timing from other LISS studies. Individuals are supposed to use documentation on income taxes in the previous year to fill out the survey questions. Therefore, the survey asks individuals in period t about sources of income in the calendar year $t-1$. For example, the 2008 survey asks about all income received in 2007. We classify individuals as employees who report receiving only income from employment over the whole year. Individuals with income from both employment and self-employment are classified as self-employed, together with those individuals who report only income from self-employment.[6] An individual is classified as self-employed if indicating at least one type of entrepreneurial work activity. The activities that the income survey covers are (part-time) work as an entrepreneur or freelancer, SSE, owning a company (including a private limited liability company or a limited partnership), or participating in a partnership (either a so called *maatschap* or *vennootschap onder firma, VOF*) and, lastly, making a profit (or loss) through an enterprise in some way (except as spouse or partner cooperating in the business). Next, we classify all individuals as unemployed who report receiving unemployment benefits and no other source of income, ignoring other social benefits. Because this will only classify individuals who are unemployed for a whole calendar year as unemployed, the unemployment definition is rather strict, covering a smaller number of individuals than the work and schooling based definition which refers to one point in time. Finally, individuals with no income from any of these sources are classified as not in the labour force. A comparison of this classification with the work and schooling based classification is given in Appendix A.

[6] About 40% of all self-employed have income from both employment and self-employment.

2.2.2 Sample selection

The surveys in the LISS panel generally have a response rate between 75 and 80%. Thus, every year we have answers of around 5000 – 6500 individuals, of which a few are incomplete. Individuals who leave the panel (i.e. stop answering the surveys all together) are replaced in later waves through refreshment samples. Most of these are stratified to improve the representativeness of the panel, aiming at oversampling difficult to reach groups with below-average response rates. For the surveys that we use, there are in total more than 14,000 individuals across the 11 waves and some 81,000 observations. We have information on the labour status for some 60,000 observations if we use the income based classification.

We restrict the sample to individuals from age 25 up to and including age 60, i.e. individuals' prime working years. We choose the lower bound at 25 because, first, the minimum wage increases with the worker's age until 23 in some sectors. As a result of this it seems that young workers in these sectors may have a higher risk of becoming unemployed close to their birthdays (Kabátek, 2020). Second, students who are finishing their education are harder to classify. They may hold a (side) job, while studying, and can also be considered first time job seekers. By age 25 most individuals should no longer be students. The age limit at 60 years stems from the idea that individuals older than 60 may have access to (early) retirement.The age restriction reduces the sample size to approximately 32,000 individuals.

Furthermore, we limit ourselves to individuals for whom the basic covariates, such as age, gender, household status, and education, are observed (only very few observations are dropped due to this restriction). The final sample restriction that we have to make is model based. In the dynamic models, we want to model labour market state outcomes based on individuals' past labour market state. Hence we can only use individuals for whom we have at least two consecutive observations. Moreover, we have to discard observations made after an individual has not responded for one or more years. For example, if an individual answers the income survey in the years 2008-2012 and again from 2014-2018, we do not include the 2014-2018 block. These restrictions make us lose approximately 15% of the observations.

As shown in Appendix 2.A the final restriction potentially creates (or worsens) attrition bias in the sample. Because a large share of the dropped observations belongs to individuals that participate in more than one wave, we correct for breaks in sequences with information from the work survey. The details are described in Appendix 2.A.2.

Figure 2.2 compares the self-employment share in the income sample with CBS population data already seen in Figure 2.1. It shows that the LISS panel replicates neither the magnitude nor the time trend of the share of self-employment. This difference can be due to selection bias, i.e. self-employed individuals are less likely to respond when invited for the LISS panel, or attrition bias, i.e. self-employed individuals are more likely to stop responding in a later wave. Most likely, the problem is a mixture of both. Examining the evolution of self-employment shares over time by recruitment wave shows that they all display similar patterns. They start with relatively high

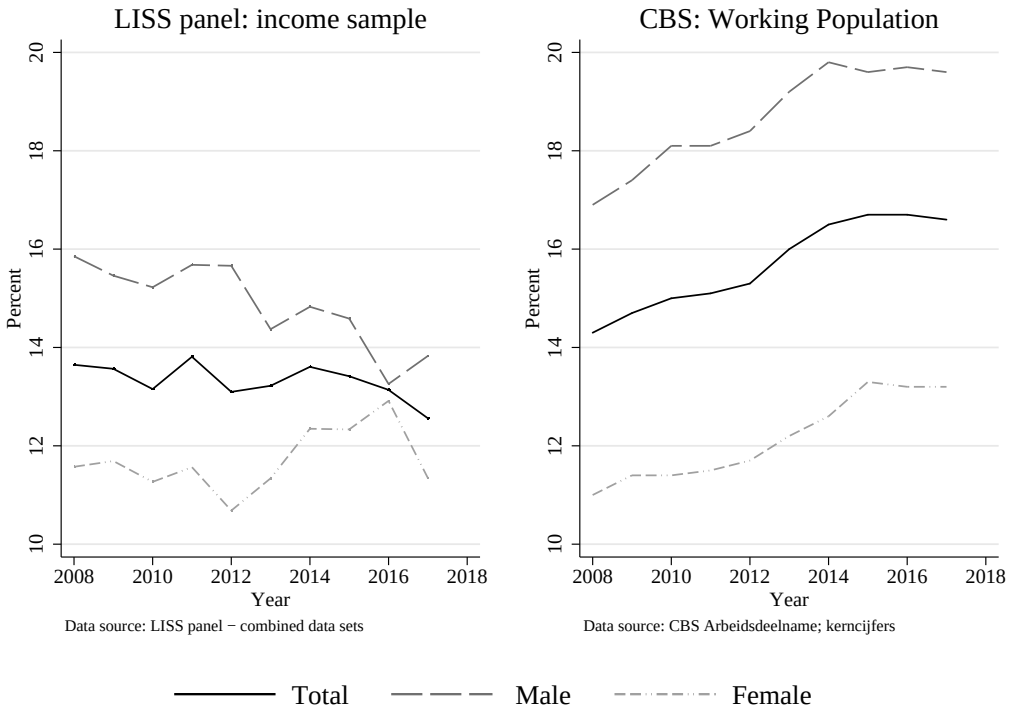


Figure 2.2: Comparison of self-employment shares in regression sample and population

shares at recruitment, followed by a drop, and ending with a flat or slightly downward trend. The fact that shares are initially closer to population figures suggests that the trend is more affected by attrition bias than by initial selection.

Both selection and attrition bias are of less concern if they are driven by observables like age, gender, or eduction. In such a case one can correct for the bias by weighing the observations accordingly. We therefore tried weighing observations using weights based upon the observable characteristics that enter our model.[7] Weighing only leads to an increase of one percentage point in self-employment shares overall and does not change the trend. Furthermore, Wooldridge (2007, p. 1293) cautions about using weighting in panels. We therefore decided not to use these weights. To correct for selection on unobservables, we would have to impose a structure on the selection process. As this would require strong assumptions, and because the self-employment shares are initially not that different from population shares, we refrain from correcting for selection bias and focus on attrition bias only. Correcting for attrition bias requires weaker assumptions, especially if we can use an exclusion restriction. We test and correct for attrition bias in section 2.3.2.

[7]The weights are determined using population data downloaded from CBS Statline.

2.2.3 Explanatory variables

Because personality traits are less commonly used, we discuss them in some detail. The *Personality* core study of the LISS panel focusses on respondents' "personality and characteristics". Its questions are based on established questionnaires from the field of psychology that each have a different focus. One of these questionnaires is the short 50 question set for the Big-Five factor markers by Goldberg (1992). Individuals answer these questions on a 1 to 5 scale. We code their answers according to the corresponding International Personality Item Pool key.[8] For each of the five factors we then sum up the points on an individual basis and standardise these values with the mean and standard deviation of the complete LISS sample for each year, allowing us to interpret coefficients of the factors in terms of changes relative to the standard deviation. The Entrepreneurship-Prone Big-Five Personality Profile Distance (EP distance) measure is calculated using the non-standardised factor values following Obschonka et al. (2013).

To reduce the number of questions asked to individuals, the LISS panel only poses the Big-Five questions every second year. In the other years the Big-Five related questions are only asked to new entrants. Furthermore, the personality survey was not asked to participants in 2016. We find that the personality traits in our sample remain rather stable across time, with a between variation that is two to three times larger than the within variation for all factor markers. This is in line with Cobb-Clark and Schurer (2012) who found that personality traits are stable over time. We therefore fill in the gaps in Big-Five factor markers and EP distance by computing individual means over all observations available and substituting missing values in gap years with those means.

Since health variables are more standard in the literature, we refer to Appendix 2.D for the construction of the health index. In addition, we include the individual characteristics age, gender, and dummy variables for medium level education (VMBO, VWO, or MBO diploma) and higher education (university (WO) and applied science university (HBO) degrees), which have been shown to have some correlation with the choice to be self-employed.[9] In addition, we use household specific variables: dummy variables controlling for whether an individual lives with a partner and/or has children, as well as the size of the household. These variables also have been found to have explanatory power in regressions explaining the decision to be self-employed; see, e.g., the overview of research on entrepreneurship by Blanchflower (2000).

Table 2.1 reports the means by labour market status for all covariates. Overall we see that women are slightly over-represented in the sample as they make up 55% of all observations. Approximately half of the individuals in the sample have at least one child and the majority lives with a partner. Unsurprisingly, we find that those not in the labour force are mostly women, and that the majority of them lives with a partner. The two education dummy variables account for 96% of the sample, implying that only 4% of the sample has the lowest education level.

[8] https://ipip.ori.org/newBigFive5broadKey.htm retrieved on July 6, 2018.

[9] See Bosch et al. (2012), Bosch (2014), CBS (2014) and Bolhaar et al. (2016).

Table 2.1: Means and standard deviations (in brackets) of covariates

	Employee		Self-employed		Unemployed		Not in LF		All	
Age	44.08	(9.71)	45.93	(9.25)	47.52	(9.71)	48.31	(9.67)	45.07	(9.79)
Female	0.52	(0.50)	0.45	(0.50)	0.62	(0.48)	0.73	(0.44)	0.55	(0.50)
Lives with partner	0.76	(0.43)	0.78	(0.41)	0.54	(0.50)	0.74	(0.44)	0.75	(0.43)
Has children	0.55	(0.50)	0.57	(0.50)	0.44	(0.50)	0.46	(0.50)	0.53	(0.50)
Female with partner	0.38	(0.49)	0.35	(0.48)	0.33	(0.47)	0.58	(0.49)	0.41	(0.49)
Female with children	0.29	(0.45)	0.27	(0.44)	0.31	(0.46)	0.36	(0.48)	0.30	(0.46)
Medium education	0.58	(0.49)	0.52	(0.50)	0.72	(0.45)	0.72	(0.45)	0.60	(0.49)
Higher education	0.40	(0.49)	0.45	(0.50)	0.21	(0.41)	0.19	(0.39)	0.36	(0.48)
Household size	2.83	(1.33)	3.03	(1.43)	2.46	(1.37)	2.71	(1.36)	2.82	(1.35)
Health index	-0.12	(1.03)	-0.13	(1.04)	-1.18	(1.57)	-1.03	(1.53)	-0.30	(1.21)
F1: extraversion	0.00	(0.98)	0.15	(1.01)	-0.16	(0.94)	-0.18	(0.97)	-0.01	(0.99)
F2: agreeableness	-0.02	(0.95)	-0.09	(1.02)	0.05	(0.99)	0.09	(1.00)	-0.01	(0.97)
F3: conscientiousness	0.12	(0.90)	0.01	(0.97)	-0.03	(0.99)	-0.01	(0.96)	0.08	(0.92)
F4: emotional stability	0.08	(0.95)	0.06	(0.94)	-0.39	(1.05)	-0.34	(1.02)	-0.01	(0.97)
F5: openness	0.08	(0.93)	0.31	(1.01)	0.03	(0.98)	-0.11	(0.96)	0.07	(0.95)
EP distance	18.71	(5.12)	18.14	(5.11)	20.74	(5.95)	20.98	(5.78)	19.08	(5.34)

Source: LISS panel, own calcuations.

Table 2.2: Observed transition probabilities (in %) by gender

Labour market state	Men				Women			
past \ current	0	1	2	3	0	1	2	3
0: employee	93.96	1.65	1.04	3.36	92.23	1.47	1.28	5.02
1: self-employed	8.15	89.35	0.51	1.99	9.34	84.73	0.71	5.22
2: unemployed	24.22	2.34	45.70	27.73	19.05	1.90	42.38	36.67
3: not in labour force	24.26	2.09	7.13	66.52	14.28	1.82	6.06	77.85
LISS total	74.67	13.08	2.51	9.75	66.01	8.82	3.46	21.72
CBS population	71.09	16.88	3.68	8.36	65.84	9.69	4.07	20.39

Based on 12014 and 14496 observation pairs respectively. Probabilities are unconditional on other individual characteristics.

Source: LISS Panel and CBS Arbeidsdeelname; kerncijfers, own calculations.

Furthermore, we can see that the distribution of the two dummy variables varies between the working and non-working population — the higher educated are much more likely to do paid work. The distribution of the health index is left skewed with the mode at 0.76, and we can see directly that individuals in the working population have a higher health status than those not working. There are also differences in Big-Five factor markers between the working and non-working individuals.

Comparing across the labour status groups, we find that the means of all variables except for the health index and the fourth personality factor marker are statistically different between employees and the self-employed. This difference is in line with the literature: men are more likely to be self-employed than women, the self-employed are on average older (compared to the employed), and the self-employed more frequently have a higher level of education. They are also more likely to have children. The argument for the EP distance by Obschonka et al. (2013) predicts that entrepreneur prone individuals are more extraverted, less agreeable, and more open to new experiences. We find all of this reflected in the differences of the means of the three factor markers. However, the theory also argues that entrepreneurs should be more conscientious, and we find the opposite. Still, the EP distance measure has a lower average value for the self-employed than for employees, as theory would predict.

2.2.3.1 Observed labour market state dynamics

In Table 2.2 the observed transition probabilities of the labour market states by gender are shown.[10] We see that employees are more likely to remain in the same state than the self-employed, whose probabilities to exit to working as an employee are 8 to 10% per year. This already suggests that the assumption that self-employment is persistent has its limitations. The

[10]The transition matrix is, of course, affected by the imputations for the gaps in the data. The changes are relatively small, however. See Table 2.E.1 in Appendix 2.E.

majority of the self-employed who do not continue as such switch to employment, and since the number of employees is much larger than the numbers in other labor market states, the largest contribution in numbers to entrants into self-employment are individuals making the transition from employment.

Comparing the transition matrices of men and women, we see that women are less likely to remain in self-employment than men are, and much more likely to remain out of labour force. Because of the substantial differences between men and women, we estimate all models separately by gender. Finally, we observe, in percent as well as in absolute numbers, very few changes from unemployment to self-employment and vice versa. This may be due to our categorisation approach in the income based definition of self-employment, or to sample selection. Finally, comparing with CBS population figures shows that self-employed are under represented, as already discussed.[11]

2.3 Model

This section presents the empirical model and the estimation procedure. Both are similar to the econometric specifications used by Gong et al. (2004) and Been and Knoef (2017). Note that the static multinomial model is nested in the dynamic model. We will therefore focus the discussion on the dynamic model, treating the static model as a special case. In the final subsection, we address the issue of attrition bias.

2.3.1 Dynamic multinomial model of labour states

We model the observed labour market state of an individual as the outcome of a utility maximisation process. Each individual re-evaluates the potential states every period, and chooses the labour state j that maximises utility for that period. In terms of the econometric specification we thus consider a discrete choice model where an individual i derives utility y^*_{ijt} from state j at time t. In other words:

$$y_{ijt} = \begin{cases} 1 & \text{if } y^*_{ijt} > y^*_{ikt} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } j,k = 0,1,2,3; j \neq k; i = 1,\ldots,N; t = 2,\ldots,T \tag{2.1}$$

where $y_{it} = (y_{i0t} \ldots y_{i3t})$ is a column vector with a 1 in the position that corresponds to individual i's labour market state at time t and zeros everywhere else. Due to data limitations, we do not include higher order lags.

Utility y^*_{ijt} from choosing state j is unobserved. It is assumed to be given by

$$y^*_{ijt} = X_{it}\beta_j + y^{\mathsf{T}}_{i,t-1}\gamma_j + \alpha_{ij} + \epsilon_{ijt} \tag{2.2}$$

Here $y_{i,t-1}$ is a vector of dummy variables describing the individual's labour state in the previous period, X_{it} is a vector of k observed strictly exogenous variables, and the coefficient

[11] The CBS shares presented in Table 2.2 are calculated for the 25 to 60 year olds, taking the average for 2008–2017.

vectors γ_j and $\beta_j, j = 1,2,3$, are to be estimated; γ_0 and β_0 are normalized to zero. The variables in X_{it} are individual as well as household characteristics that may influence the utility y^*_{ijt}. These variables have been discussed in section 2.2.3. X_{it} also includes time dummies, to control for macro-economic effects.

The error terms ϵ_{ijt} are identically and independently distributed, independent of X_{it} and α_{ij} and drawn from a Type 1 extreme value distribution. This implies that the labor market state probabilities, given X_{it}, α_{ij} and $y_{i,t-1}$, are the well-known multinomial logit probabilities (see Appendix 2.C for details). Because equation (2.2) includes lagged dependent variables, an initial conditions problem arises (Heckman, 1981a). For most individuals, we have no information on how and when they entered the labor market and the first observation $y_{i,0}$ is typically some time after labor market entry. To account for the fact that $y_{i,0}$ may well be correlated with the time persistent individual effects $\alpha_{ij}, j = 0.\ldots,3$, we model $(\alpha_{i0},\ldots,\alpha_{i3})$ as follows, following Wooldridge (2005):

$$\alpha_{ij} = y_{i0}^{\mathsf{T}}\delta_j + \mu_{ij} \tag{2.3}$$

Here $\mu_i = (\mu_{i0},\ldots,\mu_{i3})$ is independent of y_{i0} and all X_{it} and ϵ_{ijt}. Like Gong et al. (2004) and Been and Knoef (2017), we assume that μ_i is drawn from a J-dimensional multivariate normal distribution with mean zero and covariance W (see Appendix 2.C for details on how this is implemented). The δ_j are 4-dimensional parameter vectors to be estimated. This essentially boils down to including the vector of dummies y_{i0} as additional regressors when estimating the model.[12] Note that due to the presence of the unobserved heterogeneity terms, the independence of irrelevant alternatives (IIA) assumption is not imposed. This IIA assumption is often seen as a drawback of the standard multinomial logit model. The estimates of the covariances will give an indication whether individuals who prefer one labour state are also more likely to prefer any particular other labour state. For example, if the covariance for states 1 and 2 (self-employed or unemployed) is positive, we should expect an individual, ceteris paribus, to have a higher probability of choosing self-employment when he or she has a high individual parameter for unemployment.

Static multinomial model of labour states

As already mentioned, the static model is a special case of the dynamic model. That is, in the static model we exclude the past period's labour state and as a consequence also drop the initial conditions. Equations (2.2) and (2.3) can then be rewritten as

$$y^*_{ijt} = X_{it}\beta^s_j + \alpha^s_i + \epsilon^s_{ijt} \qquad \text{with} \qquad \alpha^s_i = \mu^s_i \tag{2.4}$$

[12] Here $t = 0$ is the first time individual i is observed, which varies with i due to the refreshment samples in the LISS panel. See Albarrán et al. (2019) for an extensive discussion of the plausibility of this way of modeling the initial condition for an unbalanced panel.

where superscript s indicates that the coefficients are for the static regression. Detailed assumptions and likelihood contributions are similar to those for the dynamic model.

Estimation

The probabilities implied by the static and dynamic model discussed above have to be simulated. Following Bhat (2001) and Train (2009, chapter 9.3.3) we use Halton draws to simulate the multivariate normal distribution of μ_i. See Appendix 2.C.3 for details and reasoning.

2.3.2 Attrition Bias

As noted by Verbeek and Nijman (1992, p. 681), it is well known since Heckman (1976, 1979) that "inferences based on either the balanced sub-panel or the unbalanced panel without correcting for selectivity bias, may be subject to bias if the nonresponse is endogenously determined". We therefore want to analyze whether self-employed individuals are more likely to leave the LISS panel, and thus contribute to the unbalanced nature of the panel, and, if they do so, whether this leads to biased estimates for the model above or not. In order to test for attrition bias we use a variation of the variable addition test of Verbeek and Nijman (1992). They consider three possible variables that can be included in the regression: the number of waves an individual participates in the panel, an indicator whether the individual participated in all waves, and an indicator whether an individual was observed in the previous period. Because we also want to make use of the refreshment samples, the first and third are not applicable, and the second has a different interpretation. Instead, we construct a variable that measures the ratio of the number of periods in which an individual participated and the maximum number of periods they could have participated. This is still a function of the response indicator and thus follows the idea of the variable addition test. If attrition was independent of the unobservables in the model, this additional variable should not enter the model significantly under the null hypothesis of no attrition bias.

In our benchmark specification of the model, we find that this variable enters the multinomial model significantly for men, providing evidence that the model for men suffers from attrition bias, but not for women (p-values are 0.1% for men and 0.4132 for women). We therefore estimate an extension of the model adding a Heckman correction term (estimated in a first stage). Formally the attrition model extends equations (2.1) and (2.2) as follows:

$$A_{it} = \begin{cases} 1 & \text{if } A^*_{it} > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 1,\ldots,N; t = 2,\ldots,T \tag{2.5}$$

$$A^*_{it} = X_{it-1}\beta^H + y^{\intercal}_{it-1}\gamma^H + z^{\intercal}_{it-1}\theta + \nu_i + \psi_{1,it} \qquad i = 1,\ldots,N\ ;\ t = 2,\ldots,T \tag{2.6}$$

where A^*_{it} is latent and A_{it} indicates whether an individual is observed in the LISS panel at time t or not.[13] Together, equations (2.5) and (2.6) model the attrition process in the first stage of estimation. X_{it-1} contains the same vector of regressors as X_{it} in equation (2.2) but one period lagged. y_{it-1} is a vector of dummies indicating the labour state of the individual one period past. Lastly, z_{it-1} is a vector of variables entering the attrition equation but not the other equations in the model (the "exclusion restrictions"). For the exclusion restriction we follow Cheng and Trivedi (2015) and use the number of days that individuals took to answer the last wave of the *Economic Situation: Income* core study after they received the invitation. We further include a dummy that controls for whether the individual answered within the deadline of the first call to participate in the survey, or only after the reminder, as well as an interaction term of the dummy with the days. The random term v_i is assumed to be a time invariant random effect while $\psi_{1,it}$ is assumed to be iid standard normally distributed.

Stage one is estimated separately as a panel probit model with random effects. The second stage model is then given by equations (1) and (2), but with the inverse Mills ratio $\phi(\cdot)/\Phi(\cdot)$, estimated using the panel probit model, as an additional regressor in equation (2).[14]

2.4 Estimation results

We estimated four different models: the baseline model with the basic personal and household characteristics, and models in which we add the (lagged) health index, the EP distance measure, and the Big-Five factor markers. We first present the first stage results used to construct the Heckman correction term. In our discussion of the results, we will focus on the partial implied effects, keeping all other observed and unobserved characteristics constant, and averaging over the complete sample. We use separate models for men and women, since some of the the estimated coefficients differ substantially between men and women.[15]

2.4.1 First stage — Heckman correction

In the first stage regression the sample also includes all individuals in the selected age range for whom we only have only one observation.[16] Detailed results for women and men are presented in Tables 2.E.4 and 2.E.5 in Appendix 2.E.

For women, we find rather weak effects of the labour market state indicators for self-employment and not participating in the labour force on the chances to remain in the sample.

[13] We exclude individuals from this step if they leave our sample because they turned 61 years of age. The first stage only aims at correcting for an individual's own choice to participate in the survey or not; it does not correct for the sample selection choices we made (see Section 4.2).

[14] Since the error terms in equation do not follow a normal distribution but a Type 1 extreme value distribution, our specification differs from the original Heckman correction model. Consequently the coefficients on the inverse Mills ratio cannot be interpreted as the covariances.

[15] The pooled regression results are available in Appendix 2.E

[16] This does not include new entrants to the income survey from the last year of data collection as we do not know yet whether they will return or leave in the next wave.

These effects are larger and statistically more significant for men. We find the same signs for both genders though: Compared to employees, self-employed individuals are less likely to be observed in the following period, and individuals not in the labour force are more likely to be observed again. In particular for men, and in combination with the findings of the variable addition test, this confirms the need to correct for attrition bias: attrition is at least part of the explanation why the self-employed are underrepresented in later waves of the panel.

The first stage estimates for the basic personal and household characteristics are robust across different model specifications. Only age has a statistically significant positive coefficient, implying that older individuals are more likely to continue participating in the LISS panel. For women, we do not find a statistically significant effect for any of the other basic covariates. For men, several educational dummies are statistically significant, showing that men are more likely to continue participating in the LISS panel the higher their education.

The variables that are excluded from the main model ("the exclusion restrictions") have the expected signs in the first stage. The more days individuals take to answer after the invitation to participate in the survey has been issued, the less likely they are to return in the following year. For women, this effect is particularly strong for those who answer after the first invitation for the survey, as shown by the significant interaction terms.[17]

We do not find significant effects of the respondents' contemporary health status on the probability to be observed in the next period. The EP distance measure is not significant for women either, but it is significantly negative at the 5% level for men. Recall that the EP distance's interpretation is that the lower its value, the more likely an individual is self-employed. The results show that conditional on employment or self-employment status, less entrepreneurial individuals are less likely to stay in the sample, perhaps because they are more pressed with time.

More conscientious individuals are more likely to continue participating in the LISS panel, as expected. For men, we there are no other individually significant effects, but the Big-Five factor markers are jointly significant (at a 2% level).For women, emotional stability and openness for experience have marginally significant negative effects on the probability to stay in the sample, which is not what we would have expected.[18]

2.4.2 Second stage — static models

To compare the static and dynamic models, we restrict ourselves to the sample of the dynamic model and estimate the static model excluding those individuals for whom we only have a single observation.

[17] For each of the surveys the LISS panel collects data in two calendar months. A reminder is sent to all those panel members that did not complete the questionnaire during the first month.

[18] The Big-Five factor markers are also jointly statistically significant with a p-value of 0.000 in the regression for women.

Table 2.3: Static model with Big-Five factor markers, women

	Self-employed	Unemployed	Not in Labour Force
Constant	-17.84***	-5.96***	-4.65***
	(1.2323)	(1.0431)	(0.8464)
Age	0.13***	0.07***	0.1***
	(0.0169)	(0.0135)	(0.012)
Has partner	0.45	-0.53**	0.75***
	(0.305)	(0.2302)	(0.2087)
Has child	-0.64	-0.46*	-0.73***
	(0.3988)	(0.277)	(0.2374)
Middle education	-0.24	-1.28***	-1.51***
	(0.6524)	(0.4414)	(0.3927)
High education	0.39	-2.83***	-2.92***
	(0.6965)	(0.4805)	(0.4226)
Household size	0.47***	0.26**	0.34***
	(0.1511)	(0.1203)	(0.0958)
F1: extraversion	0.13	-0.09	-0.07
	(0.1555)	(0.1128)	(0.0955)
F2: agreeableness	-0.05	0.18	0.19*
	(0.1572)	(0.1118)	(0.0972)
F3: conscientiousness	0.09	-0.58***	-0.65***
	(0.1581)	(0.121)	(0.0998)
F4: emotional stability	-0.24*	-0.47***	-0.32***
	(0.1417)	(0.1059)	(0.0864)
F5: openness for experience	0.36**	0.36***	0.17*
	(0.145)	(0.1097)	(0.0987)
Inverse Mills Ratio	10.49***	-6.63***	-8.06***
	(1.2557)	(1.3516)	(1.0045)
L	6.99***		
	(0.3628)		
	2.48***	3.46***	
	(0.2235)	(0.2033)	
	3.23***	3.65***	1.55***
	(0.2143)	(0.1653)	(0.1167)
Covariance $W = LL^{\intercal}$	48.8702	17.3339	22.5515
	17.3339	18.1315	20.6267
	22.5515	20.6267	26.1073
Observations: 14435			
Nr. of Individuals: 3267			
Loglikelihood: -7704.03			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.4: Static model with Big-Five factor markers, men

	Self-employed	Unemployed	Not in Labour Force
Constant	-22.07***	-10.28***	-2.5***
	(1.4174)	(1.5229)	(0.9608)
Age	0.2***	0.13***	0.07***
	(0.0181)	(0.019)	(0.0127)
Has partner	0.08	-0.77**	-0.76***
	(0.3333)	(0.3629)	(0.2579)
Has child	-1.65***	-0.38	-0.73**
	(0.348)	(0.4375)	(0.309)
Middle education	0.67	-1.46***	-2.45***
	(0.7667)	(0.5278)	(0.3789)
High education	1.6**	-2.6***	-3.83***
	(0.7987)	(0.5816)	(0.4332)
Household size	0.21	-0.03	0.16
	(0.1608)	(0.1919)	(0.1335)
F1: extraversion	0.39**	-0.07	-0.06
	(0.1608)	(0.1678)	(0.1028)
F2: agreeableness	-0.32**	0.13	0.05
	(0.1311)	(0.1309)	(0.0938)
F3: conscientiousness	0.02	-0.3**	-0.46***
	(0.1423)	(0.1407)	(0.1037)
F4: emotional stability	-0.17	-0.6***	-0.67***
	(0.1417)	(0.1422)	(0.0988)
F5: openness for experience	0.73***	0.34**	0.23**
	(0.1359)	(0.1464)	(0.1009)
Inverse Mills Ratio	21.18***	-0.23	-7.86***
	(1.4209)	(2.2355)	(1.2027)
L	6.37***		
	(0.3519)		
	2.23***	3.15***	
	(0.273)	(0.1969)	
	2.33***	2.44***	1.53***
	(0.2468)	(0.1824)	(0.1421)
Covariance $W = LL^{\mathsf{T}}$	40.5225	14.2217	14.8488
	14.2217	14.9076	12.8843
	14.8488	12.8843	13.7103
Observations: 11967			
Nr. of Individuals: 2647			
Loglikelihood: -5244.5			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

For the static model, we find similar effects of personal and household characteristics in all specifications. See e.g. the results for the model with Big-Five factor markers for women in Table 2.3 and for men in Table 2.4.[19] For women, only few of the personal and household characteristics are individually significant in the self-employment equation. In contrast to this, almost all of them are highly statistically significant (i.e., most at the 1%-, and some at the 5%-level) in the other two equations. In other words, personal characteristics do not help us much to explain the difference between being self-employed or working as an employee, but they are helpful in explaining the difference between employment and unemployment or not participating in the labour market. The only variables that have a significant coefficient in the self-employment equation are age and household size (both at the 1%-level). We find that the self-employed are on average older, and that the larger the household, the more likely women are to choose self-employment over wage-employment.

For men the coefficients on personal and household characteristics are also very similar across different models, but they differ from what we found for women. We still find that age is significant at the 1%-level and has a positive sign, implying that also for men, the chances that an individual is self-employed increase with age. However, we do not find a significant coefficient for household size. Unlike women, men with high education are significantly more likely to be self-employed and men with at least one child have a significantly lower probability to be self-employed.

For both genders, we find that the lagged health index does not enter the self-employment equation significantly. Health does have significant effects in the other equations though, and the coefficients are also jointly significant. The negative signs suggest that individuals who had bad health one period earlier are more likely to be observed in unemployment or out of the labour force.Similarly, we find an insignificant effect of the EP distance for women, although the sign is, as we would expect, positive in the other two equations. For men on the other hand, we find a significant effect (p-value < 0.01) also in the self-employment equation. The negative sign is in line with our expectations: as the EP distance decreases, an individual is more likely to be self-employed.

For the Big-Five factor markers, effects differ by gender. For both women and men, individuals who are more open to experiences are also more likely to be self-employed than employees. The effect is however twice as strong in magnitude for men. For women, we find that also emotional stability is significant at the 10%-level. The negative sign implies that higher emotional stability reduces the chances to become self-employed. For men on the other hand, we find that high scores for extraversion increase the likelihood of being self-employed, and that high scores for agreeableness reduce it. These effects are in line with our expectations.

In all static models, the inverse Mills ratio is highly significant for the self-employed and for the equation explaining "not in the labour force." It has a positive sign in the self-employment

[19]The pooled regression results for the model with Big-Five factor markers are in Appendix 2.E, Table 2.E.7. The results for all other models are available upon request.

equation and a negative sign in the equation for not in the labour force. For women, it is also significant and negative in the unemployment equation. This suggests that, keeping observed characteristics constant, attrition is correlated with the unobserved factors driving someone's labor market status, implying that it is important to correct for attrition bias.

Looking at the estimates for unobserved heterogeneity in the mixed logit model (the matrix L driving the covariance matrix), we see that all coefficients are highly significant. The variance of the unobserved heterogeneity in self-employment is much larger than for the other two states. In particular for men the covariances have about the same magnitude as the variance for unemployment or being out of the labour force. For women, the covariances differ more from each other and we see that, for given observed characteristics, self-employed individuals are also more likely to be out of the labour force.

2.4.3 Second stage — dynamic models

The dynamic models always outperform their static counterparts: Likelihood ratio tests reject the null hypothesis that the dynamic factors play no role, i.e. the lagged labour states are always jointly significant (p-value of 0.0000 for all).

Second, the Big-Five factor markers are jointly significant. Moreover, if we replace the big five with the EP distance or the health index, we find that the EP distance or the health index enter significantly (using LR tests, even at a 0.1% level).Hence adding either personality traits or information on an individual's health improves the model fit compared to a model with only the core personal and household characteristics. Comparing the two models with personality traits, Akaike's information criterion suggests that we should choose the model with the Big-Five factor markers over the one with the EP distance for both men and women.[20] The lagged health index does not enter significantly in the self-employment equation of the dynamic model either. In the following, we will therefore focus on the model including the Big-Five. [21]

Third, when we test for the joint significance of the inverse Mills ratio we fail to reject the null hypothesis that the coefficients are jointly equal to zero at any conventional significance level (p-values 0.188 for women and 0.395 for men). The results for the regressions without the Heckman correction are shown in Table 2.5 for women and Table 2.6 for men.[22] They indeed do not differ much from the results with the Heckman correction. This suggests that correcting for attrition bias is not essential once we estimate the dynamic model.

How do the results change from the static models once we include dynamic effects? For women, age loses its significance in the self-employment equation while for men, age remains significant

[20]See the corresponding regression results in Appendix 2.E, Table 2.E.12 (pooled), 2.E.13 (women), and 2.E.14 (men). The coefficient on the EP distance is always insignificant in the self-employment equation.

[21]The results for the model with the lagged health index can be found in Appendix 2.E, Table 2.E.15 (pooled), 2.E.16 (women), and 2.E.17 (men).

[22]The regression resuls with the Heckman correction are in Appendix 2.E, Tables 2.E.8 (women), 2.E.9 (men), and 2.E.10 (pooled). The pooled regression results without the Heckman correction are in Table 2.E.11.

Table 2.5: Dynamic model with Big-Five factor markers and no Heckman correction, women

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.98***	-6.56***	-5.49***
	(0.8184)	(0.6488)	(0.4526)
Age	0.01	0.03***	0.05***
	(0.0096)	(0.0082)	(0.0063)
Has partner	0.29	-0.67***	0.35**
	(0.2415)	(0.1886)	(0.1692)
Has child	-0.34	-0.23	-0.43**
	(0.299)	(0.2331)	(0.192)
Middle education	-0.11	-0.8**	-0.9***
	(0.5784)	(0.3503)	(0.2933)
High education	0.18	-1.53***	-1.41***
	(0.6016)	(0.3807)	(0.3122)
Household size	0.13	0.04	0.05
	(0.1201)	(0.0977)	(0.0812)
F1: extraversion	0.03	-0.06	-0.04
	(0.0972)	(0.0895)	(0.0665)
F2: agreeableness	-0.07	0.03	0.01
	(0.1107)	(0.0907)	(0.0698)
F3: conscientiousness	-0.05	-0.14*	-0.14**
	(0.099)	(0.0839)	(0.0653)
F4: emotional stability	-0.1	-0.35***	-0.21***
	(0.0894)	(0.0827)	(0.0609)
F5: openness for experience	0.25**	0.23***	0.11
	(0.1004)	(0.0895)	(0.0713)
Last state: self-employed	4***	1.07**	1.13***
	(0.2327)	(0.4764)	(0.2906)
Last state: unemployed	0.13	2.96***	1.76***
	(0.5087)	(0.2809)	(0.2164)
Last state: not in LF	0.61**	1.56***	2.14***
	(0.2611)	(0.1973)	(0.1042)
Initial state: self-employed	4.92***	1.41**	2.3***
	(0.5324)	(0.5596)	(0.36)
Initial state: unemployed	3.47***	3.35***	3.18***
	(0.693)	(0.5308)	(0.4772)
Initial state: not in LF	2.84***	3.17***	4.27***
	(0.3794)	(0.2931)	(0.2381)
L	2.26***		
	(0.2119)		
	1.12***	1.26***	
	(0.2288)	(0.2282)	
	1.25***	1.43***	0.63***
	(0.1803)	(0.1641)	(0.1701)
Covariance $W = LL^{\intercal}$	5.0893	2.5314	2.8172
	2.5314	2.8507	3.2082
	2.8172	3.2082	4.002
Observations: 14435			
Nr. of Individuals: 3267			
Loglikelihood: -6089.36			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.6: Dynamic model with Big-Five factor markers and no Heckman correction, men

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.82***	-8.77***	-6.25***
	(0.7965)	(0.9056)	(0.5853)
Age	0.02**	0.08***	0.07***
	(0.0099)	(0.0117)	(0.0087)
Has partner	-0.08	-0.34	-0.32
	(0.2537)	(0.3139)	(0.2259)
Has child	-0.28	-0.07	-0.11
	(0.3129)	(0.4071)	(0.2749)
Middle education	-0.09	-0.62	-1.12***
	(0.5254)	(0.4082)	(0.3191)
High education	0.27	-1.38***	-1.99***
	(0.5303)	(0.4576)	(0.3549)
Household size	0.18	-0.1	-0.07
	(0.1278)	(0.1811)	(0.1269)
F1: extraversion	0.28**	-0.03	-0.03
	(0.1097)	(0.1277)	(0.0855)
F2: agreeableness	-0.14	0.02	-0.05
	(0.0927)	(0.1225)	(0.0823)
F3: conscientiousness	-0.19**	-0.12	-0.15*
	(0.0958)	(0.1119)	(0.0884)
F4: emotional stability	-0.25**	-0.33***	-0.36***
	(0.1026)	(0.116)	(0.0825)
F5: openness for experience	0.24**	0.15	0.08
	(0.1035)	(0.1187)	(0.085)
Last state: self-employed	4.17***	0.36	0.74*
	(0.2254)	(0.7462)	(0.3905)
Last state: unemployed	0.24	2.65***	1.26***
	(0.5737)	(0.3338)	(0.3108)
Last state: not in LF	0.57*	0.66**	1.59***
	(0.3281)	(0.2959)	(0.1751)
Initial state: self-employed	4.75***	2.7***	2.28***
	(0.5798)	(0.7191)	(0.4486)
Initial state: unemployed	2.6***	4.13***	3.76***
	(0.9068)	(0.5922)	(0.5508)
Initial state: not in LF	2.07***	4.04***	4.31***
	(0.5227)	(0.4126)	(0.3508)
L	2.1***		
	(0.2383)		
	1.48***	1.27***	
	(0.3285)	(0.3264)	
	1.17***	1.62***	0.23
	(0.2578)	(0.2006)	(0.4156)
Covariance $W = LL^{\intercal}$	4.4139	3.1113	2.454
	3.1113	3.7993	3.7779
	2.454	3.7779	4.0285
Observations: 11967			
Nr. of Individuals: 2647			
Loglikelihood: -4067.57			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

at the 5%-level. The coefficient on household size for women, and on the dummy for having at least one child for men are no longer significant in the self-employment equation either.

Concerning the factor markers, we find that only the fifth factor, openness for experience, remains significant for women. It still has a positive sign. We find an increase in the probability for a woman to be self-employed over being an employee by a factor of 1.28 if her score on openness for experience increases by one standard deviation. For men, we also find that the coefficients on the factor markers decrease in magnitude, with the largest change occurring in the fifth factor, translating to a 40% smaller increase in the relative probability for self-employment. Its impact is now also of approximately the same size as for women, and the p-value increases to 0.023. For men we also find a change in the other factor markers: Agreeableness is no longer significant but emotional stability and conscientiousness are. Their signs are negative and thus opposite to what we would expect based on the arguments underlying the EP distance. Regarding conscientiousness this is, however, not entirely surprising, considering that we already saw in section 2.2.3 that the self-employed in the LISS panel are on average scoring lower on conscientiousness than employees. Finally, the coefficient on extraversion remains statistically significant (also at the 5% level) and has, as expected, a positive sign. For men, an increase in the score for extraversion by one standard deviation increases the relative probability to be self-employed by a factor of 1.32 compared to being employed.

Looking at the lagged labour market state variables we find persistence for both genders: Having been in a given labour state one period earlier increases the probability to be in that state – i.e. the diagonal in the block of coefficients for lags shows the largest values. This is what we would expect given the pattern of the transition probabilities in Table 2.2. The coefficient on lagged self-employment in the self-employment equation stands out as the largest of all, implying stronger state dependence in self-employment than in other labor market states.

We find that all coefficients for the lags are positive, implying that given that non-employees are more likely to end up in any of the other three states than employees. It should also be noted that in terms of the relative size of coefficients, the lagged labor market state dummies are much more important than household characteristics or personality traits. [23]

Finally, the estimated variances of the unobserved heterogeneity terms become substantially smaller once we include the dynamics. The variance in the unobserved heterogeneity for the self-employed is still larger than for the other two states but by a much smaller factor. The same holds for the estimated covariances. These show that, keeping observed explanatory variables constant, self-employed women are more likely to be out of the labour force than the unemployed, whereas the opposite holds for men.

[23] Initial values are also strongly significant with substantial coefficients. This indicates that the individual effects are correlated with the initial observation, as expected.

Table 2.7: Simulated transition probabilities (in %), men

Labour state	All individuals				45 to 60 year old			
past \ current	0	1	2	3	0	1	2	3
0: employee	93.58	1.59	1.29	3.54	92.15	1.53	1.63	4.69
1: self-employed	7.33	90.11	0.82	1.74	5.95	90.81	1.14	2.1
2: unemployed	22.22	2.78	44.7	30.3	22.18	2.33	43.58	31.91
3: not in labour force	25.03	2.51	8.68	63.77	22.62	2.1	9.71	65.57
Total	74.14	12.99	3.06	9.8	69.01	14.32	3.91	12.76

Based on dynamic model with Big-Five factor markers, 15310 observation pairs (n=2694).
Source: LISS Panel, missing values for personal/household characteristics are extrapolated.

2.5 Simulations

This section presents two different types of simulations. We first show the simulated transition probabilities for the complete LISS panel. Then we show simulated employment paths for benchmark individuals, to illustrate the implications of mobility into and out of self-employment. We only present simulations based on the model specification with the Big-Five factor markers.

2.5.1 Transition probabilities

To correct for attrition, we simulate under the counterfactual assumption that none of the individuals leave the sample. For those who do in reality, we need to impute the values of the covariates. We assume that, apart from age, personal and household characteristics remain the same.[24] The missing Big-Five factor marker values are completed with the same mean values used to fill in the initial gaps as described in section 2.2.3. We then, for each individual i, draw one vector of unobserved heterogeneity components μ_i from the multivariate normal distribution (with mean zero and covariance matrix given by $\hat{L}\hat{L}^{\intercal}$). In each time period for each individual and labour state, we then draw independent error terms ϵ_{ijt} from a Type 1 extreme value distribution. Taking the first labour market state that we observe for an individual as given, we then simulate individuals' labour market state outcomes for the following time periods.

Tables 2.7 and 2.8 show the simulated transition probabilities for men and women, based on the dynamic model with Big-Five factor markers and no Heckman correction.[25] The left panels show the results for all men and women and the right panels show the results for the 45 to 60 year old subsamples. Comparing the total shares of all labour market states with the shares in Table 2.2, we find that the simulations for the whole sample lead to results that are similar to what we observed initially in the data.

We also see that our simulations, in particular for women, overestimate the probability that individuals remain self-employed. As a consequence, the transitions out of self-employment

[24] This assumption is not too farfetched. All covariates have relatively small within variation.

[25] We choose the model without Heckman correction since we could not reject the null hypothesis that the correction terms all have coefficient zero.

Table 2.8: Simulated transition probabilities (in %), women

Labour state	All individuals				45 to 60 year old			
past \ current	0	1	2	3	0	1	2	3
0: employee	91.61	1.57	1.34	5.48	89.98	1.5	1.62	6.9
1: self-employed	8.18	87.54	0.81	3.47	7.12	86.97	1.33	4.58
2: unemployed	18.39	1.61	40.65	39.35	16.9	1.94	38.78	42.38
3: not in labour force	15.72	1.82	6.81	75.66	14.47	1.83	6.22	77.48
Total	66.72	8.87	3.68	20.73	60.63	8.91	4.17	26.29

Based on dynamic model with Big-Five factor markers, 19138 observation pairs (n=3325).
Source: LISS Panel, missing values for personal/household characteristics are extrapolated.

into other labour market states are underestimated but the general pattern seen in the data is nevertheless reproduced. Overall, it looks like we fare slightly better for men than for women in terms of replicating the observed transition probabilities.

The right panels with the results for 45 to 60 year old individuals show that when we consider older individuals, we observe approximately the same probabilities as for the complete samples to remain in self-employment. Hence, even for this older age group, where one generally assumes that projections are less prone to errors due to the smaller time horizon on which forecasts are made, we find that the assumption that the self-employed stay self-employed has clear limitations.

2.5.2 Individual simulations

By simulating the employment paths for the chosen benchmark individuals, we show how the transition probabilities translate into individual probabilities of remaining in the same labour market state for a longer time period. If we would simply take the probability of remaining in self-employment calculated above to the power 10, we would conclude that 38.14% of men and 24.76% of women remain self-employed for 10 years. This, however, would ignore that the probabilities depend with covariates that change over time (e.g., age, time dummies).

We therefore simulate the employment paths from 2008 until 2017, for benchmark individuals who are self-employed in 2007. We fix age to 45 years at the start and set other personal and household characteristics equal to the median by gender for the self-employed in 2007. Thus the benchmark individuals have a partner and one child. The self-employed male benchmark has high education, the female has medium education. The median values for the Big-Five factor markers corresponding to these personal characteristics (for age ranging from 45 to 54 years) are presented in Table 2.E.3 in Appendix 2.E.

Next we take 500 draws of the benchmark individual's unobserved heterogeneity vector μ_i, and then proceed in the same way as with the forward simulations for the transition probabilities, drawing the ϵ_{ijt} from a Type 1 extreme value distribution once in each of the time periods for

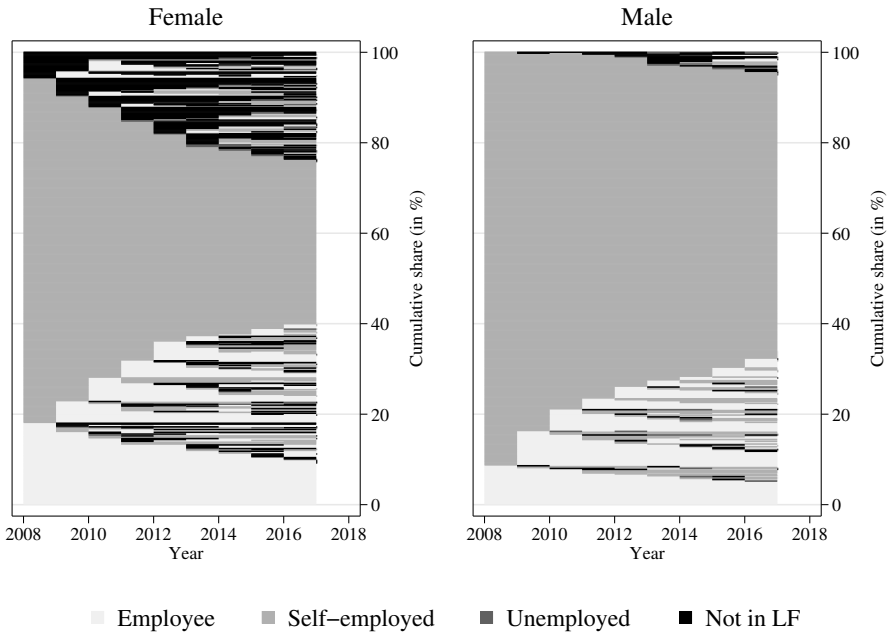


Figure 2.3: Simulated employment paths by gender: self-employed in 2007 (median characteristics)

each labour market state. The resulting employment paths for the self-employed benchmark man and woman are shown as sequence index plots in Figure 2.3.

The sequence index plots show all simulated employment paths. The paths are stacked vertically on top of each other and each path is a (thin) horizontal line on which each year is coloured according to the labour market state the path takes in that year. We order the paths by labour market states, starting with the state in the first period, followed by the second period, etc. The scale of the y-axis reflects cumulative shares (in percent). On the righthand side of Figure 2.3, we can see that the benchmark male has a chance of approximately 60% to remain self-employed throughout the ten years. For the benchmark female, this probability is only 35%. The main reason for the difference is the larger probability to leave the labour force for women.

Are these the probabilities we need to evaluate mobility? We view them as lower bounds on the probabilities of "persistent self-employment," because individuals can also leave self-employment for a short time and then return. If we are concerned about pension savings (our initial motivation), which tend to be much lower for the self-employed than for employees, the paths where self-employment is interrupted by a short employee spell are close to "persistent self-employment," since occupational pension wealth accumulated during the short employee spell will be small. A static micro-simulation which keeps everyone who is self-employed always in self-employment will then be a good approximation. We therefore also define an upper bound

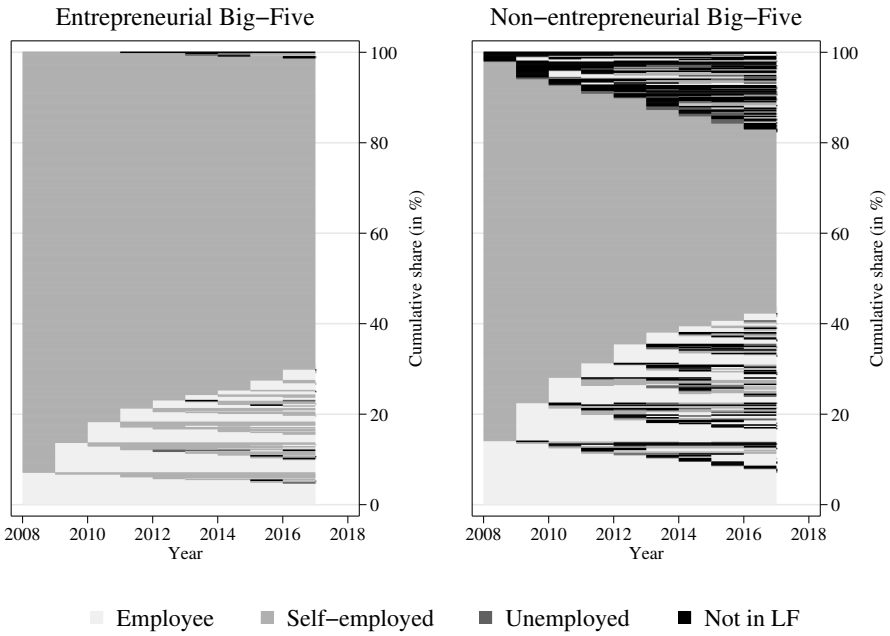


Figure 2.4: Simulated employment paths: male, self-employed in 2007

on the probability of "persistent self-employment," counting all paths where at least half of the time, five years, is spent in self-employment. This gives probabilities 80% for the male and 58.6% for the female benchmark individual starting as self-employed in 2007.

These probabilities obviously depend on the characteristics of the benchmark individuals. For example, if we change the male benchmark's education level to intermediate education, the lower and upper bounds for the male benchmark would fall to 48.6% and 72.2%. Similarly, if we would change the female benchmark's education to high education, her lower and upper bounds of persistent self-employment would rise to 45.0% and 67.8%.

Finally, since we are interested how personality affects individual probabilities to remain self-employed, we follow Obschonka et al. (2013) and define an "entrepreneurial" profile with high levels of extraversion, conscientiousness, emotional stability, and openness, and a low level of agreeableness. Taking the standardization of the factor markers into account we assign high and low values 2.5 and -2.5. In addition, we define a "non-entrepreneurial" profile that takes the opposite values. Figures 2.4 and 2.5 shows the sequence index plots for the corresponding male and female benchmark individuals who are self-employed in 2007 (other characteristics are the same as before).

As expected, the benchmark individuals with "entrepreneurial" Big-Five factor markers have a larger probability to remain self-employed than the "non-entrepreneurial" ones. The differences are large: the probabilities to stay self-employed throughout the ten year period is 67.6%

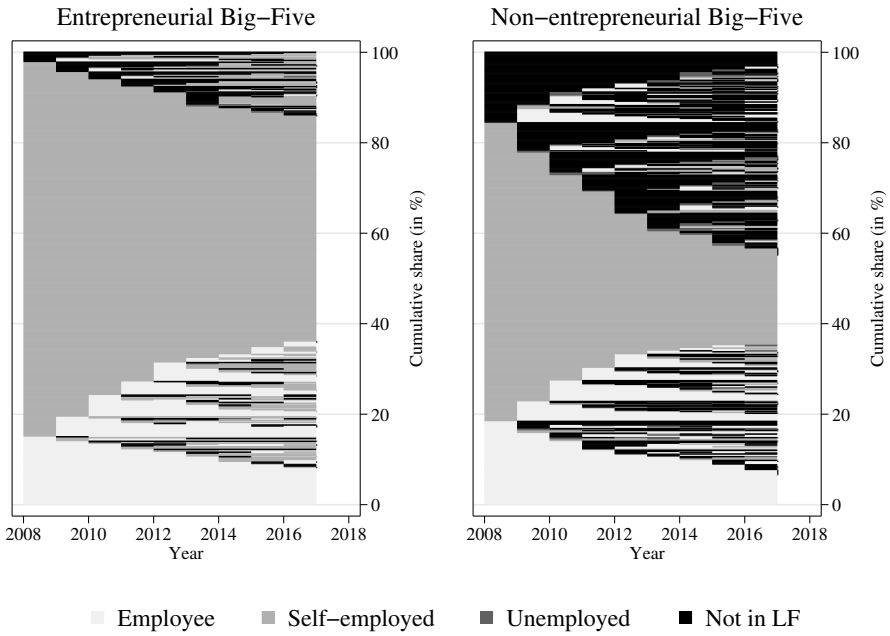


Figure 2.5: Simulated employment paths: female, self-employed in 2007

(59.2%) for the entrepreneurial male (female) and 38.4% (22.2%) for the non-entrepreneurial male (female). The changes in the probabilities compared to Figure 2.3 are larger for the "non-entrepreneurial" than for the "entrepreneurial" individuals, in line with the entrepreneurial characteristics of the self-employed benchmark individuals used in these earlier figures.

Table 2.9 summarises the simulation results for each combination of education level and personality profile. In the two rightmost columns, the table shows the lower and upper bounds based on our earlier definitions. The other columns show the share of employment paths in self-employment at a specific point in time.[26] Note that the probability of self-employment is always lower for a female individual in comparison to a male benchmark individual, also in cases where both have exactly the same characteristics. Furthermore, a shift from high to medium education increases the difference in probabilities between the "entrepreneurial" and "non-entrepreneurial" case. In addition to this, the difference between the lower and upper bound (the final two columns) within each case also increases. Finally, we also find that the share of paths observed in self-employment in a year stabilizes after the fifth to sixth year in the simulation.

These results point at the limitations of a static microsimulation approach for pension wealth and income projections. If, for example, we take the most stable of the benchmark individuals, a

[26]See Figure 2.F.1 in Appendix 2.F for a graphic representation of the evolution of the shares over time and type of benchmark individual.

Table 2.9: Probability (in %) that different benchmark individuals who are self-employed in 2007, are self-employed in later years

Big-Five	in 2008	in 2009	in 2012	in 2017	10 years	5+ years
Male: high education						
median	91.4	83.8	77.6	72.0	61.0	80.0
entrepreneurial	93.0	86.8	82.2	77.6	67.6	84.4
non-entrepreneurial	83.8	72.2	60.4	51.4	38.4	62.4
—— medium education						
median	88.2	78.6	69.4	63.4	48.6	72.2
entrepreneurial	90.8	83.0	77.2	70.8	59.4	79.2
non-entrepreneurial	77.8	62.6	43.8	32.0	22.2	47.0
Female: high education						
median	80.4	75.0	59.8	64.0	45.0	67.8
entrepreneurial	86.4	81.8	73.2	74.2	59.2	78.4
non-entrepreneurial	73.6	66.2	45.6	47.0	29.0	54.2
—— medium education						
median	76.2	68.6	51.6	54.0	35.4	58.6
entrepreneurial	82.8	77.2	64.0	69.0	49.2	72.0
non-entrepreneurial	66.0	56.4	33.4	34.0	19.6	39.4

Based on dynamic model with Big-Five factor markers and no Heckman correction (n=500).
The benchmark individual is self-employed in 2007. The rightmost column gives the probability that the individual is self-employed in at least 5 of the 10 years.

male with high education and "entrepreneurial" Big-Five factor marker levels (left panel of Figure 2.4), we still get chance of 12.6% to spend at least seven of the next ten years as an employee. Given the context of the Dutch pension system, such an individual would accumulate pension savings in the second pillar during at least seven of the ten years – and would therefore end up in a better financial position than what a model assuming that self-employment is an absorbing labour market state would predict. That is, a static micro-simulation would over-predict the share of self-employed with (too) low pension savings.

2.5.3 The impact of the macro economy

As the number of self-employed in the Netherlands has increased in recent years, one might wonder if changes in the macro economy have been drivers of this increase. Our model can answer this question insofar that we estimate time effects for each year that capture the macro-economic changes. The estimates of the time effects for the dynamic model with the Big-Five factor markers and no Heckman correction are shown in Table 2.10. We find large positive effects for the unemployment probability, in line with the notion that 2008 was the star of the economic crisis. For self-employment, we find that most of the years do not differ significantly from the base year (2008). The only significant time effects on self-employment are a negative effect for

Table 2.10: Time effects; dynamic model with Big-Five factor markers and no Heckman correction

	Women			Men		
	Self-employed	Unemployed	Not in LF	Self-employed	Unemployed	Not in LF
2009	0.26	0.66	-0.15	-0.58**	0.12	-0.05
	(0.2429)	(0.4434)	(0.1574)	(0.2700)	(0.5510)	(0.2134)
2010	-0.04	0.85**	-0.13	-0.40	0.62	0.09
	(0.2845)	(0.4040)	(0.1764)	(0.3328)	(0.5133)	(0.2223)
2011	0.22	1.21***	0.17	-0.06	0.93*	-0.04
	(0.3479)	(0.4116)	(0.1747)	(0.3396)	(0.5072)	(0.2675)
2012	-0.20	1.17***	-0.00	-0.34	0.84	0.05
	(0.3374)	(0.4299)	(0.1778)	(0.3596)	(0.5202)	(0.2404)
2013	0.41	1.65***	0.50***	-0.47	1.07**	0.41
	(0.3451)	(0.4025)	(0.1727)	(0.3629)	(0.5355)	(0.2490)
2014	0.51	2.16***	-0.09	-0.18	1.33***	0.41*
	(0.3241)	(0.3719)	(0.1631)	(0.3354)	(0.4988)	(0.2329)
2015	0.38	2.05***	0.26	-0.05	1.14**	0.11
	(0.2800)	(0.3789)	(0.1638)	(0.2794)	(0.5290)	(0.2359)
2016	0.77***	1.98***	0.15	-0.45	1.11**	0.18
	(0.2807)	(0.3863)	(0.1733)	(0.3237)	(0.5486)	(0.2491)
2017	0.55**	1.81***	-0.20	-0.51	1.55***	-0.21
	(0.2759)	(0.3794)	(0.1809)	(0.3315)	(0.5140)	(0.2637)

Time effects from the regressions shown in Table 2.5 and 2.6. 2008 is the base year.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

men in 2009, and positive effects for women in the final two years (2016 and 2017). Keeping the trend as shown in Figure 2.1 in mind, this finding is actually as expected as there is no visible impact on self-employment by e.g. the Great Recession.

We can study the impact of macro variables further by repeating the simulation but leaving out the time effects. This has the same effect as assuming that the macroeconomic situation in 2008 holds for all years, which might be particularly interesting considering that the Great Recession started in the second quarter of that year. Table 2.11 shows the simulation results for this exercise.[27] Comparing them with the results in Table 2.9 we observe that the changes for women and men generally go in opposite directions: Without the year effects, the male benchmark individuals would all have a higher probability of remaining self-employed while the female individuals would have a lower probability. This is in line with the negative time effects for men and the mainly positive time effects for women in Table 2.10. Still, the differences between the probabilities of persistent self-employment (the final columns of Tables 2.11 and 2.9) accounting and not accounting for the macro-economic developments are rather small compared to the differences that can be explained by characterstics such as education level or an entrepreneurial individual attitude.

[27] See Figure 2.F.2 in Appendix 2.F for a graphical representation.

Table 2.11: Probability (in %) that different benchmark individuals who are self-employed in 2007, are self-employed in later years: macro-economic situation as in 2008

Big-Five	in 2008	in 2009	in 2012	in 2017	10 years	5+ years
Male: high education						
median	91.4	86.2	82.8	76.4	66.4	82.6
entrepreneurial	93.0	88.6	84.8	81.0	70.4	86.4
non-entrepreneurial	83.8	76.0	65.0	60.2	46.0	69.6
—— medium education						
median	88.2	82.8	74.2	69.8	56.6	77.2
entrepreneurial	90.8	85.8	81.0	75.8	56.6	83.2
non-entrepreneurial	77.8	67.6	52.4	45.4	33.2	55.8
Female: high education						
median	80.4	72.2	60.0	56.0	42.2	64.6
entrepreneurial	86.4	80.6	72.8	69.0	55.2	76.2
non-entrepreneurial	73.6	62.6	45.8	37.0	26.2	49.6
—— medium education						
median	76.2	67.4	50.6	44.0	32.6	55.0
entrepreneurial	82.8	76.4	65.2	61.6	47.4	69.0
non-entrepreneurial	66.0	53.4	32.8	26.2	17.2	36.4

Based on dynamic model with Big-Five factor markers and no Heckman correction (n=500).
The benchmark individual is self-employed in 2007. The rightmost column gives the probability that the individual is self-employed in at least 5 of the 10 years.

2.6 Conclusion

We have analyzed transitions into and out of self-employment and other labor market states for individuals of working age in the Netherlands. We have used a dynamic multinomial logit model with random individual effects. Our explanatory variables include the "Big Five" personality characteristics, and we find that an entrepreneurial or non-entrepreneurial personality has a large effect on the probabilities to stay or become self-employed. Expanding the static model to a dynamic framework where the current labor market state has a causal effect on the next labor market state, substantially improves the fit of the model and reduces the estiated importance of unobserved heterogeneity. Still, we find substantial unobserved heterogeneity, particularly among the self-employed.

There is clear evidence of persistence (state dependence) in self-employment, but self-employment is not an absorbing state. The probability to remain self-employed in the next year is on average around 90% for men and slightly lower for women. But the probabilities to remain self-employed for a longer time period are much smaller, as illustrated by simulated employment paths over multiple years for benchmark individuals. Our benchmark self-employed male has an 80% chance to spend the majority of the next ten years in self-employment. This

probability falls to 62% if the individual does not have an entrepreneurial personality, or even to 47% if, in addition, he has medium education rather than high education (the benchmark). For the self-employed benchmark woman, the chances to spend the majority of the next ten years in self-employment are only 59%, and this falls to below 40% if the woman does not have an entrepreneurial personality.

With the ongoing pension reforms, there is a lot of recent interest in pension adequacy of a heterogeneous population, with a focus on vulnerable groups such as the self-employed. Our results suggest that future work on projecting pension incomes and pension adequacy should account for the labour market dynamics and the transitions between labour market states in which individuals do or do not (sufficiently) accumulate pension wealth. Combining the type of model and dynamic simulations here with administrative data on how much pension wealth is accumulated in a given labor market state, seems a fruitful avenue for future research.

Appendices

2.A Self-employed individuals in the LISS panel

Among all longitudinal surveys, there are three instances within the LISS panel through which we can identify self-employed individuals. First, information about an individual's labour market status is stored in the household box. When asked about their primary occupation the survey responder is prompted with fourteen options. Of these, the third, "Autonomous professional, free-lancer, or self-employed", helps us to identify the self-employed. There are three problems with using this information to define an individual's labour market state. First, it does not enable us to distinguish between SSEs and other self-employed. Second, the questionnaire gives no instruction on how "primary" is defined. Hence individuals with more than one occupation may rank these by hours, or by how much they identify themselves with each of them.[28] Lastly, this is further compounded by the fact that the survey is filled out by the household's contact person.[29] This can potentially lead to conflicting answers vis a vis individuals who answered themselves, both for e.g. part-time (self-)employed and in particular also for DGAs.[30]

Second, we can identify self-employed individuals using either one of two annual core studies: the *Work and Schooling* and the *Economic Situation: Income* study. Currently there are eleven waves available for both, covering the years 2008-2018. There are usually slightly fewer individuals who answer the income survey compared to the work and schooling survey.[31] Overall, however, the two samples are comparable in size and the majority of individuals answer both.

In the *Work and Schooling* survey we can classify individuals according to their primary occupation based on hours worked.[32] This implies that the working population consists of all individuals who indicate that they do paid work. We then split these individuals into employees and self-employed, based on follow-up questions on their primary occupation. The self-employed are therefore all individuals who answer that they are either *self-employed/freelancer*, or *independent professional* or *DGA*.[33] As an exception to the primary occupation rule, we also define those

[28]Even if the two coincide, the outcome may be different from how the person would be recorded in labour statistics. Take an individual who works part-time for less than 50% and works in the household for the rest of the week. This individual may well answer that the primary occupation is house work, leading to the classification "out of the labour force". But according to the official statistics of Statistics Netherlands (CBS), anyone working more than one hour per week for pay would count as part of the working population.

[29]In most households with more than one adult member, several adults participate in the LISS panel.

[30]DGAs ("directeur grootaandeelhouder" in Dutch — majority shareholder director) are individuals who work for an incorporated firm (either an NV or BV, i.e. a Ltd. or private Ltd. company in the British context) in a relatively high administrative position while holding a large part (or the majority) of the firm's shares. DGAs are treated as employees from the perspective of the Dutch tax authority while they may see themselves, in case of being (one of) the company owner(s), as self-employed.

[31]Numbers and response rates differ by survey wave. In 2014 for example, 7746 individuals were invited for the income survey and 7957 individuals for the work and schooling survey, with response rates of 78.9% and 82.6% respectively. In 2017, 6673 and 7256 individuals were selected with response rates of 80.3% and 80.4% respectively.

[32]If individuals work equal numbers of hours in two different jobs, they are asked to indicate the job that they consider more important.

[33]See Appendix 2.B.1 for a discussion of the inclusion of DGAs.

individuals as self-employed who indicate that they have their own business or have a partnership as a secondary occupation next to being employed.[34] Going back to the definition by primary occupation, anyone who is not part of the working population and receives unemployment benefits or is looking for a job is classified as unemployed; the remainder of individuals is classified as not in the labour force. Because participants are asked about their primary occupation at the time the survey is taken, this classification is a snapshot of the individuals' situation at the moment of data collection (which takes place around April and May of each calendar year).

Overall the LISS panel contains information on the labour status for almost 66,000 observations if we use the work and schooling data. Overall, we have overlapping information on approximately 45,000 observations for the income and work studies.[35]

2.A.1 Comparing the self-employment share in the two definitions

Despite the timing being different in the *Work and Schooling* classification compared to the income based classification the two should lead to a similar outcome. We expect small differences to arise because individuals generally do not switch occupations in January only. But since we have data over several years we will still pick-up the changes in labour market states in both data sets. Table 2.A.1 shows a comparison of the assigned labour market state according to the income based classification ("Income"), compared with the work and schooling based classification ("Work and schooling") as well as the classification based upon the household box with background variables ("Background variables").[36] The rows show the share of matches with labour market states in the "work and schooling" and "background variables" classifications for each labour market state in the income based classification. If the classifications would match perfectly, we would observe shares of 100% on the diagonal and zero elsewhere. As expected, we find that the match for the unemployed is not that good. Focusing on the self-employed and employees, we see that one fifth of the individuals who report income from self-employment activities are classified as employees in the work and schooling survey. Similarly, 26% of those with self-employment income is classified as an employee according to the background variables based classification. Only 63% of those with self-employment income are classified as self-employed accoring to the latter classification, compared to 71% for the work and schooling based classification.

The corresponding self-employment shares, calculated as the number of self-employed individuals per year as a fraction of the respective number of individuals in the working population (i.e. the sum of employees and self-employed), in the two samples after these restrictions are shown in Figure 2.A.1.

[34]This increases the total number of self-employed in the sample by approximately 21%, compared to a definition based on primary occupation only.

[35]Note that there are almost 11,000 observations combined for the years 2007 and 2018 that we cannot use because of the different timing of the two studies.

[36]The comparison is made for all observations in the corrected income sample excluding the corrected breaks (see section 2.A.2).

Table 2.A.1: Matching of labour market states across definitions (in % of income definition)

	Work and schooling				Background variables			
Income	0	1	2	3	0	1	2	3
0: Employee	95.36	1.22	1.64	1.79	91.81	0.76	1.88	5.55
1: Self-employed	20.30	71.02	2.74	5.95	26.26	63.77	2.71	7.26
2: Unemployed	9.18	1.39	34.21	55.22	8.94	1.59	39.53	49.94
3: Not in LF	15.69	2.66	10.18	71.48	14.31	2.61	8.21	74.87
Total	71.86	8.69	4.13	15.33	69.75	7.77	4.07	18.41

Source: LISS panel, own calcuations.

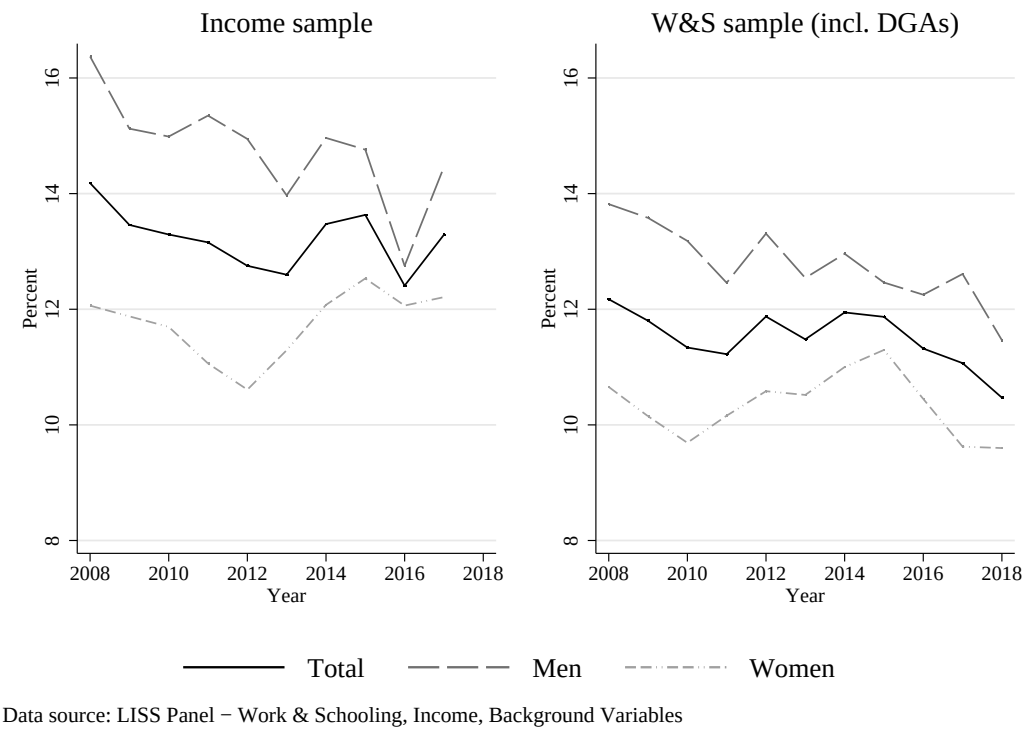


Figure 2.A.1: Comparison of self-employment shares in working population of 25-60 year olds across samples.

There are two observations that we can make based on this figure. First, the share of self-employed in the working population is lower than in the actual population shown in Figure 2.1. While the share in the income based classification looks more or less equal in 2008, we have to take into consideration that CBS classifies individuals with multiple occupations based on the one in which they work most hours. Hence, as we also include part-time self-employed, the share in the LISS panel, if it were representative, should actually be higher than the population share.[37] Second, we observe an upward trend in self-employment shares according to none of the classifications in the LISS samples. These two observations hint that we may have to deal with both initial selection (lower initial shares) and attrition bias (lack of upward trend) in the LISS panel. We will discuss these issues in more detail below.

2.A.2 Filling in the gaps, selection, and attrition

The model based sample restriction, i.e. having at least two consecutive observations, not only leads to a loss of data but also to a change in the evolution of employment shares over time in the new sample. In both the income and work based sample we see a clear downward trend after discarding the individuals with no sequence as well as any observations following a break in a sequence. The left panel of Figure 2.A.2, in comparison with the left panel in Figure 2.A.1 illustrates this for the income survey based classification. Thus the restriction of the sample to sequences creates (or worsens) attrition bias.

A closer look at the observations that are discarded because of the sequence restriction reveals that only a third of these observations belong to individuals who only participate in one wave of the income survey. That is, two thirds of the observations could be retained if we can correct for the break in the corresponding sequence. Furthermore, these are the observations (rather than the single wave answers) that are driving the downward trend shown in the left panel of Figure 2.A.2.

We take a conservative approach to filling in the breaks in sequences by imputing plausible values. That is, we only fill in one-year breaks, which account for the majority of all the breaks in the sequences. We however do not limit the correction to one break per individual but may fill in several breaks as long as these are only one period long. Nor do we extend a series beyond what we observe — that is, if an individual answers the work and schooling survey for more periods than the income survey, we do not use these other periods from the work and schooling survey to top up the income survey series. The procedure is as follows: we only consider the labour status information from the work based definition to fill in gaps in the income sample and vice versa, ignoring the information that is available in the background variables. We choose to do this because the number of correct matches between our definitions and a classification based on the background variables (as shown in Table 2.A.1) is quite low and we want to avoid generating

[37] Differences should not stem from the denominator — CBS includes anyone in the working population who works at least one hour per week. Our definitions for the working population should therefore be comparable.

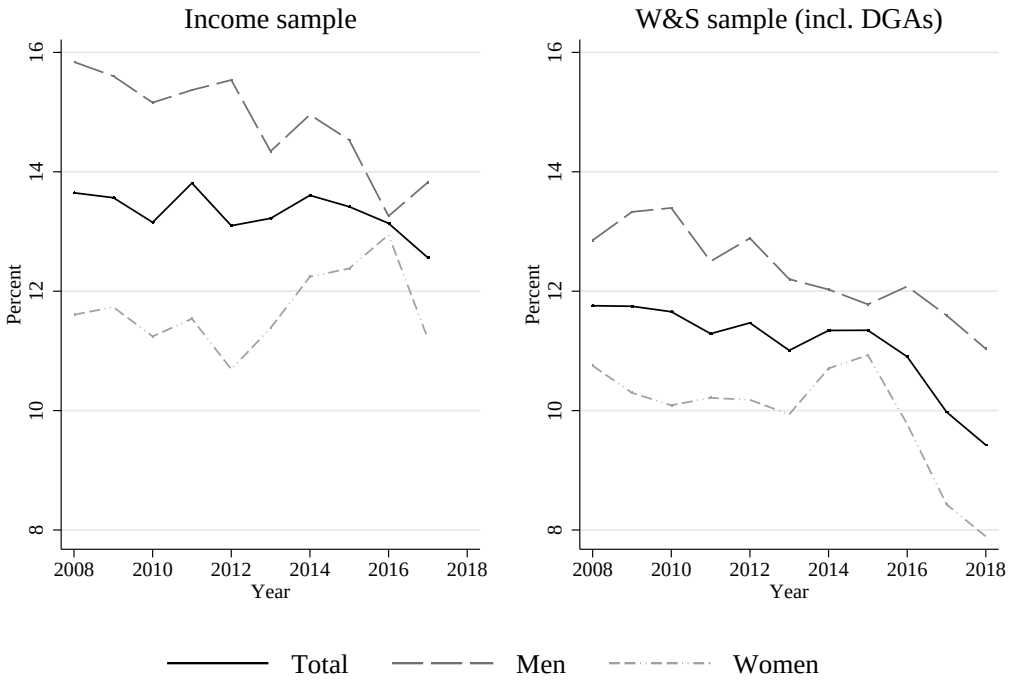
Data source: LISS Panel − Work & Schooling, Income, Background Variables

Figure 2.A.2: Comparison of self-employment shares based on income before/after correction

false labour market state transitions due to filled in gaps. In the income survey sample, we additionally make use of the question asked to self-employed individuals whether they were also self-employed in $t-2$. We only use this information to adjust for self-employment because we know that the work and schooling based sample misses 20% of the individuals that we categorise as self-employed in the income classification. To avoid selection bias, we do not use this question to fill in gaps for individuals without an answer in the work and schooling survey, because the question is only asked to individuals who are self-employed in $t-1$.

Based on this strategy we end up filling up at least one gap for more than one thousand individuals in the income survey sample. This in turn helps us to retain more than three thousand observations that would otherwise have to be discarded. More importantly, the downward trend in the self-employment shares is less pronounced after the correction. Figure 2.A.2 illustrates this for the income based classification. The impact of the correction on the work based classification is not that strong. This is likely because the number of additional observations that can be retained is rather low compared to the number of observations that is affected by the sequence restriction.

Figure 2.A.3 shows the corrected regression samples for both definitions. We can see that both samples display a slightly stronger downward trend in self-employment shares compared to the raw data in Figure 2.A.1. This negative trend also seems to worsen during later years and



Data source: LISS Panel − Work & Schooling, Income, Background Variables

Figure 2.A.3: Comparison of self-employment shares in corrected income and work panel

the effect is stronger in the work based classification. Because of this, and also due to the lower self-employment shares, we choose the income based classification for our analysis. The results discussed below, unless otherwise mentioned, are therefore for the income based classification (and the sample corresponding to it).

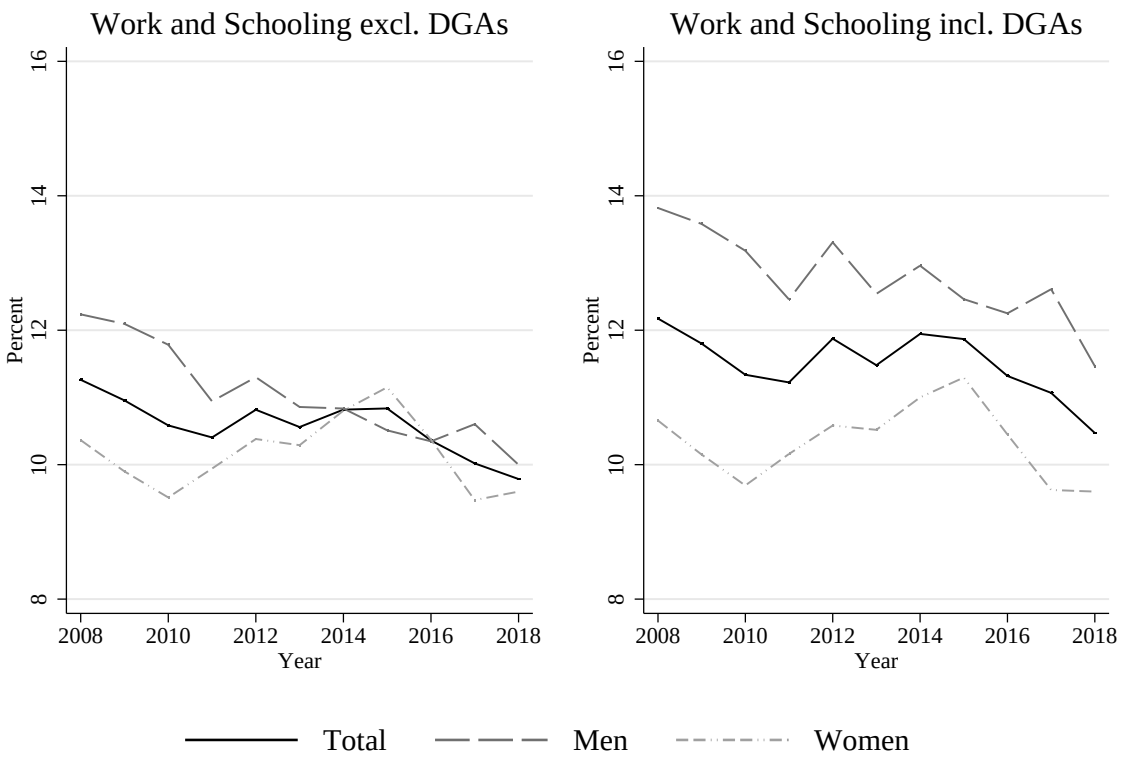
2.B DGAs and the self-employed

2.B.1 DGAs in the work and schooling survey

We use question cwxxx121, where "xxx" denotes the indicator for the year and wave, to classify individuals within the working population. The question's text explains to the survey takers that *[a] director of a limited liability or private limited company (Dutch: NV or BV, respectively) is generally on the payroll of that company. In that case, please enter that you [are an] employee in permanent or temporary employment. A majority shareholder director, also, generally receives an income as an employee. Nevertheless, if this [applies] to you, we request that you indicate that you [are] a (majority shareholder) director.*[38] Hence, self-employed individuals who have incorporated their company and act as its DGA should answer that they are DGAs, whereas e.g.

[38]Taken from the English version codebook of the Work and Schooling core study wave 10.

the directors of Shell or Unilever should answer with the option that they are *director of a limited liability or private limited company*. Therefore, including DGAs based on question cwxxx121 in the self-employment definition should not lead to a mistake in our definition of the self-employed.



Data source: LISS Panel − Work & Schooling, Income, Background Variables

Figure 2.B.1: Comparison of self-employment shares with and without DGAs in raw sample

Figure 2.B.1 illustrates this further. By comparing the left and the right panel we see that the inclusion of DGAs shifts the shares more or less equally across all periods, which is consistent with evidence presented in CBS (2014, p.13, figure 3.2.1). It does so mostly through increasing the share of self-employed among the working men, which is consistent with the finding reported by CBS in the same report that 80% of DGAs were men in 2012.[39]

2.B.2 DGAs in the income survey

In the income survey, question cixxx008 individuals are asked: *Did you receive income as employee in [t-1]?* including the explanation that *A [DGA] generally receives income as an employee as well*

[39]The shares presented in Figure 2.B.1 are calculated for all individuals from 25 to 60 years of age. The sample is further restricted to those for whom the covariates of the regression analysis are not missing. It is not conditional on observations being available for at least two periods consecutively.

and so please answer YES here. The question block on self-employment on the other hand makes no special mention of DGAs. The section is however vague when asking inviduals *[w]hich work situation as described [. . .] in [t − 1] applied to you (or for a part of [t − 1])?* and offering statements such as *work as an entrepreneur or as a freelancer (alongside a job)*, or *a company owner*, or *make profit (or losses) through enterprise in some other way [...]* in questions cixxx37 to cixxx044. DGAs are therefore only included in our definition of the self-employed if they self-identify as one of these options, otherwise they will be categorised as employees.

It is theoretically possible to identify DGAs through their wages because there is a legal requirement for a DGA to earn at least 45000 euro (gross). This minimum salary for DGA is substantially higher than the CPB's estimate of model income at 34000 euro for 2017[40] and we could therefore use reported income as an identification strategy. However, many individuals do not report their gross income in the survey, and the income brackets provided are not indicative enough as one bracket runs from 36000 to 48000.

2.C Model details

2.C.1 Correlation among the random effects

Following Train (2009, chapter 9), we use a Choleski transformation for the multivariate normals. As Train (2009, p.238) writes, the advantage of using this approach is that "for any pattern of covariance, there is some set of loadings from independent components that reproduces that covariance". We thus only have to make an assumption concerning the distribution of the unobserved heterogeneity but not of the covariance. Hence,

$$\mu_i = \xi_i^{\mathsf{T}} L^{\mathsf{T}} \tag{2.7}$$

where ξ_i is a $J \times 1$ vector of independent standard normal distributed variables, and L is the lower triangular Cholesky factorization of μ_i's covariance matrix W, such that $LL^{\mathsf{T}} = W$.

Substituting (2.7) and (2.3) in (2.2), and writing the utilities in vectorised form we have

$$y_{it}^* = X_{it}\beta + y_{it-1}\gamma + y_{i0}\delta + \xi_i^{\mathsf{T}} L^{\mathsf{T}} + \epsilon_{it} \qquad i = 1,\ldots,N\ ;\ t = 2,\ldots,T \tag{2.8}$$

where y_{it}^* is a $1 \times J$ row vector of indirect utilities for individual i at time t, γ and δ are $J \times J$ matrices of parameters, β is a $k \times J$ matrix of parameters, and L contains the parameters of the covariance structure. The elements of β, γ δ, and L have to be estimated (apart from normalizations).

[40] See https://www.cpb.nl/publicatie/macro-economische-verkenning-mev-2019

2.C.2 Likelihood function

The probability to observe a particular individual choosing labour state j at time t conditional on ξ_i in the multinomial logit model is then given by

$$Prob(y_{it} = j|X_{it}, y_{it-1}, y_{i0}, \xi_i) = \frac{exp(y_{it-1}\gamma_j + y_{i0}\delta_j + X_{it}\beta_j + \xi_i^{\intercal} L_j^{\intercal})}{\sum_{k=0}^{J} exp(y_{it-1}\gamma_k + y_{i0}\delta_k + X_{it}\beta_k + \xi_i^{\intercal} L_k^{\intercal})} \tag{2.9}$$

where L_j is the jth row of L.

It follows that the conditional probability of observing a sequence of choices for individual i is

$$Prob(y_i|X_i, y_{i0}, \xi_i) = \prod_t \prod_j Prob(y_{it} = j|X_{it}, y_{it-1}, y_{i0}, \xi_i)^{\mathbb{D}_{ijt}} \tag{2.10}$$

where $\mathbb{D}_{ijt}$ is an indicator function denoting whether state j is chosen by the individual. The unconditional probability, or likelihood function, is then given by

$$Prob(y_i|X_i, y_{i0}) = \int_{\xi_i} Prob(y_i|X_{it}, y_{it-1}, y_{i0}, \xi_i) f(\xi_i) d\xi_i \tag{2.11}$$

where $f(\xi_i)$ denotes the multivariate normal distribution of ξ_i with means zero. The log likelihood function to be estimated is thus

$$log\mathscr{L} = \sum_{i=1}^{N} logProb(y_i|X_i, y_{i0}) \tag{2.12}$$

2.C.3 Maximum Simulated Likelihood

The results of this section hold for both the static and the dynamic model. Note that in its current form the multinomial logit model described by equations (2.9) – (2.12) is not identified as there are too many parameters. For identification purposes we therefore take $j = 0$, that is wage-employment, as the base category. β_0, γ_0, δ_0 and the first column in L are normalised to zero. The parameters for the other alternatives $j = 1, 2, 3$ must be interpreted relative to the base category.[41] For example, the covariance matrix is given by

$$W = \begin{pmatrix} l_{11}^2 & \cdot & \cdot \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & \cdot \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}$$

and we estimate $J - 1 = 3$ coefficients for each variable.

Furthermore, the probabilities given by equation (2.11), i.e. $\int_{\xi_i}$ in particular, have to be simulated. As Train (2009, chapter 10) writes, $log\hat{P}$ is not an unbiased estimator for $logP$ because of the non-linear log operation, even if $\hat{P}$ is an unbiased estimator of P. Thus the bias in the simulator for $logProb(y_i|X_i, y_{i0})$ translates into a bias in the maximum simulated likelihood

[41] Note that dimensions are now reduced by 1, and all vectors and matrices are now of dimensions $(\cdot \times J - 1)$.

estimator. This bias however diminishes as more draws are used in the simulation. Using a large number of draws for ξ_i on the other hand increases the computational burden of the estimation. One way to reduce this burden is, instead of using independent random draws, to use an alternative method that provides better coverage of the support of the distribution of the individual and therefore leads to greater accuracy for a given number of draws. This can be achieved using Halton draws. Both Bhat (2001) and Train (2009, chapter 9.3.3) have shown that e.g. 100 Halton draws can provide more precise results than 1000 random draws.

In order to simulate $\int_{\xi_i}$, we take 150 draws for each individual from a $J-1$-dimensional Halton sequence in which we closely follow the method described in Train (2009, chapter 9.3.3). Based on the discussion above, 150 draws for each individual should be sufficient. Furthermore the panel is based on more than six thousand individuals, i.e. large in itself, which should also lower the need for more draws. In addition we randomise the Halton draws following the procedure described by Bhat (2003). As a base of the Halton sequences we use the primes 11, 13 and 7.

Returning to the model, the probability of observing an individual's observed sequence of labour state choices in the simulation is given by

$$Prob(y_i|X_i, y_{i0}) = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=2}^{T}\left(\prod_{j=1}^{J} Prob(y_i|X_{it}, y_{it-1}, y_{i0}, \xi_i)^{\mathbb{D}_{ijt}}\right)^{\mathbb{S}_{it}}$$

The simulated loglikelihood function is thus given by

$$log\mathscr{SL} = \sum_{i=1}^{N} log\left[\frac{1}{R}\sum_{r=1}^{R}\prod_{t=2}^{T}\left(\prod_{j=1}^{J} Prob(y_i|X_{it}, y_{it-1}, y_{i0}, \xi_i)^{\mathbb{D}_{ijt}}\right)^{\mathbb{S}_{it}}\right] \tag{2.13}$$

where R is the number of Halton draws taken to simulate $\int_{\xi_i}$, and $\mathbb{S}_{ijt}$ is an indicator function to control for the unbalanced panel, denoting whether an individual's observation enters in the estimation. That is, $\mathbb{S}_{ijt} = 1$ if y_i, y_{it-1}, X_{it} are observed.

All model specifications for the correlated random effects models reported in section 2.4 are estimated using own code written in Matlab 2017a. We solve them as an unconstrained minimisation problem using KNITRO as a solver and supplying the gradient as defined in the next subsection. As the best guess for the starting values we first estimate a pooled multinomial regression on the same covariates, including the lags, for each specification. The only difference is the non-inclusion of the unobserved individual heterogeneity. For this step we make use of the function "mnrfit" in Matlab's statistics toolbox.

2.C.4 First order derivatives

Let for notational ease $\mathsf{P}(y_{ijtr}) \equiv Prob(y_{ijtr}|X_{it}, y_{it-1}, y_{i0}, \xi_i)$ and similarly for all other probabilities. The simulated loglikelihood function from equation (2.13) is then given by

$$log\mathscr{SL} = \sum_{i=1}^{N} log\left[\frac{1}{R}\sum_{r=1}^{R}\prod_{t=2}^{T}\left(\prod_{j=1}^{J} \mathsf{P}(y_{ijtr})^{\mathbb{D}_{ijt}}\right)^{\mathbb{S}_{it}}\right]$$

Furthermore let vector θ be a vector with κ elements containing the stacked columns of all the coefficient matrices β, γ, δ, as well as a vector of all the elements of the Cholesky factorisation $[l_{11}, l_{21}, l_{22}, l_{31}, l_{32}, l_{33}]'$. And lastly, let V_{it} be a vector of the covariates, where $X_i t, y_{it-1}, y_{i0}$ are stacked accordingly to the elements in θ. Note that e.g. for a covariate in X_{it} its index κ in V_{it} is jointly defined by k and j. The last rows in V_{it} are given by the vector $[\xi_1, \xi_1, \xi_2, \xi_1, \xi_2, \xi_3]'$.

Then using section 2.C.2 the first order derivative can be written as

$$\begin{aligned}\frac{\partial log \mathscr{SL}}{\partial \theta_\kappa} &= \sum_{i=1}^{N}\left[\frac{1}{\mathsf{P}(y_i)}\left(\frac{1}{R}\sum_{r=1}^{R}\mathsf{P}(y_{ir})\left(\sum_{t=1}^{T}\mathbb{S}_{it}\frac{\mathsf{P}(y_{itr})'}{\mathsf{P}(y_{itr})}\right)\right)\right] \\ &= \sum_{i=1}^{N}\left[\frac{1}{\mathsf{P}(y_i)}\left(\frac{1}{R}\sum_{r=1}^{R}\mathsf{P}(y_{ir})\left(\sum_{t=1}^{T}\mathbb{S}_{it}[\mathbb{D}_{i\kappa t} - \mathsf{P}(y_{i\kappa tr})]V_{i\kappa t}\right)\right)\right]\end{aligned}$$

2.D Health index

In order to explore whether an individual's general health has an impact on their labour status, we want to construct an objective measure of health. The survey-based LISS panel provides us with a subjective measure of overall health: Individuals are asked to asses their own health status in the *Health* core study. The answer-options provided are on a five-point Likert scale ranging from poor to excellent. Studies have shown, however, that own health assessment in surveys can be biased; see, e.g., Jürges (2007). Hence we do not use the self-assessed health status directly. Instead we make use of the richness of the health survey and generate a linear index based on an ordered probit regression with random effects: We regress the survey answers on dummy indicators for the individual's perceived change in health relative to the last year, BMI based indicators for under- and overweight, as well as a set of indicators for self-assessed difficulties with daily tasks, regularly taken medication, health problems and hospital visits in the past year, etc. We run the regression with a sample consisting only of individuals in our chosen age range that are also in the final regression sample. As a sensitivity check, we repeat this including observations of those without a sequence (but still within the age range). The results are robust to the sample chosen, and we choose to use the health index based on the larger sample.

The health survey was not asked in 2014. We also miss some years for some individuals in the sample. We fill in these gaps but take a different approach than with the factor markers. We assume that health follows a dynamic process. We regress the health index, as an AR(1) process, on its lag and control for unobserved heterogeneity. We then take the estimated coefficients to calculate missing values. Because health may be endogenous to labour market state choices, we will use the one period lag of the health index in our model.

Table 2.D.1: Regression results for health index

	Regression sample only Health in general	incl. single observations Health in general
Health compared to $t-1$: poor	-1.555***	-1.536***
	(0.0692)	(0.0674)
— moderate	-0.585***	-0.581***
	(0.0238)	(0.0231)
— very good	-0.0120	-0.0166
	(0.0225)	(0.0217)
— excellent	0.455***	0.421***
	(0.0443)	(0.0421)
Hinder in daily life (index)	-0.290***	-0.282***
	(0.0114)	(0.0110)
Long-standing disease	0.264***	0.279***
	(0.0210)	(0.0203)
Regularly pain in joints	-0.101***	-0.107***
	(0.0200)	(0.0193)
— Hearth problems	-0.185***	-0.194***
	(0.0471)	(0.0456)
— Breathing problems	-0.166***	-0.163***
	(0.0362)	(0.0350)
— Coughing, flu, etc.	-0.208***	-0.194***
	(0.0217)	(0.0210)
— Stomach/intestinal problems	-0.207***	-0.210***
	(0.0249)	(0.0241)
— Headaches	-0.107***	-0.113***
	(0.0209)	(0.0202)
— Fatigue	-0.253***	-0.245***
	(0.0198)	(0.0191)
— Sleeping problems	-0.0814***	-0.0864***
	(0.0219)	(0.0213)
Other recurrent complaints	-0.249***	-0.244***
	(0.0259)	(0.0251)
No recurrent complaints	0.119***	0.113***
	(0.0231)	(0.0223)
Taking medicine for:		
— high blood pressure	-0.123***	-0.131***

Table 2.D.1: Regression results for health index (continued)

	Regression sample only Health in general	incl. single observations Health in general
	(0.0291)	(0.0283)
— diabetes	-0.311***	-0.327***
	(0.0507)	(0.0495)
— joint pain/infection	-0.166***	-0.173***
	(0.0355)	(0.0345)
— hormonal osteoporosis	-0.371***	-0.325**
	(0.137)	(0.135)
Not taking any medicine	0.231***	0.214***
	(0.0252)	(0.0243)
Hospital stay	-0.0646**	-0.0749***
	(0.0256)	(0.0248)
Female	0.0583***	0.0653***
	(0.0155)	(0.0151)
Age	-0.00781***	-0.00756***
	(0.000829)	(0.000798)
Underweight	-0.140**	-0.0958*
	(0.0567)	(0.0542)
Overweight	-0.162***	-0.170***
	(0.0163)	(0.0158)
Obese	-0.343***	-0.342***
	(0.0233)	(0.0226)
Constant cut1	-4.397***	-4.305***
	(0.123)	(0.120)
Constant cut2	-1.964***	-1.888***
	(0.115)	(0.112)
Constant cut3	0.725***	0.778***
	(0.114)	(0.112)
Constant cut4	1.960***	2.013***
	(0.115)	(0.112)
Observations	28926	30669

Including year fixed effects and controls for depression & anxiety, difficulty with actions, other medication, and eduction. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

2.E Additional tables

Table 2.E.1: Observed transition probabilities (in %) excluding/including corrections

Labour state	Excluding corrections				Including corrections			
past \ current	0	1	2	3	0	1	2	3
0: employee	93.34	1.35	1.07	4.24	93.06	1.55	1.16	4.22
1: self-employed	7.93	87.76	0.65	3.66	8.68	87.28	0.60	3.44
2: unemployed	18.09	2.30	46.38	33.22	21.01	2.07	43.64	33.28
3: not in labour force	16.80	1.94	6.17	75.09	16.96	1.89	6.34	74.81
overall	69.92	10.30	3.03	16.75	69.94	10.75	3.03	16.29

Based on pooled samples with 24061 and 265510 observation pairs respectively (n=6019).

Source: LISS Panel, own calculations.

Table 2.E.2: Simulated transition probabilities (in %), pooled

Labour state	without Heckman correction				with Heckman correction			
past \ current	0	1	2	3	0	1	2	3
0: employee	92.43	1.57	1.40	4.59	91.45	1.56	1.49	5.49
1: self-employed	7.32	88.55	0.91	3.22	6.96	87.96	1.00	4.08
2: unemployed	21.56	1.88	43.13	33.43	20.77	1.64	40.29	37.29
3: not in labour force	16.92	2.06	6.95	74.06	16.3	2.03	6.25	75.42
Total	69.32	10.52	3.49	16.67	66	10.20	3.56	20.24

Based on dynamic model with Big-Five factor markers, 34448 pairs (n=6019).

Source: LISS Panel, missing values for personal/household characteristics are extrapolated.

Table 2.E.3: Personal characteristics for individual simulation

	Employee		**Self-employed**					
	Female	Male	Female			Male		
	median	median	median	high	low	median	high	low
Age in 2008	45	45	45			45		
Partner	Yes	Yes	Yes			Yes		
Nr. of children	1	1	1			1		
Medium education	Yes	Yes	Yes			No		
High education	No	No	No			Yes		
F1	0.0018	0.0913	0.3166	2.5	-2.5	0.2073	2.5	-2.5
F2	0.4373	-0.5206	0.2526	-2.5	2.5	-0.4403	-2.5	2.5
F3	0.0217	0.1470	0.9095	2.5	-2.5	-0.1899	2.5	-2.5
F4	0.0722	0.3549	-0.3043	2.5	-2.5	0.5378	2.5	-2.5
F5	-0.3563	-0.2132	-0.3008	2.5	-2.5	1.1217	2.5	-2.5

Note: “high” denotes the “entrepreneurial” and “low” the “non-entrepreneurial” benchmark individual.

Table 2.E.4: First stage Heckman regression results, women

Model	Baseline	Health index	EP distance	Big5 Factors
Self-employed in $t-1$	-0.14**	-0.11*	-0.13**	-0.11*
	(0.048)	(0.047)	(0.047)	(0.047)
Unemployed in $t-1$	0.10	0.12	0.10	0.10
	(0.088)	(0.088)	(0.087)	(0.087)
Not in labour force in $t-1$	0.07	0.08*	0.07	0.07*
	(0.037)	(0.037)	(0.037)	(0.037)
Age	0.02***	0.02***	0.02***	0.01***
	(0.002)	(0.001)	(0.001)	(0.002)
Has partner	-0.02	-0.03	-0.02	-0.05
	(0.041)	(0.040)	(0.040)	(0.040)
Has child	-0.04	-0.06	-0.03	-0.03
	(0.050)	(0.049)	(0.049)	(0.049)
Middle eduction	0.07	0.05	0.03	0.04
	(0.081)	(0.079)	(0.080)	(0.080)
High education	0.08	0.05	0.03	0.06
	(0.084)	(0.081)	(0.082)	(0.084)
Household size	-0.02	0.00	-0.01	-0.00
	(0.021)	(0.021)	(0.020)	(0.020)
Days until answered	-0.01***	-0.01***	-0.01***	-0.01***
(within call)	(0.002)	(0.002)	(0.002)	(0.002)
Answered in first call	0.32***	0.31***	0.31***	0.30***
	(0.048)	(0.047)	(0.047)	(0.047)
Interaction term	-0.01***	-0.01**	-0.01**	-0.01**
days x first call	(0.003)	(0.003)	(0.003)	(0.003)
F1: extraversion				-0.01
				(0.016)
F2: agreeableness				-0.01
				(0.018)
F3: conscientiousness				0.12***
				(0.016)
F4: emotional stability				-0.04**
				(0.016)
F5: openness for experience				-0.04*
				(0.018)
Health index		0.01		
		(0.013)		
EP distance			0.00	
			(0.003)	
Constant	0.39***	0.46***	0.46***	0.47***
	(0.117)	(0.113)	(0.128)	(0.116)
Observations	16,868	16,605	16,679	16,679
Number of nomem_encr	4,045	3,848	3,869	3,869
ρ	0.145	0.0870	0.102	0.0979
σ_u	0.412	0.309	0.336	0.329
LL	-6559	-6252	-6336	-6307

Dependent variable: Indicator variable for labour state observed in t.
Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 2.E.5: First stage Heckman regression results, men

Model	Baseline	Health index	EP distance	Big5 Factors
Self-employed in $t-1$	-0.23***	-0.19***	-0.21***	-0.19***
	(0.048)	(0.047)	(0.047)	(0.048)
Unemployed in $t-1$	-0.12	-0.05	-0.09	-0.08
	(0.111)	(0.113)	(0.110)	(0.111)
Not in labour force in $t-1$	0.12*	0.13*	0.15*	0.16**
	(0.063)	(0.062)	(0.063)	(0.063)
Age	0.02***	0.02***	0.02***	0.02***
	(0.002)	(0.002)	(0.002)	(0.002)
Has partner	0.05	0.03	0.06	0.05
	(0.051)	(0.050)	(0.050)	(0.050)
Has child	-0.00	-0.01	0.00	0.00
	(0.064)	(0.062)	(0.062)	(0.062)
Middle eduction	0.21*	0.20*	0.22*	0.20*
	(0.089)	(0.086)	(0.087)	(0.087)
High education	0.27**	0.23**	0.26**	0.25**
	(0.091)	(0.089)	(0.089)	(0.091)
Household size	-0.06*	-0.05	-0.05*	-0.05
	(0.026)	(0.025)	(0.025)	(0.025)
Days until answered	-0.01***	-0.01***	-0.01***	-0.01***
(within call)	(0.003)	(0.003)	(0.003)	(0.003)
Answered in first call	0.19***	0.21***	0.17**	0.17**
	(0.056)	(0.055)	(0.056)	(0.056)
Interaction term	-0.00	-0.00	-0.00	-0.00
days x first call	(0.004)	(0.004)	(0.004)	(0.004)
F1: extraversion				-0.01
				(0.019)
F2: agreeableness				-0.01
				(0.019)
F3: conscientiousness				0.06**
				(0.019)
F4: emotional stability				0.02
				(0.019)
F5: openness for experience				-0.01
				(0.021)
Health index		0.03		
		(0.016)		
EP distance			-0.01*	
			(0.003)	
Constant	0.23*	0.33**	0.43***	0.34***
	(0.131)	(0.127)	(0.142)	(0.131)
Observations	13,824	13,578	13,675	13,675
Number of nomem_encr	3,242	3,059	3,104	3,104
ρ	0.203	0.137	0.158	0.160
σ_u	0.505	0.398	0.433	0.436
LL	-5046	-4766	-4865	-4860

Dependent variable: Indicator variable for labour state observed in t.
Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 2.E.6: First stage Heckman regression results, pooled

Model	Baseline	Health index	EP distance	Big5 Factors
Self-employed in $t-1$	-0.18***	-0.15***	-0.16***	-0.15***
	(0.034)	(0.033)	(0.033)	(0.033)
Unemployed in $t-1$	0.02	0.06	0.03	0.04
	(0.069)	(0.069)	(0.068)	(0.068)
Not in labour force in $t-1$	0.08**	0.09**	0.09**	0.10**
	(0.032)	(0.032)	(0.032)	(0.032)
Age	0.02***	0.02***	0.02***	0.02***
	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.05	-0.08	-0.04	-0.08
	(0.047)	(0.045)	(0.046)	(0.046)
Has partner	0.03	0.00	0.04	0.03
	(0.046)	(0.045)	(0.045)	(0.045)
Has child	-0.05	-0.06	-0.04	-0.05
	(0.049)	(0.048)	(0.048)	(0.048)
Female x partner	-0.04	-0.02	-0.05	-0.05
	(0.057)	(0.054)	(0.055)	(0.055)
Female x has child	0.03	0.04	0.04	0.05
	(0.049)	(0.047)	(0.047)	(0.047)
Middle eduction	0.14*	0.12*	0.12*	0.12*
	(0.060)	(0.058)	(0.059)	(0.059)
High education	0.17**	0.14*	0.14*	0.15*
	(0.061)	(0.060)	(0.060)	(0.061)
Household size	-0.03	-0.02	-0.03	-0.02
	(0.016)	(0.016)	(0.016)	(0.016)
Days until answered	-0.01***	-0.01***	-0.01***	-0.01***
(within call)	(0.002)	(0.002)	(0.002)	(0.002)
Answered in first call	0.26***	0.27***	0.25***	0.25***
	(0.036)	(0.036)	(0.036)	(0.036)
Interaction term	-0.01**	-0.01**	-0.01**	-0.01**
days x first call	(0.002)	(0.002)	(0.002)	(0.002)
F1: extraversion				-0.01
				(0.012)
F2: agreeableness				-0.00
				(0.013)
F3: conscientiousness				0.09***
				(0.013)
F4: emotional stability				-0.02
				(0.012)
F5: openness for experience				-0.03*
				(0.013)
Health index		0.01		
		(0.010)		
EP distance			-0.00	
			(0.002)	
Constant	0.34***	0.44***	0.46***	0.47***
	(0.090)	(0.087)	(0.096)	(0.089)
Observations	30,692	30,183	30,354	30,354
Number of individuals	7,287	6,907	6,973	6,973
ρ	0.170	0.107	0.126	0.124
σ_u	0.452	0.346	0.379	0.376
LL	-11614	-11026	-11212	-11183

Dependent variable: Indicator variable for labour state observed in t.
Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 2.E.7: Static model with Big-Five factor markers, pooled

	Self-employed	Unemployed	Not in Labour Force
Constant	-19.65***	-7.79***	-4.15***
	(0.9777)	(0.8237)	(0.628)
Age	0.17***	0.09***	0.09***
	(0.0124)	(0.0103)	(0.0085)
Female	-1.22***	0.61**	0.63**
	(0.4002)	(0.2832)	(0.2654)
Has partner	0.21	-0.81***	-1.04***
	(0.3082)	(0.2764)	(0.2508)
Has child	-1.55***	-0.96***	-1.06***
	(0.2993)	(0.292)	(0.2435)
Female x partner	-0.39	0.15	1.77***
	(0.3974)	(0.3123)	(0.2928)
Female x has child	1.26***	0.84***	0.49*
	(0.3829)	(0.3041)	(0.2633)
Middle education	0.01	-1.43***	-1.98***
	(0.5194)	(0.3255)	(0.263)
High education	0.95*	-2.8***	-3.43***
	(0.5407)	(0.3519)	(0.2905)
Household size	0.37***	0.16	0.3***
	(0.1122)	(0.1027)	(0.0756)
F1: extraversion	0.26**	-0.09	-0.06
	(0.1113)	(0.0908)	(0.0684)
F2: agreeableness	-0.24**	0.12	0.07
	(0.0995)	(0.0807)	(0.0651)
F3: conscientiousness	0.14	-0.35***	-0.54***
	(0.1048)	(0.0903)	(0.0705)
F4: emotional stability	-0.31***	-0.59***	-0.5***
	(0.0976)	(0.0826)	(0.0628)
F5: openness for experience	0.5***	0.33***	0.21***
	(0.0989)	(0.086)	(0.0691)
Inverse Mills Ratio	15.53***	-4.13***	-8.04***
	(0.9191)	(1.1389)	(0.7769)
L	6.95***		
	(0.2621)		
	2.16***	3.37***	
	(0.1778)	(0.143)	
	2.47***	3.33***	1.69***
	(0.1594)	(0.1156)	(0.0826)
Covariance $W = LL^{\intercal}$	48.3517	14.9999	17.1594
	14.9999	15.9799	16.5171
	17.1594	16.5171	20.0146
Observations: 26402			
Nr. of Individuals: 5914			
Loglikelihood: -12989.71			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.8: Dynamic model with Big-Five factor markers, women

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.72***	-6.12***	-4.5***
	(1.0592)	(0.9536)	(0.6593)
Age	0.01	0.03***	0.04***
	(0.0111)	(0.0106)	(0.0078)
Has partner	0.3	-0.65***	0.38**
	(0.2423)	(0.1892)	(0.1699)
Has child	-0.34	-0.22	-0.41**
	(0.3001)	(0.2332)	(0.1926)
Middle education	-0.12	-0.81**	-0.92***
	(0.5824)	(0.3518)	(0.2951)
High education	0.18	-1.55***	-1.44***
	(0.6075)	(0.3824)	(0.3139)
Household size	0.13	0.05	0.06
	(0.1203)	(0.0984)	(0.0816)
F1: extraversion	0.03	-0.06	-0.03
	(0.0974)	(0.0899)	(0.0667)
F2: agreeableness	-0.07	0.04	0.01
	(0.1112)	(0.0916)	(0.0701)
F3: conscientiousness	-0.07	-0.18*	-0.23***
	(0.1161)	(0.1021)	(0.0776)
F4: emotional stability	-0.09	-0.34***	-0.19***
	(0.0926)	(0.0853)	(0.0627)
F5: openness for experience	0.26**	0.25***	0.14*
	(0.1032)	(0.0913)	(0.0728)
Last state: self-employed	4.02***	1.11**	1.21***
	(0.2397)	(0.4859)	(0.2955)
Last state: unemployed	0.11	2.92***	1.69***
	(0.5094)	(0.2832)	(0.218)
Last state: not in LF	0.6**	1.54***	2.09***
	(0.2675)	(0.2003)	(0.1068)
Initial state: self-employed	4.93***	1.42**	2.3***
	(0.5339)	(0.5624)	(0.362)
Initial state: unemployed	3.48***	3.35***	3.18***
	(0.6971)	(0.5318)	(0.4812)
Initial state: not in LF	2.85***	3.18***	4.28***
	(0.3806)	(0.2937)	(0.2389)
Inverse Mills Ratio	-0.59	-0.9	-2.05**
	(1.2873)	(1.3931)	(1.002)
L	2.26***		
	(0.2128)		
	1.13***	1.26***	
	(0.2295)	(0.229)	
	1.26***	1.43***	0.63***
	(0.1802)	(0.1649)	(0.1711)
Covariance $W = LL^{\intercal}$	5.1287	2.5575	2.8429
	2.5575	2.8603	3.2239
	2.8429	3.2239	4.0307
Observations: 14435			
Nr. of Individuals: 3267			
Loglikelihood: -6086.96			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.9: Dynamic model with Big-Five factor markers, men

	Self-employed	Unemployed	Not in Labour Force
Constant	-7***	-7.74***	-5.11***
	(1.2323)	(1.4295)	(0.9316)
Age	0.03*	0.07***	0.06***
	(0.0139)	(0.0166)	(0.0115)
Has partner	-0.08	-0.36	-0.34
	(0.2575)	(0.3142)	(0.2256)
Has child	-0.28	-0.08	-0.12
	(0.3137)	(0.4079)	(0.2754)
Middle education	-0.07	-0.75*	-1.26***
	(0.5538)	(0.4333)	(0.3313)
High education	0.29	-1.53***	-2.15***
	(0.5674)	(0.492)	(0.3739)
Household size	0.18	-0.07	-0.03
	(0.1304)	(0.185)	(0.1285)
F1: extraversion	0.28**	-0.03	-0.03
	(0.1105)	(0.1276)	(0.0855)
F2: agreeableness	-0.15	0.03	-0.04
	(0.093)	(0.1227)	(0.0822)
F3: conscientiousness	-0.19*	-0.16	-0.19**
	(0.1)	(0.1281)	(0.092)
F4: emotional stability	-0.25**	-0.33***	-0.37***
	(0.104)	(0.116)	(0.0827)
F5: openness for experience	0.24**	0.15	0.08
	(0.1036)	(0.1209)	(0.0848)
Last state: self-employed	4.15***	0.48	0.86**
	(0.2447)	(0.7504)	(0.4071)
Last state: unemployed	0.23	2.69***	1.3***
	(0.5791)	(0.3341)	(0.312)
Last state: not in LF	0.57*	0.59*	1.5***
	(0.3311)	(0.3118)	(0.1862)
Initial state: self-employed	4.75***	2.7***	2.29***
	(0.5798)	(0.7324)	(0.4488)
Initial state: unemployed	2.61***	4.14***	3.76***
	(0.9214)	(0.5955)	(0.5501)
Initial state: not in LF	2.1***	4.03***	4.3***
	(0.5243)	(0.4125)	(0.3502)
Inverse Mills Ratio	0.27	-1.91	-2.1
	(1.4686)	(2.2414)	(1.3447)
L	2.1***		
	(0.2382)		
	1.49***	1.25***	
	(0.3318)	(0.3344)	
	1.18***	1.6***	0.24
	(0.2575)	(0.2017)	(0.4127)
Covariance $W = LL^{\intercal}$	4.4261	3.1269	2.4758
	3.1269	3.7779	3.7523
	2.4758	3.7523	4.0018
Observations: 11967			
Nr. of Individuals: 2647			
Loglikelihood: -4066.08			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.10: Dynamic model with Big-Five factor markers, pooled

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.81***	-7.17***	-4.89***
	(0.7856)	(0.7682)	(0.5285)
Age	0.02*	0.05***	0.05***
	(0.0084)	(0.0087)	(0.0063)
Female	-0.04	0.48**	0.33*
	(0.262)	(0.2256)	(0.1885)
Has partner	0.05	-0.35	-0.37**
	(0.2356)	(0.2375)	(0.1859)
Has child	-0.25	-0.28	-0.23
	(0.2619)	(0.2505)	(0.1949)
Female x partner	0.22	-0.3	0.73***
	(0.3025)	(0.268)	(0.2181)
Female x has child	-0.16	0.19	-0.1
	(0.2704)	(0.2548)	(0.1969)
Middle education	-0.13	-0.79***	-1.12***
	(0.3972)	(0.2647)	(0.2138)
High education	0.21	-1.53***	-1.78***
	(0.4078)	(0.2901)	(0.2322)
Household size	0.17*	0	0.03
	(0.0876)	(0.0846)	(0.0663)
F1: extraversion	0.15**	-0.04	-0.03
	(0.0728)	(0.0714)	(0.0515)
F2: agreeableness	-0.11	0.04	-0.02
	(0.0708)	(0.0695)	(0.0502)
F3: conscientiousness	-0.14*	-0.15*	-0.21***
	(0.077)	(0.0772)	(0.0577)
F4: emotional stability	-0.17***	-0.36***	-0.28***
	(0.0655)	(0.0652)	(0.0481)
F5: openness for experience	0.26***	0.19***	0.11**
	(0.0711)	(0.0716)	(0.0538)
Last state: self-employed	4.06***	0.93**	1.14***
	(0.1651)	(0.3819)	(0.2277)
Last state: unemployed	0.3	2.75***	1.55***
	(0.3483)	(0.2023)	(0.1664)
Last state: not in LF	0.63***	1.19***	1.89***
	(0.1985)	(0.1604)	(0.0885)
Initial state: self-employed	4.95***	1.76***	2.16***
	(0.3848)	(0.3905)	(0.2708)
Initial state: unemployed	2.92***	3.78***	3.44***
	(0.5375)	(0.3869)	(0.3484)
Initial state: not in LF	2.51***	3.53***	4.27***
	(0.2842)	(0.23)	(0.1894)
Inverse Mills Ratio	-0.36	-1.03	-2.06***
	(0.96)	(1.1517)	(0.7955)
L	2.2***		
	(0.1536)		
	1.11***	1.48***	
	(0.1756)	(0.147)	
	1.1***	1.52***	0.65***
	(0.1372)	(0.1216)	(0.1206)
Covariance $W = LL^{\intercal}$	4.8371	2.4439	2.4218
	2.4439	3.4189	3.4722
	2.4218	3.4722	3.9505
Observations:	26402		
Nr. of Individuals:	5914		
Loglikelihood:	-10201.11		

Regression incl. year fixed effects; (Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.11: Dynamic model with Big-Five factor markers and no Heckman correction, pooled

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.98***	-7.67***	-5.89***
	(0.5708)	(0.5313)	(0.3637)
Age	0.02**	0.05***	0.06***
	(0.0067)	(0.0066)	(0.0049)
Female	-0.05	0.46**	0.28
	(0.2604)	(0.223)	(0.1864)
Has partner	0.05	-0.34	-0.35*
	(0.2342)	(0.2374)	(0.1855)
Has child	-0.26	-0.29	-0.26
	(0.2604)	(0.2499)	(0.1942)
Female x partner	0.21	-0.32	0.7***
	(0.3007)	(0.2678)	(0.2174)
Female x has child	-0.16	0.2	-0.07
	(0.2689)	(0.2547)	(0.1959)
Middle education	-0.12	-0.75***	-1.04***
	(0.3894)	(0.2615)	(0.2114)
High education	0.23	-1.47***	-1.69***
	(0.3974)	(0.2856)	(0.2292)
Household size	0.17*	-0.01	0.01
	(0.0873)	(0.0832)	(0.0657)
F1: extraversion	0.15**	-0.05	-0.03
	(0.0724)	(0.0713)	(0.0514)
F2: agreeableness	-0.11	0.04	-0.02
	(0.0707)	(0.0692)	(0.0501)
F3: conscientiousness	-0.13*	-0.11*	-0.14***
	(0.0692)	(0.0659)	(0.0511)
F4: emotional stability	-0.17***	-0.37***	-0.29***
	(0.0651)	(0.0647)	(0.0477)
F5: openness for experience	0.26***	0.18***	0.09*
	(0.0704)	(0.0705)	(0.0533)
Last state: self-employed	4.04***	0.86**	1.04***
	(0.1567)	(0.3751)	(0.2224)
Last state: unemployed	0.29	2.77***	1.58***
	(0.3482)	(0.2017)	(0.166)
Last state: not in LF	0.64***	1.22***	1.95***
	(0.1961)	(0.1567)	(0.085)
Initial state: self-employed	4.95***	1.77***	2.16***
	(0.3845)	(0.3913)	(0.2694)
Initial state: unemployed	2.91***	3.77***	3.43***
	(0.5352)	(0.3865)	(0.3468)
Initial state: not in LF	2.5***	3.52***	4.26***
	(0.2837)	(0.2298)	(0.1888)
L	2.2***		
	(0.1535)		
	1.12***	1.47***	
	(0.1759)	(0.1482)	
	1.1***	1.52***	0.66***
	(0.1371)	(0.1221)	(0.1204)
Covariance $W = LL^{\intercal}$	4.8245	2.4551	2.4063
	2.4551	3.4135	3.4555
	2.4063	3.4555	3.9316
Observations:	26402		
Nr. of Individuals:	5914		
Loglikelihood:	-10204.73		

Regression incl. year fixed effects; (Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.12: Dynamic model with EP distance, pooled

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.4***	-7.59***	-5.26***
	(0.8093)	(0.7696)	(0.5423)
Age	0.01	0.04***	0.04***
	(0.0084)	(0.0086)	(0.0063)
Female	-0.07	0.42*	0.23
	(0.2531)	(0.2191)	(0.1849)
Has partner	0.02	-0.43*	-0.45**
	(0.2352)	(0.2338)	(0.1843)
Has child	-0.26	-0.25	-0.21
	(0.2613)	(0.2499)	(0.1942)
Female x partner	0.19	-0.24	0.76***
	(0.3007)	(0.2651)	(0.2175)
Female x has child	-0.21	0.16	-0.12
	(0.2694)	(0.253)	(0.1961)
Middle education	-0.15	-0.8***	-1.16***
	(0.3865)	(0.2634)	(0.2141)
High education	0.31	-1.45***	-1.79***
	(0.391)	(0.2839)	(0.2275)
Household size	0.17*	-0.01	0.03
	(0.0869)	(0.0847)	(0.0663)
EP distance	-0.01	0.04***	0.04***
	(0.0107)	(0.01)	(0.0077)
Last state: self-employed	4.05***	0.93**	1.15***
	(0.1656)	(0.3738)	(0.227)
Last state: unemployed	0.32	2.78***	1.54***
	(0.3499)	(0.2008)	(0.1665)
Last state: not in LF	0.63***	1.18***	1.89***
	(0.1961)	(0.1592)	(0.0878)
Initial state: self-employed	5.03***	1.81***	2.21***
	(0.3845)	(0.3856)	(0.2715)
Initial state: unemployed	3.03***	3.87***	3.52***
	(0.5154)	(0.391)	(0.3535)
Initial state: not in LF	2.6***	3.63***	4.35***
	(0.2821)	(0.2321)	(0.191)
Inverse Mills Ratio	-0.16	-1.06	-2.07***
	(0.9515)	(1.1191)	(0.784)
L	2.21***		
	(0.1536)		
	1.14***	1.49***	
	(0.176)	(0.1484)	
	1.13***	1.56***	0.6***
	(0.1338)	(0.1172)	(0.1222)
Covariance $W = LL^{\intercal}$	4.9029	2.5242	2.5039
	2.5242	3.5324	3.6228
	2.5039	3.6228	4.0719
Observations: 26402			
Nr. of Individuals: 5914			
Loglikelihood: -10235.85			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.13: Dynamic model with EP distance, women

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.56***	-6.67***	-4.93***
	(1.1146)	(0.955)	(0.6893)
Age	0.01	0.02**	0.04***
	(0.011)	(0.0104)	(0.0077)
Has partner	0.26	-0.68***	0.33**
	(0.2409)	(0.188)	(0.1683)
Has child	-0.36	-0.22	-0.42**
	(0.295)	(0.2315)	(0.1916)
Middle education	-0.1	-0.79**	-0.93***
	(0.5728)	(0.3503)	(0.2922)
High education	0.31	-1.41***	-1.39***
	(0.5916)	(0.3738)	(0.3054)
Household size	0.13	0.02	0.06
	(0.1194)	(0.0984)	(0.0814)
EP distance	-0.01	0.04***	0.03***
	(0.0149)	(0.0121)	(0.0096)
Last state: self-employed	4.05***	1.11**	1.24***
	(0.2398)	(0.4806)	(0.2958)
Last state: unemployed	0.13	2.93***	1.67***
	(0.5064)	(0.2805)	(0.2167)
Last state: not in LF	0.62**	1.53***	2.09***
	(0.2641)	(0.198)	(0.1058)
Initial state: self-employed	4.88***	1.39**	2.29***
	(0.5319)	(0.5559)	(0.3634)
Initial state: unemployed	3.49***	3.45***	3.26***
	(0.6774)	(0.5306)	(0.4821)
Initial state: not in LF	2.84***	3.29***	4.34***
	(0.3742)	(0.2959)	(0.2398)
Inverse Mills Ratio	-0.35	-0.68	-2**
	(1.2709)	(1.3529)	(0.9855)
L	2.25***		
	(0.2135)		
	1.16***	1.31***	
	(0.2288)	(0.2228)	
	1.26***	1.48***	0.58***
	(0.1822)	(0.1639)	(0.1757)
Covariance $W = LL^{\intercal}$	5.0511	2.6181	2.8361
	2.6181	3.0796	3.4134
	2.8361	3.4134	4.1256
Observations: 14435			
Nr. of Individuals: 3267			
Loglikelihood: -6103.72			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.14: Dynamic model with EP distance, men

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.41***	-8.03***	-5.5***
	(1.2027)	(1.4117)	(0.9209)
Age	0.02	0.06***	0.05***
	(0.0138)	(0.0163)	(0.0114)
Has partner	-0.14	-0.45	-0.43*
	(0.2534)	(0.309)	(0.2221)
Has child	-0.31	-0.1	-0.1
	(0.308)	(0.4019)	(0.2749)
Middle education	-0.13	-0.79*	-1.34***
	(0.542)	(0.4341)	(0.3337)
High education	0.3	-1.53***	-2.22***
	(0.5467)	(0.485)	(0.3705)
Household size	0.2	-0.05	-0.03
	(0.1243)	(0.1839)	(0.1286)
EP distance	-0.02	0.05**	0.05***
	(0.0164)	(0.0202)	(0.0143)
Last state: self-employed	4.13***	0.52	0.89**
	(0.2476)	(0.7188)	(0.4014)
Last state: unemployed	0.26	2.71***	1.34***
	(0.5869)	(0.3345)	(0.3126)
Last state: not in LF	0.56*	0.59*	1.49***
	(0.3306)	(0.3079)	(0.185)
Initial state: self-employed	4.88***	2.76***	2.37***
	(0.5861)	(0.7407)	(0.4359)
Initial state: unemployed	2.71***	4.22***	3.81***
	(0.8558)	(0.5976)	(0.5513)
Initial state: not in LF	2.24***	4.1***	4.4***
	(0.5159)	(0.4169)	(0.3523)
Inverse Mills Ratio	0.4	-2.39	-2.45*
	(1.4459)	(2.1704)	(1.3444)
L	2.12***		
	(0.2385)		
	1.48***	1.26***	
	(0.3368)	(0.3401)	
	1.2***	1.57***	0.41
	(0.2376)	(0.2028)	(0.3464)
Covariance $W = LL^{\intercal}$	4.5066	3.1374	2.5555
	3.1374	3.7761	3.7657
	2.5555	3.7657	4.1003
Observations: 11967			
Nr. of Individuals: 2647			
Loglikelihood: -4086.93			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.15: Dynamic model with health index, pooled

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.59***	-7.15***	-4.7***
	(0.7895)	(0.7363)	(0.5164)
Age	0.01	0.04***	0.04***
	(0.0085)	(0.0083)	(0.0062)
Female	-0.18	0.38*	0.2
	(0.2633)	(0.2148)	(0.1829)
Has partner	0.08	-0.33	-0.38**
	(0.2354)	(0.2251)	(0.1789)
Has child	-0.21	-0.35	-0.3
	(0.2619)	(0.239)	(0.1901)
Female x partner	0.1	-0.25	0.74***
	(0.3062)	(0.2561)	(0.212)
Female x has child	-0.11	0.27	-0.03
	(0.2716)	(0.2429)	(0.1915)
Middle education	-0.03	-0.65**	-1.02***
	(0.4027)	(0.256)	(0.2116)
High education	0.48	-1.16***	-1.53***
	(0.4098)	(0.2767)	(0.2248)
Household size	0.17*	0.02	0.06
	(0.0881)	(0.0825)	(0.0665)
Health Index, lagged	-0.09	-0.35***	-0.31***
	(0.0538)	(0.0381)	(0.0333)
Last state: self-employed	4.02***	0.83**	1.14***
	(0.1637)	(0.369)	(0.2253)
Last state: unemployed	0.24	2.87***	1.5***
	(0.3559)	(0.1968)	(0.1669)
Last state: not in LF	0.67***	1.17***	1.92***
	(0.2016)	(0.1588)	(0.0877)
Initial state: self-employed	5.15***	1.95***	2.2***
	(0.3908)	(0.3749)	(0.2668)
Initial state: unemployed	2.88***	3.43***	3.24***
	(0.5244)	(0.3676)	(0.3439)
Initial state: not in LF	2.5***	3.42***	4.16***
	(0.2853)	(0.2266)	(0.1894)
Inverse Mills Ratio	-0.42	-0.47	-1.77**
	(0.9601)	(1.0979)	(0.7858)
L	2.23***		
	(0.1565)		
	1.16***	1.36***	
	(0.1688)	(0.1458)	
	1.06***	1.61***	-0.33*
	(0.14)	(0.1138)	(0.1987)
Covariance $W = LL^{\intercal}$	4.9878	2.5883	2.3697
	2.5883	3.181	3.4133
	2.3697	3.4133	3.8278
Observations: 26319			
Nr. of Individuals: 5866			
Loglikelihood: -10143.69			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.16: Dynamic model with health index, women

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.72***	-6.4***	-4.58***
	(1.0301)	(0.9)	(0.6442)
Age	0.01	0.02**	0.03***
	(0.0107)	(0.01)	(0.0076)
Has partner	0.28	-0.6***	0.38**
	(0.2386)	(0.1841)	(0.1686)
Has child	-0.27	-0.18	-0.39**
	(0.2992)	(0.2266)	(0.1924)
Middle education	-0.14	-0.63*	-0.79***
	(0.5481)	(0.341)	(0.2925)
High education	0.27	-1.14***	-1.15***
	(0.5643)	(0.3677)	(0.3071)
Household size	0.12	0.06	0.08
	(0.1221)	(0.0966)	(0.0828)
Health Index, lagged	-0.11	-0.37***	-0.3***
	(0.0687)	(0.0492)	(0.0451)
Last state: self-employed	4.03***	1.02**	1.21***
	(0.2416)	(0.477)	(0.2947)
Last state: unemployed	0.02	3.02***	1.63***
	(0.5083)	(0.2589)	(0.2162)
Last state: not in LF	0.56**	1.51***	2.07***
	(0.2706)	(0.1971)	(0.1058)
Initial state: self-employed	4.88***	1.56***	2.34***
	(0.5248)	(0.5445)	(0.3602)
Initial state: unemployed	3.51***	3.19***	3.12***
	(0.6573)	(0.4869)	(0.4818)
Initial state: not in LF	2.86***	3.11***	4.22***
	(0.3674)	(0.2854)	(0.2365)
Inverse Mills Ratio	-0.61	-0.29	-1.91**
	(1.2789)	(1.312)	(0.9636)
L	2.22***		
	(0.2066)		
	1.28***	1.09***	
	(0.204)	(0.2046)	
	1.3***	1.53***	0.26
	(0.1761)	(0.1551)	(0.3168)
Covariance $W = LL^{\intercal}$	4.9092	2.8448	2.8775
	2.8448	2.845	3.3432
	2.8775	3.3432	4.1
Observations: 14399			
Nr. of Individuals: 3248			
Loglikelihood: -6054.79			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

Table 2.E.17: Dynamic model with health index, men

	Self-employed	Unemployed	Not in Labour Force
Constant	-6.43***	-7.44***	-4.86***
	(1.2339)	(1.4121)	(0.9146)
Age	0.01	0.06***	0.05***
	(0.0138)	(0.0163)	(0.0113)
Has partner	-0.04	-0.38	-0.33
	(0.2483)	(0.3068)	(0.2179)
Has child	-0.29	-0.17	-0.16
	(0.3058)	(0.3943)	(0.2729)
Middle education	0.11	-0.73*	-1.26***
	(0.5824)	(0.4305)	(0.322)
High education	0.61	-1.29***	-2.01***
	(0.5893)	(0.4788)	(0.3591)
Household size	0.21*	-0.02	-0.01
	(0.1262)	(0.1811)	(0.1282)
Health Index, lagged	-0.07	-0.3***	-0.31***
	(0.0843)	(0.0664)	(0.0556)
Last state: self-employed	4.18***	0.54	0.95**
	(0.2419)	(0.7238)	(0.4074)
Last state: unemployed	0.42	2.68***	1.35***
	(0.5715)	(0.3352)	(0.311)
Last state: not in LF	0.73**	0.63**	1.55***
	(0.3211)	(0.3019)	(0.1798)
Initial state: self-employed	4.87***	2.5***	2.05***
	(0.5846)	(0.7152)	(0.4329)
Initial state: unemployed	2.32**	3.79***	3.38***
	(0.9402)	(0.5778)	(0.5303)
Initial state: not in LF	1.76***	3.85***	4.21***
	(0.5156)	(0.406)	(0.3476)
Inverse Mills Ratio	-0.3	-1.76	-1.61
	(1.4966)	(2.2136)	(1.3821)
L	2.09***		
	(0.2363)		
	1.23***	1.38***	
	(0.3575)	(0.2794)	
	0.89***	1.66***	0.44
	(0.2452)	(0.1836)	(0.3343)
Covariance $W = LL^{\intercal}$	4.3711	2.5745	1.8555
	2.5745	3.4159	3.3856
	1.8555	3.3856	3.7517
Observations: 11920			
Nr. of Individuals: 2618			
Loglikelihood: -4042.31			

Regression including year fixed effects.
(Non-robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.10

2.F Additional figures

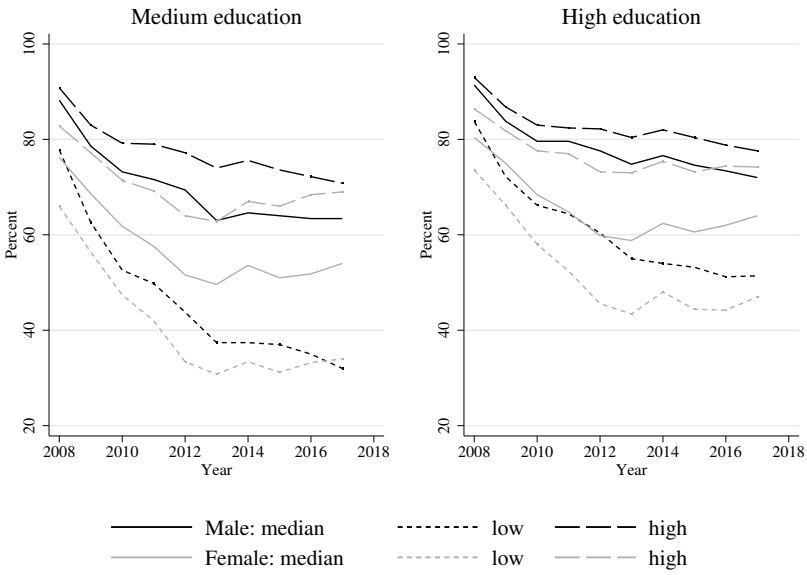


Figure 2.F.1: Share of employment paths spent in self-employment for different Big-Five factor markers (benchmark individual starts as self-employed in 2007)

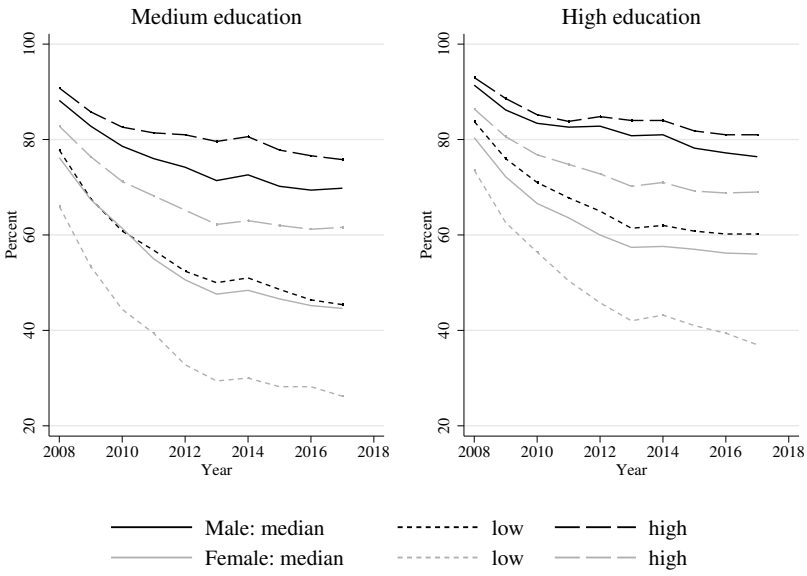


Figure 2.F.2: Share of employment paths spent in self-employment for different Big-Five factor markers (benchmark individual starts as self-employed in 2007, no year effects)

CHAPTER 3

LABOUR MARKET TRAJECTORIES OF THE SELF-EMPLOYED IN THE NETHERLANDS

This chapter is based on the identically entitled article published in De Economist, co-authored with Arthur van Soest

This paper employs sequence analysis to study the labour market trajectories of the self-employed. Using Dutch administrative data on more than 50,000 individuals including 13,000 with self-employment experience between 1989 and 2017, we find seven different clusters with distinct life-cycle patterns of several types of self-employment, wage employment, and non-employment. We find large heterogeneity across clusters in terms of income, wealth, and pension accumulation. In particular, the clusters of individuals with short self-employment spells but little labour market attachment in other periods are an economically vulnerable group, whereas those who are persistently self-employed are not worse off than employees.

3.1 Introduction

The self-employment rate in the Netherlands has risen substantially in recent years, from 11.6% in 2003 to 16.8% in 2016.[1] This increase has raised interest in the income vulnerability of the self-employed and their impact on the social security system. In particular, they are an important potentially vulnerable group when it comes to retirement income, since, like in many other countries, they are not obliged to accumulate the same occupational pensions that employees

[1] OECD data; see https://data.oecd.org/emp/self-employment-rate.htm.

accumulate.[2] Policy makers are therefore concerned that the self-employed are not building up enough pension wealth until their retirement (See SZW, 2016, section 3.1). Indeed, Mastrogiacomo (2016) shows that the self-employed have the same pension ambitions as employees but tend to fall short on their goals. Similar results are found by other studies on retirement preparedness of the Dutch; see, e.g., De Bresser and Knoef (2015) or Knoef et al. (2016).

Mastrogiacomo et al. (2016) emphasize the large heterogeneity in wealth holdings of the self-employed. This raises the question how to characterize them, dividing them, for example, into groups that differ distinctly in their income, wealth or pension entitlements. Population data published by Statistics Netherlands show that the rise in the number of self-employed is due to the growth in solo self-employed (SSE), self-employed that work on their own and have no employees in their business. Earlier studies like Bosch and Van Vuuren (2010) or SER (2010) found that the SSE are themselves quite heterogeneous in terms of personal characteristics and background. This is in line with the international literature on entrepreneurship (Blanchflower, 2000; Parker, 2004). So far only few studies have differentiated between SSE and other self-employed. An exception is Mastrogiacomo et al. (2016) who show that heterogeneity in business wealth of the self-employed is mostly driven by SSE, while entrepreneurs who own a larger firm are more homogeneous. Zwinkels et al. (2017) focus on the heterogeneity of SSE in particular: the variation in their calculated pension wealth is larger than for employees. Given the heterogeneity of SSE it is questionable if a bisection of the self-employed into SSE and other self-employed individuals is sufficient to understand the heterogeneity in how the self-employed prepare for retirement. Bolhaar et al. (2016) split the SSE into two groups based on their business activity but do not find many differences. Zwinkels et al. classify the SSE into several groups based on income sources and household composition. They find that households consisting of only SSE in particular will often have a low replacement rate after retirement.

Unlike most of the existing studies, the current paper analyses individuals labour market trajectories over a large number of years. It uses sequence analysis to construct different clusters of self-employment based on these trajectories. Sequence analysis was introduced in sociology by Abbott (1983) and became a popular way to study e.g. life-course and career trajectories in the social sciences. It is particularly well suited for the analysis of self-employment and pensions, since pension savings are typically determined by individuals' complete labour market trajectories. We show for example that a large share of the self-employed remain in self-employment for only a few years, while studies on pension preparedness of the Dutch often assume that individuals remain in the same labour market state going forward. The only exception we are aware of is Mastrogiacomo et al. (2016) who distinguish long-term from short-term entrepreneurs.

Some related studies also use sequence analysis. Zacher et al. (2012) use data from the German Socio-Economic Panel and find that compared to those who are only self-employed for a short time, individuals who remain continuously in self-employment are more likely to be

[2]See online Appendix 3.D for more information on the institutional background of the Dutch pension system.

male, from older cohorts, and have a higher risk taking propensity. Humphries (2016) studies the 1970 cohort of Swedish men and distinguishes between incorporated and unincorporated self-employment. He finds clusters that differ both in income and wealth accumulation, in line with Levine and Rubinstein (2017) who found that in the US, the legal form chosen by self-employed explains most of the variation in income. Visser (2018) uses sequence analysis to investigate whether careers in the Netherlands have become more complex for individuals born in the early 1940s compared to those born in the 30s. Munnell et al. (2019) follow individuals in the US from age 50 to 62 to assess how they use non-traditional jobs. They find that individuals who are more frequently employed in non-traditional work during their late career are more likely to have lower retirement income and higher rates of depression. Tophoven and Tisch (2016) analyse the implications of work trajectories for accrued statutory pension entitlements by the age of 42 for different cohorts in Germany. Their study does not offer any insights for the self-employed, since they are not included in the statutory pension system. Lastly, Madero-Cabib and Fasang (2016) study cohorts born in 1920–1950 in Germany and Switzerland to examine how work and family life affect financial well-being in retirement. Analysing labour market and family trajectories jointly, they find that breadwinner policies (i.e. supporting a system where men do and women do not do paid work) in combination with liberal pension policies later in life, as in Switzerland, intensify pension penalties for typical female work-family life courses.

Our study is descriptive and does not aim at identifying causal mechanisms. Our main aim is to analyse the heterogeneity of labour market careers involving self-employment. In order to partition individuals into distinct groups, we use optimal matching to first calculate dissimilarity measures for all observed employment paths. In a second step we cluster the individuals based on the dissimilarity measures. We then study the characteristics of the clusters, i.e. their income, wealth, and pension investments. Like other studies on the Dutch self-employed, we use data from Statistics Netherlands' Income Panel (IPO), following nine five-year birth cohorts over 29 years. Our sample consists of more than 50,000 individuals, of which more than 13,000 self-employed. Our definition of self-employment is similar to the one of Zwinkels et al. (2017), but there are small differences. We include all freelancers and all those with income from their own enterprise, even those with low income.

We find the following clusters involving self-employment: (1) self-employed with weak labour market attachment, (2) self-employed who spend a large part of their trajectory as benefit recipients, (3) employees with short self-employment spells, (4) employees that switch to self-employment later in their career, (5) always pure self-employed, (6) always hybrid self-employed (combining self-employment with income from another source), and (7) DGAs ("directeur-grootaandeelhouder", directors and majority share holder of their firms). We find that DGAs are the "positive" outliers among the self-employed. They earn more, have higher wealth, and save more for retirement through the (voluntary) third pillar. At the other end we find the first three clusters, whose members do not spend a long period in self-employment. They have

large gaps in their occupational pension accumulation and are not more likely than employees to participate in a third pillar solution. Similar findings hold for the late career switchers. The pure and hybrid self-employed have a slightly higher disposable income than employees and accumulate also significantly more non-pension wealth than employees. Like DGAs, these two clusters have a rather low risk of not having sufficient means after retirement.

Our study is structured as follows. In Section 4.2 we describe the different data sources and give our definition of self-employment. Section 4.3 introduces the methods and presents the self-employment clusters that we find. Section 3.4 discusses the differences between the clusters with respect to demographic characteristics, income and wealth, while Section 3.5 investigates pension accumulation across clusters. Section 4.5 concludes.

3.2 Data

This paper combines several sources of Dutch administrative data provided by Statistics Netherlands (CBS). These are all longitudinal data sets with annual-level data. The core is the Income Panel Study (IPO), which follows a representative sample of the Dutch population over time. Attrition only occurs if individuals decease or move abroad. As such, IPO contains information on more than 90,000 individuals of all ages. These are the so called "core individuals" (in Dutch: "kernpersonen").[3]

The main advantage of IPO over other data such as integral income data sets covering the complete Dutch population, is the long time period covered. IPO starts in 1989[4] and ends in 2014 – when it was replaced by integral population data – and hence covers a period of 26 years, whereas available integral income data currently only cover 15 years. This allows us, for example, to differentiate between those who have always been self-employed and those who switch to self-employment later in life. We extend the time horizon of IPO by three years (2015-2017) by linking the IPO individuals to their records in the integral income data set (INPATAB).[5] Hence our analysis of trajectories covers the years 1989–2017.

For all these years IPO records detailed information on the individuals' income from different sources and demographic characteristics such as age, gender, and marital status. There are two breaks in the series due to tax reforms, but neither of them has an impact on our analysis.[6] The integral data set with which we extend IPO also provides more detailed information on the self-employed starting from 2005: the type of self-employment (i.e., whether the individual is solo self-employed (SSE) or not), and the industry in which the individual is active. We also also link

[3]In addition, IPO contains information on all members of the core persons' households. Because these additional individuals are no longer tracked if they leave the core person's household, we do not include them in the analysis.

[4]There are earlier waves in 1977, 1981, and 1985, but the 1977 sample is unrelated to the later years and the samples in 1981 and 1985 are much smaller.

[5]We also replace the (provisional) income data in IPO 2014 by the (definitive) data from INPATAB The differences matter for the self-employed in particular, as their tax information becomes available later than that of employees.

[6]Our models include year dummies to account for a general shift in taxable income. See online Appendix 3.E for more details on the income data and the breaks in IPO.

Table 3.1: Overview of cohorts in sample

Cohort	Birth years	Age	Years sequenced	T	Cohort N	SE: N (%)
1	1936–1940	49/53–62/66	1989–2002	14	4275	721 (17)
2	1941–1945	44/48–62/66	1989–2007	19	4744	988 (21)
3	1946–1950	39/43–62/66	1989–2012	24	6233	1552 (25)
4	1951–1955	34/38–62/66			5530	1476 (27)
5	1956–1960	29/33–57/61	1989–2017	29	5947	1774 (30)
6	1961–1965	24/28–52/56			6432	2028 (32)
7	1966–1970	24/28–47/51	1994–2017	24	6732	2074 (31)
8	1971–1975	24/28–42/46	1999–2017	19	5803	1607 (28)
9	1976–1980	24/28–37/41	2004–2017	14	5141	1142 (22)
Total					50837	13362 (26)

the individuals to data on household wealth (VEHTAB) and (for a subsample) to accrued pension rights in the second pillar ("Pensioendeelnemingen").

We divide the IPO core individuals into five-year birth cohorts and limit our analysis to those for whom we have at least fourteen years of data during their working age. We consider the range 24–66 as individuals' working age. This allows us to form symmetric cohorts preceding and following the three whose observation window falls completely on their working age.[7] Table 3.1 gives an overview of the nine cohorts that we retain. The oldest cohort includes individuals born in 1936–1940, and the youngest those born in 1976–1980. Cohorts 1 to 4, and partially cohort 5, fall into what Mastrogiacomo (2016) refers to as the old self-employed that are a last group of high income and wealth individuals; the remaining cohorts allow us to analyse whether the newer generations of self-employed are indeed different.

We limit our analysis to individuals who remain in the Netherlands for the whole (cohort-specific) period of interest. We exclude both immigrants and emigrants if they enter or leave the Netherlands in the years we study, since we do not observe them before or after their move. Similarly, we drop everyone who temporarily leaves the country. All in all, this leaves us with approximately 70% of all individuals from the cohorts we consider – a sample of slightly less than 51,000 individuals.

3.2.1 Definition of labour market states

In order to construct employment trajectories we need to define possible labour market states. Each individual can only be in one state in a given year. The definition of the states is similar to the socio-economic categories (SEC) that CBS provides in IPO. Like the SEC, our definition is

[7] We will take into account that individuals retire before age 66 when analysing pension savings. Note that for individuals born after 31 March 1952 the official retirement age is 66 years.

based on observed incomes. We use different categories to differentiate between those who are exclusively self-employed with non-zero income from self-employment only (referred to as "pure self-employed")[8] and "hybrid self-employed" who, in addition to income from self-employment, also receive income from another source. The SEC in IPO makes this distinction until 2000 only.

Individuals who do not receive any income are categorized as having *no income*. Individuals with non-zero income from entrepreneurial activity are categorized as *pure self-employed* or *hybrid self-employed* (self-employed individuals with additional income from employment or pension income). All the other individuals are categorized according to their largest non-zero income as either *employees* (total income from employment in the private and public sector), *freelancers* (income from "other work"),[9] *DGAs* (income as DGA[10]), *benefit recipients* (based on the sum of all social benefits received,[11], *pensioners* (sum of state and occupational pensions), or, finally, if all these incomes are zero but IPO indicates that the individual has an income,[12] as *other income recipients*.

The rightmost column in Table 3.1 shows that the shares of individuals in each cohort who have spent at least one year in self-employment vary between 17% and 32%, with an overall average of 26%. This is much larger than the average percentage in self-employment at a given point in time, which is 11%. The reason is mobility into and out of self-employment. Our analysis will focus on distinguishing between those who remain in self-employment persistently and those who only spend short periods in self-employment.

3.2.2 Descriptive statistics

Table 3.2 shows the demographic characteristics of our sample overall and by labour market state. Even though our sample selection procedure discriminates against emigrants and immigrants, these groups are still represented fairly well. For the year 2000, CBS Statline reports a share of 82.51% Dutch, 3.43% first generation and 5.18% second generation Western immigrants, and 5.59% first generation and 3.29% second generation non-Western immigrants for all ages. The only group that is strongly under-represented are second generation non-Western immigrants, perhaps because our sample excludes cohorts born after 1980 while CBS Statline includes all individuals.

Freelancers and DGAs are less frequently of non-Western origin than pure and hybrid self-employed. The self-employed are on average older than employees and are also more frequently

[8]The income ("profit from business activity") can also be negative.

[9]The difference between self-employed and freelancers is made by the Dutch tax authority. Only the former are recognised as "entrepreneurs" and can benefit from several tax rebates. See online Appendix 3.C.

[10]We apply a correction in 1992 when we define DGAs using the IPO variable SEC, because DGA income is always zero in 1992 (for unknown reasons).

[11]Unemployment benefits (WW and "wachtgeld") sickness and disability benefits ("ziekte wet", "arbeidsongeschiktheid (inkomensverzekering)", "particuliere verzekering ivm ziektekosten/arbeidsongeschiktheid"), "ANW", "ABW", and other benefits ("IOAW", "IOAZ", "Wajong", etc.). The latter includes both unemployment ("IOAZ") and disability ("Wajong") benefits and mixtures ("IOAW",).

[12]This would e.g. be the case for an individual who only has income from returns on investments.

Table 3.2: Demographic characteristics across labour market states

	Average	in %						
	Age	Men	Bread-winner	Single†	Dutch	Western	Non-Western	Sample share
No income	47.24	3.76	2.88	5.79	87.23	7.79	4.98	9.68
Employee	41.81	55.32	62.54	36.68	87.37	7.91	4.73	62.19
Pure SE	45.20	66.24	71.53	27.58	90.35	6.12	3.53	5.78
Hybrid SE	45.72	64.00	80.76	33.52	88.23	7.89	3.89	2.65
Freelancer	46.34	13.33	15.18	11.01	90.01	7.33	2.66	1.21
DGA	47.73	77.92	72.94	20.63	91.01	7.30	1.69	1.56
Benefits	47.65	43.08	66.71	54.94	75.50	10.13	14.37	11.47
Pensioner	61.72	51.56	64.70	27.84	87.79	9.40	2.81	5.07
Other	39.43	40.66	67.24	61.36	81.28	10.05	8.66	0.40
Overall	44.45	49.39	57.97	34.15	86.27	8.12	5.61	

Note: The statistics are calculated over all individual-year observations for each labour market state for a total of 1,228,203 observations. Demographic characteristics are taken from IPO 1989–2014.

† Includes singles, divorced and widowed individuals.

male. Freelancers stand out in particular: they tend to be female and most of them can rely on a partner as the household's breadwinner (defined as the adult with the highest gross income). There is less of a gender difference between pure and hybrid self-employed where roughly two thirds of the observations are for men. Among DGAs, almost three quarters are men. The three types of self-employed are more frequently than employees the breadwinner of their household and are less often single.

Figure 3.1 plots the distribution of labour market states over time for all individuals in our sample between age 24 and 60.[13] The graph confirms what we would expect: (female) labour market participation increases over time: the share of individuals without income declines. The share of self-employed increases. While the share of freelancers fluctuates somewhat over time, the shares of the other three self-employment types are all increasing. The total share of self-employed individuals in the sample rises steadily from approximately 6.6 percent in 1989 to 16.5 percent in 2017.

Figure 3.2 shows the share of self-employed individuals in the sample as a fraction of the working population (self-employed and employees), by cohort and for all cohorts.[14] The decline in the contribution of older cohorts to the self-employment share is partly by construction, as individuals older than 60 are not considered for the calculation of the shares. Hence each cohort

[13]We do not include individuals above age 60 because they, in particular for the older cohorts in our sample, have the option to choose early retirement.

[14]As in Figure 3.1 the analysis is limited to individuals between age 24 and 60. The age restriction is more important here because (early) retirement is not taken up at the same rate by employees and self-employed.

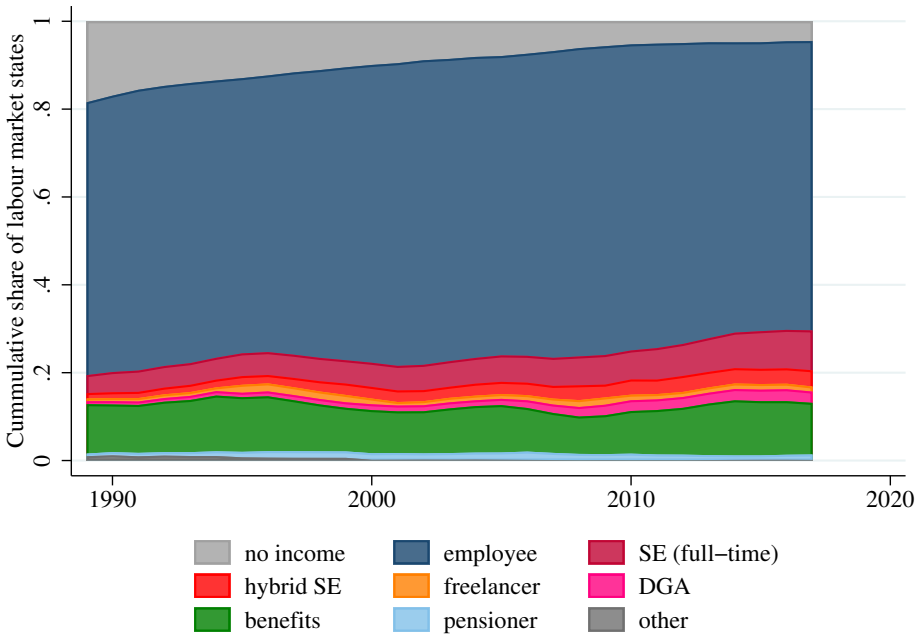


Figure 3.1: Distribution of labour market states over time

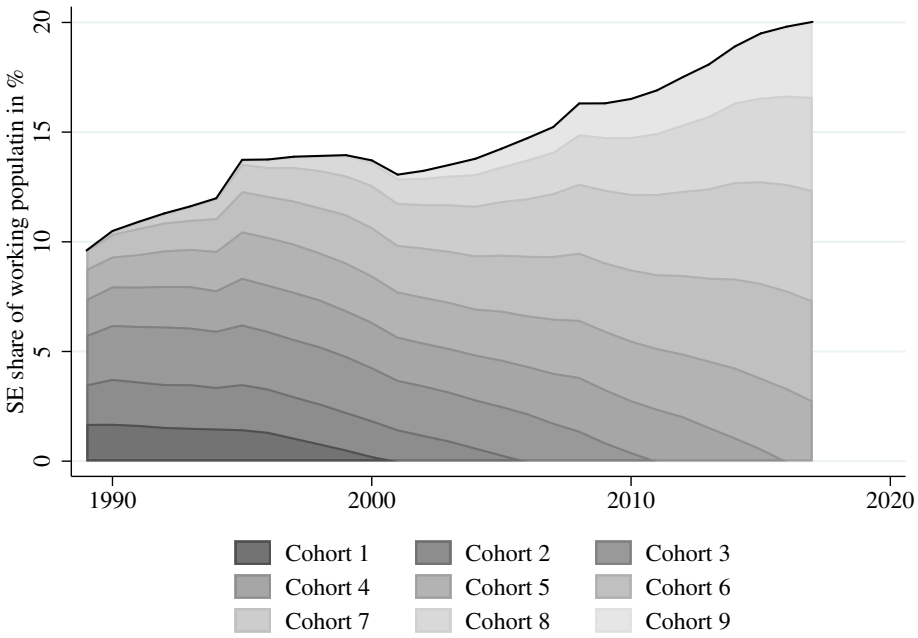


Figure 3.2: Self-employment share and contributions of cohorts

fades out over five years. Similarly, the trough in the early 2000s is due to the youngest cohort, which enters the sample later than that the oldest cohort leaves; in addition, the share of self-employed is lower among individuals at the beginning of their careers. We have no explanation for the sharp increase in the self-employment rate in 1996. The plot shows that until the beginning of the 2000s, all cohorts contribute approximately equally to the self-employment share. Thereafter we see that the contribution of the younger cohorts 6 – 9 is increasing over time; these cohorts mainly drive the rise in self-employment.

How do the self-employed differ from individuals in other labour market states? Table 3.3 shows the sample distribution of taxable income at the individual level, and of disposable household income. To make values comparable, we adjust household income using the CBS equivalence scales, convert all values to euro and deflate with CPI data from The World Bank World Development Indicators (base year 2010). Two general observations can be made: First, women's taxable income is on average lower than men's. Second, women's disposable income is generally higher than their taxable income while the opposite holds for men. This is as we would expect considering that the majority is married/living with a partner, and women less often participate in the labour market and are more likely to work part-time.[15] This is also reflected in the median and the 25th and 75th percentiles of taxable income: Median taxable income for female employees, self-employed (full-time and hybrid) and DGAs is between 55 and 60% of males and similar ratios are observed for the quartiles. This can also be observed for pensioners, in line with the fact that occupational pensions are related to earnings before retirement.

Comparing taxable income of different groups shows that the DGAs earn much more than employees or other self-employed. This is partially due to the minimum salary requirement for DGAs imposed by the tax rules. Because it would be fiscally attractive for DGAs to pay themselves dividends instead of a salary, the tax rules set a minimum salary ("gebruikelijkloonregeling") that was 45000 euros in 2017.[16] Individuals would generally not incorporate their business if they could not afford to pay themselves this salary. This is in line with the findings by Levine and Rubinstein (2017) and Humphries (2016). The other types of self-employed (who are generally not incorporated), have lower median incomes than employees have. Finally, the differences between the labour states become smaller at the 75th percentile, showing that the dispersion in income among the self-employed is larger than among employees.

Closer inspection of the differences in taxable income between men and women shows that the largest difference is found for freelancers. For female freelancers, the median taxable income is less than 20% of that of employees, whereas for male freelancers the median income is 72% that of employees. This stands in stark contrast with hybrid self-employed women, whose median taxable income is 87% of that of employees and whose third quartile is even higher than that of female employees. The same holds for men, which makes the hybrid self-employed the second

[15]The income data presented in Table 3.3 is not adjusted for FTE since our data do not contain information on hours worked.

[16]Exceptions exist for new enterprises, part-time workers, or firms that make a structural loss.

Table 3.3: Taxable individual and disposable household income, wealth and sample share by gender and labour state

	Women				Men			
	Median	Q_1	Q_3	Share (%)	Median	Q_1	Q_3	Share (%)
Individual taxable income								
No income†	0	0	0	17.91	0	0	0	0.71
Employee	20457	12925	28754	59.03	33142	26253	43312	74.55
Pure SE	10701	2195	24074	4.03	19996	8193	38220	8.07
Hybrid SE	17847	8466	30015	1.91	28833	15672	46116	3.26
Freelancer	3513	1642	8840	2.12	23925	9342	43091	0.32
DGA	31161	18152	49364	0.69	56338	40593	80695	2.52
Benefits	13772	9557	16906	12.35	16570	13338	22803	9.13
Pensioner	20260	9272	31605	1.49	32393	24600	44071	1.11
Other	0	0	10909	0.47	0	0	4346	0.33
Disposable household income								
No income	17815	14083	23348	18.71	11302	1395	17606	0.68
Employee	23457	18525	29715	58.61	22944	18084	28998	75.06
Pure SE	25255	17444	36209	3.84	23950	16306	33759	7.83
Hybrid SE	25336	18061	34730	1.78	25066	17977	33551	3.23
Freelancer	20073	15792	26480	2.13	19408	13414	27244	0.31
DGA	35564	25788	48590	0.67	31219	23236	42497	2.36
Benefits	14358	11468	20234	12.25	14275	11281	19606	9.02
Pensioner	23127	17958	30108	1.50	21859	17001	28517	1.13
Other	12684	5443	18966	0.51	8643	3579	16922	0.38
Wealth (in thousands)								
No income	74	8	172	10.48	46	1	197	0.65
Employee	40	3	117	65.62	38	2	114	72.57
Pure SE	99	19	234	4.66	89	16	217	9.24
Hybrid SE	72	9	198	2.71	66	6	181	3.80
Freelancer	68	11	158	2.34	28	1	134	0.29
DGA	288	106	619	0.96	218	79	530	3.62
Benefits	2	0	44	11.29	3	0	52	8.76
Pensioner	108	17	252	1.78	107	13	230	1.02
Other	23	1	158	0.16	29	1	180	0.06

Note: All values are in euro and deflated with CPI data from The World Bank World Development Indicators (base year 2010). Disposable income and wealth are measured at the household level and additionally adjusted with the CBS equivalence scale. The statistics are calculated over all individual-year observations for each labour market state for all individuals in the sample aged 24–60. Taxable income is available for all years (563,259 and 549,279 observations for women and men respectively). Disposable household income is only available for the years 1989–2014 (518,086 and 505,583 observations for women and men respectively). Household wealth is available for 2006–2014 (168,138 and 163,067 observations for women and men respectively). No adjustment has been made to income for FTE.

† While it may seem counter intuitive, individuals with no income may still have (positive) taxable income. That is because home owners get taxed on a hypothetical income that they could earn if they were to let their homes.

Table 3.4: Share of individuals (in %) contributing to pension pillars

	Women			Men		
	2nd Pillar	3rd Pillar	N	2nd Pillar	3rd Pillar	N
Employee	67.74	6.51	332831	78.28	16.81	409571
Pure SE	0.00	12.68	22808	0.00	26.12	44584
Hybrid SE	51.08	9.57	10756	57.04	20.88	17971
Freelancer	2.97	1.41	11939	4.90	15.27	1775
DGA	3.88	10.01	3895	2.58	25.32	13845
Benefit recipient	5.62	1.58	69764	7.93	3.87	50196
Pensioner	12.83	3.94	8433	20.73	9.20	6096
Other	0.00	3.19	2791	0.00	3.94	1955

Note: Observations are counted as individuals per year per labour market state. The reported shares and observations are for all individuals in our sample aged 24–60.

most successful group of self-employed after DGAs. The inter-quartile range of taxable income is largest for the pure self-employed and for freelancers, indicating large income dispersion in these groups.

A different picture is found for disposable household income. Adjusted for household size, the values are very similar across labour states, except for DGAs and freelancers. The higher taxable income of DGAs and lower income of freelancers are also reflected in their household's disposable incomes. Disposable household income is substantially higher than personal income for women and vice versa for men, indicating that women are less frequently the breadwinner in the household. This holds in particular for the lower half of the income distribution of freelance, hybrid and pure self-employed women. Moreover, the dispersion of disposable income is larger for most self-employment types and while the median values of pure and hybrid self-employed are relatively close to the median of employees, the third quartile is much larger.

The lower panel of Table 3.3 reports household wealth adjusted for household size. Wealth is measured net of debts and includes financial wealth, real estate, firm wealth (for non-incorporated self-employed), substantial interest shares (i.e. firms in which the household owns > 5% - hence potentially wealth in firms of incorporated self-employed as well as DGAs), and other wealth. The dispersion of wealth is larger for the self-employed than for employees. Once more, DGAs are outliers, with much more wealth than any other group of self-employed.[17] For all self-employed except male freelancers, the quartiles of household wealth are higher than those of employees. Self-employed women have higher median wealth than employees.

Table 3.4 shows the shares of individuals that make a contribution to the second pillar (occupational pensions) and third pillar (voluntary private pensions). By definition, the pure

[17]Furthermore, unlike unincorporated individuals, capital in a DGA's firm does not show up in the household wealth statistics. Hence the wealth figures for DGA's are a lower bound.

self-employed make no contribution to the second pillar. Contributions by pure self-employed (e.g., voluntary payments to the pension fund of an old employer) are recorded as third pillar contributions. The same rules also apply to the other self-employed, but as we can see in Table 3.4 both DGAs and freelancers still have a small share of individuals that contribute to a pension fund – they have an employment contract with an income above the fund's franchise.

The share with contributions to the second pillar is much larger for hybrid self-employed, as expected. Still, compared to employees, a larger share of hybrid self-employed is not contributing to an occupational pension. This is because among the hybrid self-employed, there are more individuals whose earnings as an employee are below the franchise or work for a company without occupational pension. For the same reasons, we do not see a 100% share for employees, particularly for women.

Men are more likely to contribute to either pillar than women. The differences are particularly large for the third pillar. For example, 25% of male pure self-employed make a contribution to the third pillar, compared to half as many women. Contributions to the third pillar are more common among the pure self-employed than for DGAs or hybrid self-employed. The freelancers are once more the worst performing group among the self-employed, with the largest difference between genders. While female freelancers make almost no contributions to the 3rd pillar, their male counterparts still do so in 15% of all cases.

3.3 Labour market trajectories over the life cycle

This section describes the analysis of labour market trajectories. First, we show how self-employment spells vary across individuals' working lives. Then we introduce sequence analysis and explain the optimal matching algorithm with which we calculate so-called edit distances between all pairs of trajectories. Finally, we explain how clustering can be used as a tool to build a data-driven classification of labour market trajectories.

3.3.1 Visualisation of employment trajectories

An employment trajectory is an individual's sequences of labour market states over time. Figure 3.3 shows the distribution of those states over time for the individuals of the 1961–1965 birth cohort (cohort 6) who are categorised as self-employed in at least one of the 29 years in which we observe them. The figure shows that the share of individuals who are self-employed at a given point in time is increasing over time, in accordance with Figures 3.1 and 3.2. In addition, the share of benefit recipients in this subsample is smaller than in the overall population shown in Figure 3.1. Furthermore, the shares of all types of self-employed increase but the share of full-time self-employed increases most. The same patterns are also observed across other cohorts (not shown), though less pronounced in the oldest cohorts who often had attractive early retirement options. The graph does not show, however, what kind of labour market careers the

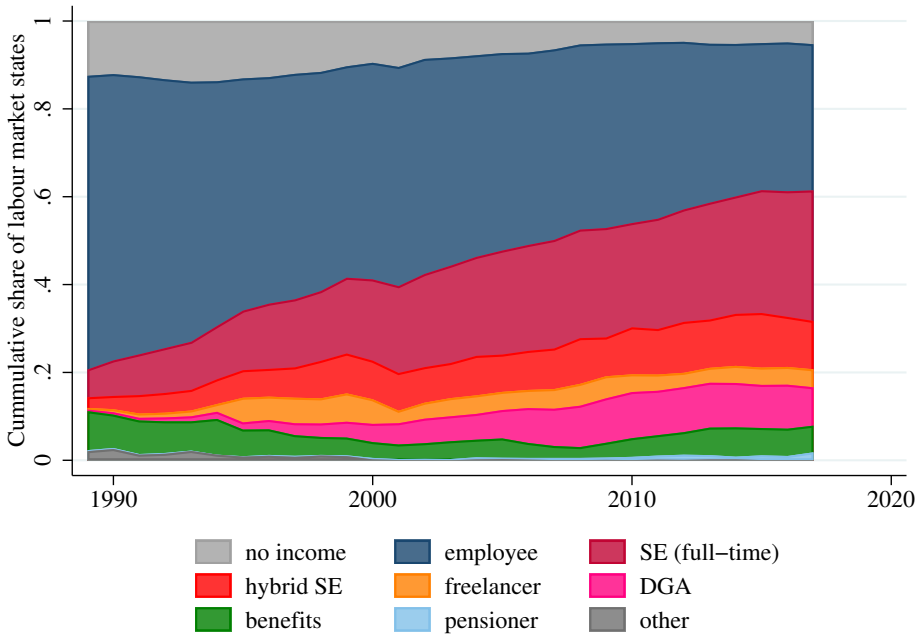


Figure 3.3: Distribution of labour market states over time of cohort born in 1961-1965 (self-employed)

self-employed have had or will have. For example, are the individuals that are self-employed early in their career also self-employed in later years? How much mobility is there into and out of self-employment? If so, what are the labor market states from which they enter self-employment, etc.?

One way to answer these questions is to plot all labour market trajectories. Figure 3.4 essentially does that and shows a so-called index plot[18] of the trajectories of the same individuals as in Figure 3.3. The individual's paths are stacked vertically on top of each other and each path is a (thin) horizontal line on which each year is coloured according to the individual's labour market state in that year.[19] The figure conveys two messages. First, only a small share spends the majority of the 29 years in self-employment. Overall, a quarter of the individuals spend no more than three years in self-employment, less than half spend more than ten years in self-employment, and only around one third spend more than half of the 29 years in self-employment.[20] Second, the trajectories vary a lot across individuals and two trajectories of different individuals are hardly ever exactly the same. We want to know whether this large variation in trajectories can

[18]Index plots were first introduced by Scherer (2001). We construct all index plots in this paper using the program SQ (Brzinsky-Fay et al., 2006) in Stata.

[19]To protect the privacy of the persons in the sample we consciously overplot all our index plots so that zooming in does not reveal the individual trajectories and individuals cannot be recognised.

[20]Note that we only use one observation per year, so that short spells or short interruptions of self-employment are not always accounted for.

help us to better understand the large variation in income, wealth and pension savings of the self-employed.

3.3.2 Sequence analysis and optimal matching

In order to have a better picture of how the trajectories differ, we turn to sequence analysis, a methodology that has been used in sociology for some years to study sequences of social events. It was first introduced in the field of social sciences by Abbott (1983) and further developed in, e.g., Abbott and Forrest (1986); Abbott and Hrycak (1990); Abbott (1995). Sequence analysis takes a holistic approach, not only taking account of the types of states that individuals experience over time, but also of their duration. It studies the complete sequence in order to understand the importance of different patterns that the trajectories may have. As argued by Studer and Ritschard (2016, p. 481), sequence analysis thus stands in contrast to, e.g., survival or event history analysis that focus on specific events rather than an overall view off the trajectories.

Sequence analysis compares all sequences pair-wise in order to find sequences that display similar patterns. In our case, the aim is to group individuals that share a similar history of labour market states at similar times and in a similar order. The first step is to compute a dissimilarity measure for all unique sequence pairs. Studer and Ritschard (2016) provide an overview of the dissimilarity measures that are used in the field. We use the most common measure for discrete sequences: Optimal Matching (OM).[21] OM provides a measure of "edit distance" between each pair of sequences: Given a pair of sequences, OM computes the least costly way in which one sequence can be converted into the other one. The operations used for this conversion are (1) substitutions, (2) deletions, and (3) insertions. Each of these is associated with a cost and the edit distance is the minimum cost at which the conversion can be achieved.

Consider the hypothetical example given by Table 3.5, which shows four individuals with different labour market trajectories. For the sake of simplicity, it covers only 8 years and three labour market states: employment (E), self-employment (S), and not working (N). Let us also assume unit costs for all operations. For trajectories 1 and 2 the computation is straightforward. The two trajectories only differ from each other in year 3. Substituting E with N at the third position in the first trajectory is the fastest manner to transform it into the second. The edit distance between the second and first trajectories is thus equal to 1. On the other hand, trajectories 2 and 3 only coincide in the last four years. Relying on substitution only would require four substitutions, resulting in a total cost of 4. Alternatively, using also insertion and deletion, we can achieve the same transformation with fewer steps. The solution with the fewest steps involves deleting the first period in trajectory 3, shifting the whole sequence one position to the left. Next the N, which is now in first position, is substituted with an E. Finally, an S is inserted in the seventh position. This gives an edit cost of 3 between trajectories 2 and 3.

[21] OM was originally developed in computer science and in biology. We use the package SADI in Stata (Halpin, 2017).

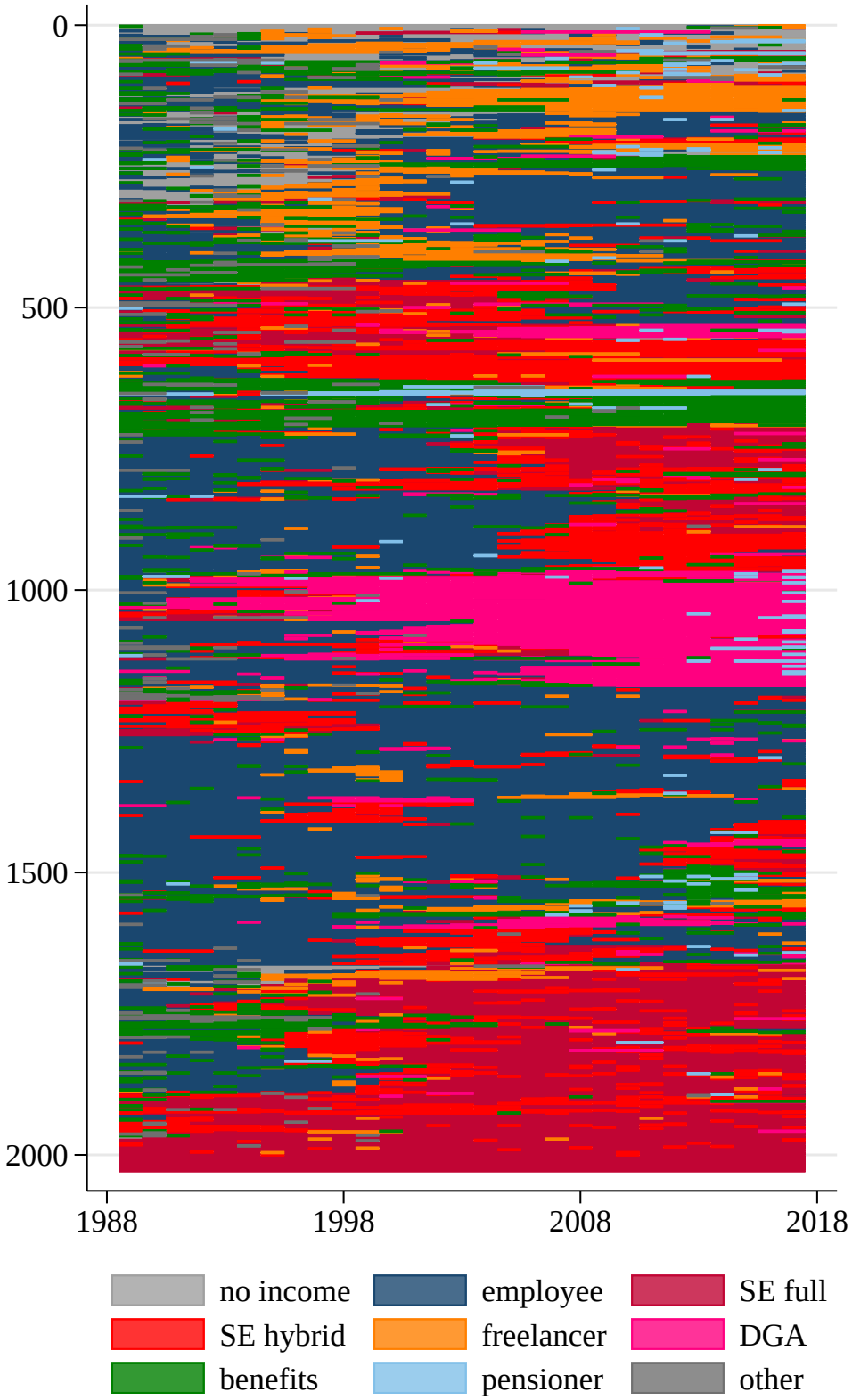


Figure 3.4: Index plot of cohort born in 1961-1965 (self-employed)

Table 3.5: Hypothetical labour market trajectories

	Year							
	1	2	3	4	5	6	7	8
Trajectory 1:	E	E	E	S	S	S	S	E
Trajectory 2:	E	E	N	S	S	S	S	E
Trajectory 3:	S	N	E	N	S	S	S	E
Trajectory 4:	S	S	S	S	E	E	E	E

Our example used unit costs for all operations. It is easy to see that the optimal outcome and the distances between trajectories will change if costs are chosen differently. This raises the issue how the costs for the operations should be defined. A popular solution is to derive the substitution costs from observed transition rates. Low costs are then assigned to frequently observed transitions, and higher costs to rare transitions. Studer and Ritschard (2016) point out, however, that observed transition rates are generally low and the resulting substitution costs are close to 2. Because of this, the approach does not produce results that differ much from those obtained with fixed state-independent costs.

There is less discussion on the setting of insertion and deletion, or "indel" costs in the literature.[22] Most applications use the same value for insertion and deletion and the only choice therefore concerns the value that should be assigned to indel versus substitution costs. To understand this, consider the last of the example trajectories in Table 3.5. If we compare trajectories 1 and 4 we can see that both individuals spend half of their time in self-employment and the rest in employment. The shortest way to transform trajectory 4 into trajectory 1 involves six steps: either six substitutions or three deletions at the end of trajectory 4 together with three insertions at the beginning. If substitution and indel costs are the same, the algorithm is indifferent between the two. If insertion and deletion however are cheaper than substitution, indel operations will be the preferred way to make these two trajectories alike, etc. Thus as indel operations become cheaper compared to substitutions, OM will place less importance on the timing and more on the sequencing of events. Moreover, note that as soon as indel costs are less than half of substitution costs, any substitution is more costly than two indel operations achieving the same result.

Based on these considerations we decided to set substitution costs equal to 2 for all labour market state transitions, and we set the indel costs equal to 1.5. We then apply OM separately to each cohort, where before OM, we split each cohort into two samples separating those individuals that are at least one year self-employed from those that are never self-employed. The latter group will be used as a control group to which we compare the self-employed.[23]

[22] Hollister (2009) offers a good discussion of indel costs. However, her proposed solution (localised OM) has its own problems as discussed by Studer and Ritschard (2016).

[23] We set the indel costs equal to 1 for the non-self-employed sub-sample. We choose the lower cost because we care

3.3.3 Clusters of self-employment

The result of OM is a matrix containing all pairwise distances. The distance measures can be used to group all individuals into clusters using machine learning; see, e.g., Zacher et al. (2012), Madero-Cabib and Fasang (2016), Tophoven and Tisch (2016), Visser (2018), Munnell et al. (2019), or Humphries (2016). We follow this approach and use Ward's Method (Ward, 1963) to cluster the trajectories.[24]

Ward's Method produces a tree of potential groupings. In practice, the researcher has to choose how many clusters to use. This is not a straightforward decision – no "hard" statistical criteria exist and different criteria often lead to different numbers of clusters. Nevertheless, the applications in the literature show that clustering has its merits. In our application, it provides a tool to separate individuals with long periods in self-employment from those with short self-employment spells and helps to automatically distinguish groups that spend their time in different types of self-employment. We will choose cluster solutions that match ex-ante expectations of self-employment types. For example, we expect different clusters for long-term (hybrid) self-employed, freelance workers, and DGAs.

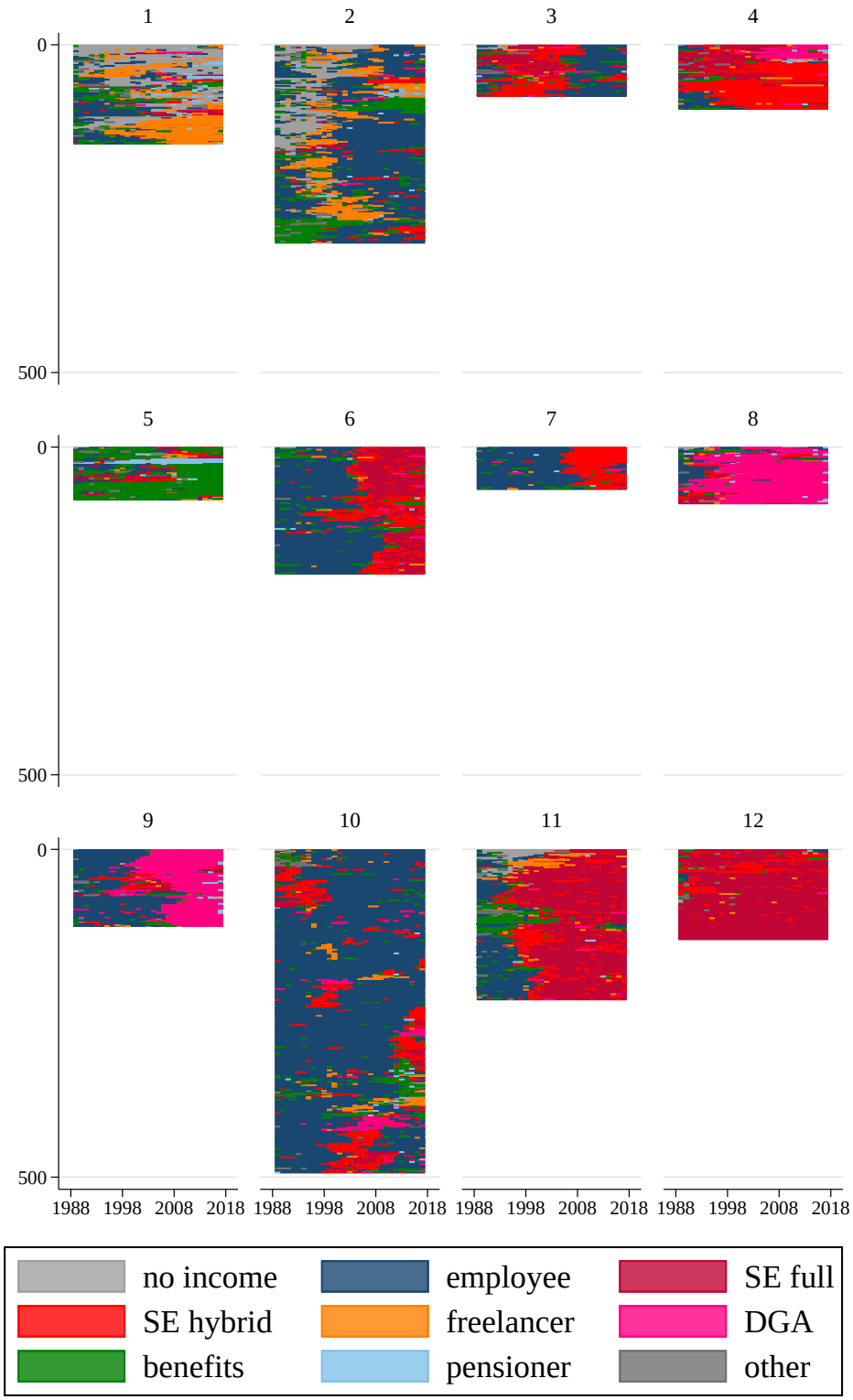
As with OM we apply the clustering algorithm to each cohort separately. Ward's Method does not split up all cohorts in the same manner. If we would choose the number of clusters based upon popular criteria, we typically would get 5 or 6 clusters according to the Calinski-Harabasz index, and 6 to 8 according to the Average Silhouette Width (see Studer (2013) for a discussion of several measures). These clusters would not be the same across cohorts. Instead of strictly following the criteria, we choose a clustering that is harmonized across cohorts. We start with a rather fine partitioning into twelve groups for each cohort (Ward's solution). This is the smallest partitioning that gives a meaningful grouping. Based on the Ward solution, we then build seven clusters of self-employment trajectories. As we will show below, this allows for a sufficient level of distinction between the different clusters while maintaining parsimony and manageability. We will use the clustering of the self-employed in cohort 6 as an example to illustrate this.[25]

Figure 3.5 presents the same trajectories as shown in Figure 3.4, but now broken down into Ward's twelve groups. Several groups mainly contain one type of self-employed: Group 4 is dominated by hybrid self-employment, groups 8 and 9 are DGAs, and groups 11 and 12 are dominated by pure self-employment. Similarly, group 5 are individuals whose main income in most years are social benefits. We also see that timing of sequences matters: group 3 consists of individuals that start as self-employed and then switch to employment later, whereas groups 6 and 7 do the opposite. Finally, groups 1, 2 and 10 consist of more volatile trajectories which have

less about the timing of e.g. unemployment spells and more about their length.

[24]Ward's Method is a hierarchical agglomerative method. The algorithm starts by taking each sequence as its own cluster. It then identifies which two clusters can be merged with the smallest increase in variance within clusters. This is repeated until all sequences are merged into one large group. See Chapter 3 in Aldenderfer and Blashfield (1984) for details and comparison to other algorithms.

[25]The mapping for each cohort from Ward's twelve groups solution to our seven clusters is given in online Appendix 4.C.



Graphs by Ward's 12 cluster solution

Figure 3.5: Index plot by clusters of cohort born in 1961-1965 (self-employed)

short self-employment spells.

This leads to seven larger clusters involving self-employment that we construct from the 12 Ward groups (denoted in "[ ]"):

1. *Weak labour market attachment / freelancer* [1, 2] – In the older cohorts this cluster predominantly consists of individuals that spend a large share of their trajectory without any income. In younger cohorts this is less the case and most trajectories either start or end with a long spell without income. The predominant self-employment type observed in these trajectories is freelancer.

2. *Benefit recipients* [5] – This cluster contains the individuals who receive some type of benefit for most of the observed period.

3. *Mostly employed with short SE spells* [3, 10] – Almost no cohorts have a distinct cluster of individuals that start as self-employed and then switch to employment. We therefore also include these individuals in the cluster with short spells.

4. *Employees that switch to SE later on* [6, 7] – We do not differentiate between hybrid and (full-time) self-employed in this cluster.

5. *Always pure SE* [11, 12] – In particular in the younger cohorts (observed from age 30 or younger) this also includes trajectories where few of the initial years are spent in employment.

6. *Always hybrid SE* [4] – This cluster is defined similarly to the preceding cluster but for hybrid self-employed.

7. *DGAs* [8, 9] – For DGAs we do not differentiate with respect to the timing. Most DGAs either spend the whole trajectory in this state or the second half. Since DGAs are a small (and special) group, we merge both types into one cluster.

For the "control group" of individuals who are never self-employed, we follow a similar strategy. Their clusters are built on the six groups solution of Ward's Method.[26] We retain four main clusters: *8) Weak labour market attachment (no SE)*; *9) Benefit recipient (no SE)*, *10) Employee with long spells as benefit recipient* and *11) Employee (including short spells as benefit recipient)*. For cohorts 1–4 we also include a separate cluster *12) Pensioner*, which consists of individuals that are pensioners for most of the years that we observe them.

Table 3.6: Demographic characteristics across clusters in sample

		in %						
Cluster	Age	Men	Bread-winner	Single†	Dutch	Western	Non-Western	Sample share
with self-employment								
1	45.10	6.12	12.06	12.33	89.32	6.98	3.70	3.97
2	45.89	43.09	60.66	45.53	77.50	9.37	13.12	1.28
3	42.99	47.59	54.98	31.95	86.47	9.25	4.28	8.55
4	41.38	59.81	69.12	39.52	89.13	6.03	4.84	3.41
5	44.33	67.26	69.58	31.65	89.86	6.49	3.65	5.65
6	43.87	58.12	70.49	31.76	88.53	7.96	3.51	1.48
7	44.48	78.07	74.27	27.06	90.85	7.39	1.77	2.68
without self-employment								
8	50.02	2.27	7.32	9.71	86.80	8.31	4.89	7.54
9	46.42	38.65	63.65	53.01	74.84	9.99	15.17	8.32
10	46.27	41.87	55.63	37.08	83.13	8.67	8.20	9.45
11	43.12	59.77	66.20	36.53	87.83	7.84	4.33	47.00
12	55.60	43.80	70.36	39.27	85.27	11.91	2.82	0.68
Overall	44.45	49.39	57.97	34.15	86.27	8.12	5.61	

Note: The statistics are calculated over all 1,228,203 individual-year observations for each cluster.
† Includes singles, divorced and widowed individuals.

3.4 Differences among self-employment clusters

In this section we compare demographic characteristics, income and wealth of individuals across the clusters identified in Section 4.3. Table 3.6 shows the demographic characteristics across the twelve clusters involving (1-7) and not involving (8-12) self-employment. The patterns are largely similar to those for the labour market states shown in Table **??**. For example, both clusters with weak labour market attachment (clusters 1 and 8) have a high proportion of women who frequently also have a partner that acts as the main income earner in their household. Similarly, women are also the majority (around 60%) in the clusters of benefit recipients (clusters 2 and 9). In the other self-employment clusters, the gender imbalance is less pronounced than in Table **??**. It is slightly in favour of women in the cluster of individuals with short self-employment spells (3), whereas the other self-employment clusters contain more men than women. In particular, the DGA cluster (7) remains predominantly male (around 80%).For late switchers (4) and hybrid self-employed (6) the share of men is lower in Table 3.6 than in Table **??**. This can be explained

[26]The Ward's four groups solution already leads to meaningful clusters for cohorts 6 and 8. To better account for individuals who retire early in cohorts 1-4, we nevertheless use a finer break down.

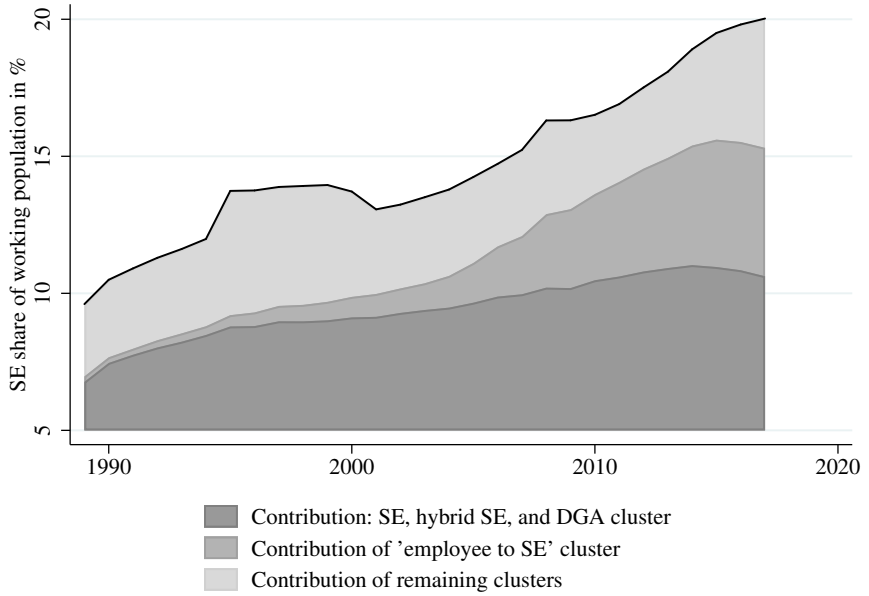


Figure 3.6: Self-employment share and contributions of clusters

by the distribution of the length of self-employment spells.[27] Men generally spend more time in self-employment than women and hence make up a larger share of the self-employed observations in Table **??**. For similar reasons we see a lower share of breadwinners in clusters 4–7, reflected in the relatively low share of breadwinners among individuals with short SE spells.

We find similar differences for marital status. Individuals in clusters 1 and 8 are mostly married. For immigration status, there is no significant difference across clusters, and the share of both western and non-western immigrants (first and second generation) is slightly larger in the low attachment and benefit recipient clusters.

Average age is similar across clusters by construction, as all individuals are observed for a similar period in their lives. Still, the increase of self-employment in younger cohorts is reflected in a lower average age for the self-employment clusters (3–7). Furthermore we see the increase in labour market participation among women in younger cohorts reflected in a higher average age of cluster 1 in comparison to cluster 8. In the latter cluster, most individuals do not participate in the labour market and that cluster's sample share is decreasing over cohorts.

Self-employment over time

Figure 3.6 shows the same overall self-employment shares as Figure 3.2 and adds the contributions that different clusters make to the self-employment rate. Only about half of the share is

[27] We consider any sequence of consecutive years in any of the four types of self-employment as a spell. See Table 3.A.1 in Appendix 3.A for a detailed overview.

due to the clusters in which individuals spend most of their working years in self-employment (hybrid and pure self-employed and DGAs).The share of individuals who become self-employed later in their career (cluster 4) is growing substantially over time, to approximately a quarter of the self-employment share in recent years. Finally, a constant share of approximately three to five percentage points is always attributed to the clusters of individuals that do not remain in self-employment (clusters 1–3). These findings are similar if we look at the cohorts separately (details available upon request).

Table 3.A.1 in Appendix 3.A shows that in most clusters, around three out of four individuals have only one self-employment spell, while roughly one in five have two. The exception is the cluster with weak labour market attachment where a larger share has more spells.In clusters 1, 2 and 3 the median spell length is only 2 to 3 years, whereas individuals in the other clusters have a median spell length of at least 7 years. For pure and hybrid self-employed, the top 25% spend almost all of the observed period in self-employment. The distribution of self-employment spell lengths hardly changes across cohorts.[28]

Income

The upper and middle sections of Table 3.7 show individual taxable income and disposable household income (adjusted for household size) by gender and cluster. Comparing the results to Table **??** we see that most of the results for the clusters are similar to those for the corresponding labour market status. We still find that self-employment is different for men and women. Women who switch to self-employment after a career as an employee have higher incomes than other self-employed clusters, with the exception of DGAs. Furthermore, women with short self-employment spells also do relatively well in terms of income. The same holds for men where this is in fact the best earning cluster.

Disposable income for the household gives smaller differences among clusters. For women the starkest contrast is between individuals with short self-employment spells and the other groups. For men, both pure and hybrid self-employed have larger interquartile ranges than switchers and individuals with short spells.

To better understand the relation between our clusters and income, we estimate the following model:

$$y_{it} = \beta_0 + X'_{it}\beta_x + L'_{it}\beta_l + W'_i\beta_c + C'_i\gamma + T'\delta + \epsilon_{it} \tag{3.1}$$

Here Y_{it} is the income variable of interest.[29] X_{it} is a vector with characteristics of individual i

[28]See Table 3.A.2 in Appendix 3.A for spell lengths across cohorts.

[29]As a robustness check we also use taxable income in box 1 (income from work and housing) as the dependent variable. Data on taxable income by source is available since 2001. Overall, the results obtained for this specification are very similar to those for total taxable income. Considering that the large majority have very little taxable income in box 2 (income from substantial business interest) or 3 (income from savings and investments), this is not surprising.

Table 3.7: Taxable individual and disposable household income, wealth and sample share (%) by gender and cluster

	Women				Men			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Taxable individual income								
Weak LM attach. (SE)	1141	0	8584	7.39	16631	939	31880	0.51
Benefit recipient (SE)	12483	5077	17226	1.44	16729	10824	26467	1.07
Short SE spells	15677	6320	25832	9.13	32622	23061	47380	8.49
Employee to SE	19032	7847	30509	2.84	29333	19381	42109	4.31
Pure SE	11219	1890	23401	3.66	20731	9249	34621	7.64
Hybrid SE	16450	6338	26802	1.26	28899	16177	45410	1.72
DGA	26444	12857	44002	1.16	50669	33345	74614	4.28
Weak LM attach.	0	0	0	13.07	759	0	18402	0.34
Benefit recipient	14045	9142	18232	9.91	18376	13729	25069	6.17
Long benefit rec.	14945	7575	22412	10.97	27625	20199	35272	7.68
Employee	21582	14253	29846	38.58	33405	26522	43299	57.31
Pensioner	21092	10535	31060	0.59	33879	26184	41556	0.47
Disposable household income								
Weak LM attach. (SE)	19033	14737	25157	7.45	18004	11953	24937	0.50
Benefit recipient (SE)	16560	12295	22428	1.45	15488	11485	22242	1.08
Short SE spells	22496	16947	29572	9.08	23251	17446	30770	8.42
Employee to SE	24604	17667	33326	2.78	22791	16929	30515	4.25
Pure SE	23559	16195	33522	3.67	23074	15848	31791	7.65
Hybrid SE	24132	17468	32340	1.25	24325	17298	32523	1.72
DGA	33470	23234	46698	1.17	30053	22136	41097	4.28
Weak LM attach.	19114	14930	24874	13.58	15221	10733	21973	0.34
Benefit recipient	15943	12112	22024	9.93	15037	11676	20211	6.23
Long benefit rec.	20039	15035	26552	11.00	19714	15130	25199	7.85
Employee	23500	18599	29695	37.99	23009	18199	28928	57.16
Pensioner	23193	18389	30735	0.64	21420	17354	26831	0.52
Household wealth (in thousands)								
Weak LM attach. (SE)	69	12	156	6.96	27	0	118	0.50
Benefit recipient (SE)	5	0	95	1.39	1	0	32	0.93
Short SE spells	58	5	150	9.71	39	1	129	8.90
Employee to SE	52	6	163	3.25	47	3	127	4.88
Pure SE	93	15	233	3.71	95	15	240	7.73
Hybrid SE	92	8	240	1.40	82	7	227	1.72
DGA	267	86	593	1.12	192	68	474	4.31
Weak LM attach.	65	5	156	8.57	44	1	138	0.28
Benefit recipient	3	0	58	8.98	2	0	25	5.12
Long benefit rec.	25	1	107	10.81	17	1	97	6.49
Employee	39	3	114	43.99	40	2	115	59.05
Pensioner	147	20	270	0.12	89	7	158	0.09

Note: All values are in euro and deflated (WDI CPI, base year 2010). Disposable income and wealth are adjusted for household size (CBS equivalence scale). The statistics are calculated over all individual-year observations by cluster for all individuals in the sample aged 24–60. Taxable income is available for all years, disposable household income for the years 1989–2014, and household wealth for 2006–2014. No adjustment has been made to income for FTE. The sample sizes are the same as in Table 3.3.

in period t such as age and its square, civil status and country of origin controls,[30] as well as dummy variables for whether the individual is their household's main income earner, whether self-employed individuals are solo self-employed (SSE).[31]

Table 3.8: Panel regression – taxable income

	taxable income (in thousands)			
	women		men	
LM state: pure SE	-1.84	-1.81	-13.27***	-12.64***
	(1.212)	(1.225)	(1.274)	(1.265)
hybrid SE	1.28	1.29	-5.94***	-6.05***
	(1.121)	(1.123)	(1.600)	(1.567)
freelancer	-1.58	-0.90	0.42	1.12
	(1.104)	(1.108)	(2.354)	(2.354)
DGA	10.61***	9.28***	14.86***	11.09***
	(1.660)	(1.753)	(2.119)	(2.137)
cluster: weak LM attach. (SE)		-9.18***		-10.49***
		(0.198)		(1.399)
benefit recipient (SE)		-6.85***		-8.54***
		(0.316)		(0.735)
short SE spells		-2.66***		3.89***
		(0.272)		(0.854)
employee to SE		-0.02		1.11
		(0.421)		(0.636)
pure SE		-2.28***		-4.64***
		(0.492)		(0.614)
hybrid SE		-1.10		2.92
		(0.900)		(1.925)
DGA		11.70***		19.43***
		(2.409)		(1.154)
is SSE	-6.39***	-6.37***	-6.51***	-6.49***
	(1.099)	(1.099)	(1.605)	(1.597)
unknown SSE status	-3.51***	-3.59***	1.28	0.92
	(1.053)	(1.056)	(1.660)	(1.664)
breadwinner	5.36***	5.31***	4.68***	4.55***

[30]We take singles as the base category and include indicators for married, widowed and divorced individuals. The base category for the country of origin variable is native Dutch (both parents born in the Netherlands); we also distinguish first and second generation immigrants of Western and non-Western origin.

[31]The SSE information is not available for all years. We also include a dummy variable that captures the cases where the SSE status is unknown.

Table 3.8: Panel regression – taxable income (continued)

	taxable income (in thousands)			
	women		men	
	(0.134)	(0.134)	(0.171)	(0.171)
main income source: profit	-0.45**	-0.52***	6.29***	6.21***
	(0.156)	(0.156)	(0.627)	(0.622)
freelancer	-0.79	-0.95	-7.74***	-8.73***
	(1.078)	(1.096)	(2.081)	(2.107)
constant	5.99*	2.14	-22.11***	-26.24***
	(2.458)	(2.335)	(4.675)	(4.577)
Non-SE clusters		✓		✓
Observations	576376	576376	563348	563348
Individuals	25965	25965	25441	25441
σ_u	8.159	8.118	18.873	18.819
σ_e	11.200	11.200	34.411	34.411
R^2 within	0.205	0.206	0.068	0.069
R^2 overall	0.299	0.327	0.118	0.129
R^2 between	0.451	0.472	0.232	0.260

Clustered standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including year and cohort fixed effects, and controlling for individual's demographic characteristics, non-self-employed labour market states, and other main income sources.

Table 3.9: Panel regression – disposable household income

	disposable household income (in thousands)			
	women		men	
LM state: pure SE	3.18***	3.14***	0.11	0.26
	(0.791)	(0.794)	(0.609)	(0.605)
hybrid SE	4.18***	4.12***	1.95**	1.87**
	(0.766)	(0.763)	(0.683)	(0.664)
freelancer	1.18	1.42*	-1.25	-0.95
	(0.724)	(0.724)	(1.029)	(1.029)
DGA	7.52***	6.17***	4.98***	3.35***
	(1.022)	(1.042)	(0.742)	(0.788)
cluster: weak LM attach. (SE)		-3.02***		-4.15***
		(0.215)		(0.828)

Table 3.9: Panel regression – disposable household income (continued)

	disposable household income (in thousands)			
	women		men	
benefit recipient (SE)		-4.25***		-4.13***
		(0.323)		(0.357)
short SE spells		-0.59**		1.07***
		(0.197)		(0.270)
employee to SE		0.99**		0.50
		(0.335)		(0.270)
pure SE		-0.80*		-1.38***
		(0.355)		(0.297)
hybrid SE		0.44		1.28
		(0.691)		(0.812)
DGA		11.30***		7.80***
		(1.098)		(0.541)
is SSE	-3.58***	-3.57***	-3.09***	-3.06***
	(0.732)	(0.731)	(0.550)	(0.547)
unknown SSE status	-3.56***	-3.62***	-1.12	-1.28*
	(0.702)	(0.701)	(0.632)	(0.636)
breadwinner	-5.13***	-5.14***	-4.75***	-4.80***
	(0.103)	(0.103)	(0.101)	(0.100)
main income source: profit	3.53***	3.49***	3.65***	3.60***
	(0.220)	(0.221)	(0.251)	(0.250)
freelancer	0.17	-0.00	-1.06	-1.47
	(0.643)	(0.651)	(1.015)	(1.024)
constant	27.09***	25.80***	23.32***	21.54***
	(1.861)	(1.811)	(1.996)	(1.953)
Non-SE clusters		✓		✓
Observations	577212	577212	564003	564003
Individuals	25965	25965	25441	25441
σ_u	7.182	7.164	7.576	7.556
σ_e	10.642	10.642	14.332	14.332
R^2 within	0.142	0.142	0.080	0.081
R^2 overall	0.178	0.191	0.109	0.119
R^2 between	0.248	0.274	0.191	0.221

Clustered standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including year and cohort fixed effects, and controlling for individual's demographic characteristics, non-self-employed labour market states, and other main income sources.

Finally, we include dummy variables indicating the household's *main income source* in each period. These include income as entrepreneur (profit) and income from "other work" (freelancers), as well as wage income (as base category), no income, benefits, pension, asset income, and other.

L_{it} is a set of *labour market state* dummies, while W_i controls for the *clusters*. The coefficients of the labour market states give the direct (current) impact, whereas the cluster coefficients say how income varies with different labour market careers, keeping current labour market state (and demographics etc.) constant. We take employees as the base category for labour market state, cluster, and main income source. Results should therefore be interpreted relative to employees. C_i and T capture cohort and year fixed effects, respectively.

As a baseline, we first estimate equation (3.1) without the cluster information. For both income measures, the year effects first increase over time and then decrease, returning to the 2006 levels in the last few years of IPO. For taxable income of males, the younger cohorts (5, 6, 7, 9) do significantly better than the older cohorts (particularly reference cohort 4). On the other hand, disposable household income is higher for cohorts 2 and 3, suggesting that older individuals have around 700–1000 euros more, ceteris paribus.

Table 3.8 shows the other regression results for men and women with taxable income as the dependent variable.[32] The results for disposable household income are shown in Table 3.9.

For all specifications, the cluster dummies are jointly significant. Furthermore, the estimates for the other coefficients (e.g. the effects of current labour market status) do not change much. DGAs have significant and positive coefficients for their cluster as well as their labour market state in all specifications. That is, individuals in the DGA cluster earn more, ceteris paribus, than other employees, even if they are not currently self-employed, and earn even more once they become self-employed. This is in line with the result on incorporated individuals in Humphries (2016). Similarly, for the clusters with weak labour market attachment or benefit recipients, we find large negative coefficients.

The picture is less clear for other self-employed clusters. For example, we find negative cluster effects for the pure self-employed (larger in absolute terms for men), but the direct effects of the labour market states differ by gender and income measure: While there is no significant effect on women's taxable income, we find a large negative effect for men. Vice versa, we find no significant effect on men's disposable household income but a strong positive effect for women, which dominates the negative cluster effect.

Hybrid self-employed generally have insignificant cluster coefficients. We find positive direct effects of hybrid self-employment for both genders on disposable household income, but a negative direct effect on men's taxable income. Similarly, we find no significant cluster effect for individuals that become self-employed later in their career except in the disposable income regression for

[32]Because of the large number of controls, we only present a subset of the coefficients. The coefficients for other labour market states and cluster variables are almost all statistically significant at the 0.001%-confidence level with expected (negative) signs; complete results are available upon request.

women where the effect is positive. This cluster effect is small compared to the direct effects we estimate.

Finally, we find opposite signs for the cluster with short self-employment spells for both income measures. Women in this cluster earn less than employees while men earn more. In periods where men are self-employed, this will generally mean that they earn a lower taxable income, as the cluster effect is smaller than the direct effect.

On average, self-employed individuals earn less in terms of individual taxable income, but with the exception of freelancers, they have a slightly larger disposable household income. Since the difference between taxable and disposable income of employees includes pension contributions, the higher disposable income of self-employed individuals may not mean they are better off if they do not save for their pension. Furthermore, solo self-employment has a large significantly negative coefficient.

Wealth

The lower panel of Table 3.7 reports household wealth adjusted for household size by cluster. As for the values by labour market status, the clusters with the self-employed have higher wealth at each quartile. Wealth still has high variation within clusters, as shown by the large inter-quartile ranges. Median household wealth of individuals with short SE spells or with a switch from employment to self-employment later in their career, is much lower than median wealth of pure and hybrid self-employed, for both men and women. Overall, differences across clusters are larger for the higher quartiles than for the first quartile. Household wealth of women in a given cluster is generally higher than that of men, probably because of the partner's income.

We estimate equation (3.1) with household wealth as the dependent variable, adding an additional control variable based on the standard industrial classification that is available for the self-employed starting from 2005: a dummy variable with value one for individuals active in the agricultural, forestry, or fishery sector. This aims at identifying farmers whose main wealth component often is their land. Table 3.10 shows the results.[33]

We find negative year effects starting around 2009/2010, which can be attributed to housing wealth – There are no significant year effects in the regression that excludes owner-occupied housing from wealth. Interestingly, we do not find any significant cohort effects on household wealth. This stands in contrast to Mastrogiacomo et al. (2016) who suggest that younger cohorts are less wealthy. While our regression sample excludes the oldest two cohorts due to data availability, we would have expected cohort effects for, e.g., cohort 3 and 4.

[33]As a robustness check, we also split wealth into wealth in owner-occupied housing and wealth without owner-occupied housing. The results are largely similar to those for total household wealth. These regressions also reveal that differences between self-employed clusters and employees are higher for non-housing wealth than for housing wealth (details available upon request).

Table 3.10: Panel regression – household wealth

	household wealth (in thousands)			
	women		men	
LM state: pure SE	78.73***	69.88***	62.32***	42.57***
	(16.862)	(17.222)	(8.308)	(8.327)
hybrid SE	72.46***	66.00***	44.73***	30.05***
	(17.759)	(17.926)	(8.641)	(8.569)
freelancer	57.83***	53.92***	21.20	6.02
	(15.564)	(15.547)	(17.625)	(17.651)
DGA	248.38***	188.13**	156.72***	105.26***
	(63.751)	(60.112)	(16.441)	(15.060)
cluster: weak LM attach. (SE)		14.50		25.82
		(8.672)		(20.602)
benefit recipient (SE)		-38.08***		-31.85***
		(10.466)		(8.220)
short SE spells		18.45**		37.28**
		(5.791)		(11.954)
employee to SE		25.94*		16.26*
		(10.589)		(7.362)
pure SE		42.41***		84.93***
		(12.457)		(7.924)
hybrid SE		47.62**		118.67***
		(16.842)		(36.037)
DGA		805.75**		312.50***
		(254.252)		(31.788)
is SSE	-52.67***	-50.13**	-19.20	-14.41
	(15.867)	(15.736)	(10.865)	(10.711)
unknown SSE status	-52.22**	-52.49**	-41.33***	-40.51***
	(19.979)	(19.775)	(11.563)	(11.626)
breadwinner	-4.25	-4.01	-5.97*	-5.82*
	(2.698)	(2.700)	(2.564)	(2.558)
main income source: profit	5.63	4.73	-0.20	-2.56
	(3.699)	(3.699)	(6.702)	(6.667)
freelancer	19.45*	18.11	-4.36	-4.88
	(9.360)	(9.501)	(15.257)	(15.356)
constant	182.10	195.62	34.23	-2.53
	(162.372)	(165.986)	(79.195)	(77.596)

Table 3.10: Panel regression – household wealth (continued)

	household wealth (in thousands)			
	women		men	
Non-SE clusters		✓		✓
Observations	168055	168055	162981	162981
Individuals	21186	21186	20621	20621
σ_u	484.498	484.428	279.246	279.119
σ_e	220.149	220.149	151.313	151.313
R^2 within	0.004	0.004	0.007	0.008
R^2 overall	0.039	0.050	0.085	0.103
R^2 between	0.044	0.062	0.099	0.123

Clustered standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including year and cohort fixed effects, and controlling for individual's demographic characteristics, agricultural sector, non-self-employed labour market states, and other main income sources.

As for income, we find that the cluster dummies are jointly significant. Unlike for income however, we find almost no differences between genders in signs and significance. All self-employed clusters except the first two have significant and positive coefficients. Current self-employment labour market states also have significantly positive coefficients, except freelancers for men. These direct effects have a clear ranking – DGAs have the highest household wealth and hybrid self-employed the lowest. The pattern of the cluster effects is clearer than for income: For both genders, pure and hybrid self-employed clusters have coefficients that are around twice as large as those of short self-employment spells or late switchers.

Compared to women, self-employed men seem to be in a better financial position. The magnitude of the coefficients for males who are always self-employed is around twice that of females. The results from the wealth regression therefore suggest that men who are always self-employed tend to accumulate more wealth than employees, and that this may partially offset shortages in their pension accumulation. The questions that remain are whether the approximately one hundred thousand euro extra that we find for them is sufficient to bridge the gap in pension accumulation, and how the other clusters fare with less wealth.

3.5 Pensions of the self-employed

In this section we analyse the pension accumulation of the self-employed, considering pension investments in the second pillar (mandatory occupational pensions) and the third pillar (voluntary private pensions). We do not consider the first pillar as our sample is restricted to individuals

Table 3.11: Percentage contributing to second and third pension pillar

	Women			Men		
	2nd Pillar	3rd Pillar	N	2nd Pillar	3rd Pillar	N
Weak LM attach. (SE)	13.14	0.95	41698	25.17	7.49	2805
Benefit recipient (SE)	13.83	2.02	8120	18.40	8.03	5892
Short SE spells	49.61	5.94	51543	67.09	17.53	46710
Employee to SE	44.91	8.69	16007	52.41	18.32	23745
Pure SE	8.59	11.24	20739	10.55	22.85	42192
Hybrid SE	39.13	6.47	7136	48.82	20.23	9496
DGA	21.89	9.94	6561	25.46	25.02	23525
Low LM attach.	5.13	0.25	73707	12.25	3.01	1861
Benefit recipient	18.18	1.76	55894	20.86	3.63	33943
Long benefit rec.	42.14	3.91	61851	53.57	10.97	42238
Employee	69.32	7.10	217481	77.96	17.07	314903
Pensioner	9.88	5.60	3340	17.77	9.52	2605

Note: Observations are counted as individuals per year per cluster. The reported shares and observations are for all individuals in our sample aged 24–60.

who spend (almost) their complete working life in the Netherlands and therefore receive the full public pension.

Table 3.11 shows the share of individuals who make contributions to the second and third pillar in a given year, by gender and cluster. Women contribute less often than men in the same cluster, in line with Table 3.4. The cluster of DGAs contributes more often than what Table 3.4 would suggest. Approximately half of the DGA cluster spends the first half of their career as employees and accumulate an occupational pension during that time (Table 3.4), but they do not pay occupational pension premiums when they are DGA (Table 3.4). Self-employed individuals in the cluster with weak labour market attachment and those who are benefit recipients during most of their working life are comparable to their non-self-employed counterparts, though self-employed men in these groups prepare somewhat better for retirement through the third pillar. For men, the difference between shares of the short self-employment spells cluster and the cluster of employees is what we would expect given the median length of their self-employment spells.[34] For women the gap of approximately 20 percentage points is larger than expected. Neither men nor women bridge the gap in the second pillar by participating in the third pillar – Their 3rd pillar participation does not differ from that of employees. The same applies to the cluster that becomes self-employed later in their career, who have an even larger participation gap in the second pillar. A similar result holds for the clusters of pure and hybrid self-employed. Men in these clusters participate in the third pillar more often, but these differences are small compared

[34]As a rough calculation we would expect a gap of around 4.5 percentage points per year in self-employment, since we have approximately 22 years of observations per individual on average.

to the differences in second pillar participation. For women, if they cannot count on a partner's pension, the outlook is even bleaker, considering that their third pillar participation is very low.

The administrative data on pension participation ("pensioendeelnemingen") contain information on second pillar entitlements. It covers the years 2005–2014 and is based on an annual survey among a representative sample of pension funds. Not all these funds are present in every year due to non-response. Because the data do not cover all pension funds we do not know whether an individual for whom we do not observe a pension account never participated or whether their pension fund is not in the sample. Moreover, if someone has more than one pension fund, the observed entitlement may not be their total entitlement. Moreover, due to the fluctuations in sampled pension funds across years, the time series of each individual gives excessive noise, preventing the use of panel data models.

To account for these data issues, we proceed as follows. We first add up the pension entitlements over all observed funds for each individual within a year. We then deflate monetary values and take the median by individual across all years. Overall, this gives pension information for about half of the individuals in our sample. Table 3.12 reports sample statistics by cluster and gender for three variables.[35] The first is pension contribution years, converted to full-time equivalents. The second variable is the pension annuity to which individuals are entitled given their accumulated pension rights at the time of reporting (calculated by the pension funds). Third, the current capital entitlement is reported (also based on pension funds' calculations).

Note that all three variables refer to pension rights accumulated at the time of the survey, which is, for most individuals, quite a long time before retirement. Pension accumulation between the survey date and retirement will most likely be lower for self-employment clusters than for employees, so that the ultimate gaps will be larger than those considered here.

Table 3.12 shows the same picture for all three variables. Employees have much higher medians than any of the self-employed clusters and women almost always have lower values than men. The two self-employed clusters with weak ties to the labour market are comparable to their non-self-employed counterparts. For all other self-employed clusters we find large gaps in comparison with the employees' values, particularly for men. For the cluster with short self-employment spells, the gap in contribution years corresponds approximately to the median self-employment spell length for women. For men the gap is much larger, in particular for the third quartile. The pure self-employed, who by definition spend at best a few years in employment, have almost no pension entitlements and will need to rely on a third pillar annuity or on other private wealth.

Table 3.13 reports the results of a regression that shows the ceteris paribus associations between the pension accrual variables and the clusters, similarly to the income and wealth regressions. We use the same demographic characteristics as in the income regressions and take their values from the last year in which we have information on an individual's second pillar

[35] Since we exclude individuals older than 60, cohorts 1 and 2 are not included.

Table 3.12: Occupational pension entitlements

	Women				Men			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Pension contribution years								
Weak LM attach. (SE)	2.40	0.88	5.40	5.23	5.25	1.31	9.62	0.41
Benefit recipient (SE)	2.50	0.96	7.02	1.22	4.66	1.69	9.13	0.84
Short SE spells	6.25	2.74	10.62	10.88	8.65	3.17	17.78	8.87
Employee to SE	6.60	2.92	11.90	3.32	9.47	3.74	17.13	4.62
Pure SE	2.55	0.94	5.80	2.10	4.98	1.77	10.00	4.80
Hybrid SE	4.87	2.11	9.00	1.48	7.23	2.75	16.80	1.56
DGA	4.57	1.49	10.63	0.85	6.86	2.25	14.64	3.23
Weak LM attach.	2.80	1.00	5.75	4.76	5.44	2.53	16.58	0.19
Benefit recipient	5.13	1.60	11.38	7.39	7.73	1.86	19.85	4.43
Long benefit rec.	6.70	2.90	12.36	11.66	10.84	3.33	23.44	7.43
Employee	9.95	5.60	15.40	50.73	14.42	5.05	27.82	63.20
Pensioner	18.43	8.17	27.91	0.39	34.99	29.84	38.25	0.42
Pension annuity (in thousands)								
Weak LM attach. (SE)	0.31	0.09	1.07	5.33	0.84	0.17	3.96	0.42
Benefit recipient (SE)	0.40	0.10	1.74	1.27	0.70	0.14	1.81	0.90
Short SE spells	1.65	0.48	3.93	10.87	3.36	0.80	8.35	8.93
Employee to SE	2.12	0.60	4.83	3.33	3.19	0.94	6.26	4.70
Pure SE	0.43	0.11	1.31	2.24	0.81	0.19	2.29	5.14
Hybrid SE	1.55	0.36	3.19	1.49	2.92	0.70	7.54	1.56
DGA	0.87	0.18	3.39	0.85	1.44	0.34	5.53	3.30
Weak LM attach.	0.36	0.11	1.13	4.92	0.54	0.17	2.85	0.20
Benefit recipient	0.95	0.16	2.94	7.57	1.05	0.17	4.19	4.55
Long benefit rec.	1.54	0.47	3.75	11.79	3.17	0.64	7.58	7.41
Employee	3.43	1.48	6.23	49.94	5.89	1.92	11.90	62.50
Pensioner	5.03	1.49	9.34	0.41	14.75	10.10	20.00	0.40
Capital entitlement (in thousands)								
Weak LM attach. (SE)	2.07	0.59	6.90	5.32	4.56	0.87	19.90	0.42
Benefit recipient (SE)	2.88	0.54	11.83	1.27	3.73	0.74	10.26	0.90
Short SE spells	10.04	2.82	25.40	10.87	15.72	3.82	46.98	8.92
Employee to SE	10.84	3.28	30.88	3.33	14.37	4.53	33.77	4.70
Pure SE	2.51	0.62	7.81	2.24	3.90	0.96	10.74	5.14
Hybrid SE	8.46	2.43	17.48	1.49	12.77	2.71	38.19	1.56
DGA	6.22	1.25	16.90	0.85	7.18	1.84	30.51	3.30
Weak LM attach.	2.40	0.66	7.70	4.92	2.22	0.56	14.22	0.20
Benefit recipient	6.41	0.89	21.84	7.57	5.46	0.83	27.54	4.55
Long benefit rec.	10.17	2.91	27.68	11.79	17.48	3.10	50.48	7.41
Employee	17.59	7.23	40.28	49.94	27.66	7.71	71.47	62.50
Pensioner	31.49	11.86	71.67	0.41	82.62	52.28	131.28	0.40

Note: All values are in euro and deflated with CPI data from The World Bank World Development Indicators (base year 2010). The statistics are calculated over individuals by cluster for the median values observed for individuals in the sample aged 24–60. (Total observation count pension contribution years: 13,648 (w), 14,450 (m), pension annuity: 14,046 (w), 14,845 (m), capital entitlement: 14,047 (w), 14, 847 (m).) Observation counts differ across measurements of pension fund participation due to not all pension funds reporting all values for all individuals' accounts. Pension contribution years are measured in FTE years.

Table 3.13: Regression – accruals in 2nd pillar

	Annuity (in thousands)		*Capital (in thousands)*	
	women	men	women	men
cluster: weak LM attach. (SE)	-3.91***	-4.94***	-26.83***	-30.38***
	(0.168)	(1.060)	(1.383)	(7.568)
benefit recipient (SE)	-3.78***	-5.69***	-26.15***	-35.10***
	(0.327)	(0.728)	(2.690)	(5.197)
short SE spells	-1.88***	-1.29***	-12.57***	-8.45***
	(0.122)	(0.243)	(1.003)	(1.734)
employee to SE	-1.35***	-2.71***	-8.42***	-14.16***
	(0.218)	(0.380)	(1.799)	(2.710)
pure SE	-3.66***	-6.04***	-24.09***	-36.91***
	(0.262)	(0.389)	(2.161)	(2.776)
hybrid SE	-3.11***	-2.41***	-22.23***	-13.33**
	(0.308)	(0.577)	(2.536)	(4.120)
DGA	-2.46***	-3.89***	-18.36***	-26.98***
	(0.392)	(0.385)	(3.228)	(2.746)
weak LM attach. (no SE)	-3.70***	-5.13***	-25.45***	-31.10**
	(0.172)	(1.535)	(1.420)	(10.959)
benefit recipient (no SE)	-2.78***	-4.21***	-17.44***	-23.60***
	(0.157)	(0.374)	(1.289)	(2.671)
long spells as benefit recipient	-2.20***	-3.49***	-14.43***	-22.61***
	(0.121)	(0.283)	(0.998)	(2.023)
pensioner	-0.01	0.61	3.89	2.91
	(0.571)	(1.099)	(4.699)	(7.848)
breadwinner	1.96***	1.73***	14.16***	10.96***
	(0.091)	(0.201)	(0.749)	(1.437)
main income source: profit	0.41***	0.70*	2.87**	3.32
	(0.120)	(0.276)	(0.987)	(1.972)
freelancer	-0.33	-2.16	-1.46	-15.58
	(0.707)	(1.469)	(5.822)	(10.491)
constant	-1.91	1.29	51.70*	126.57***
	(3.120)	(5.285)	(25.686)	(37.733)
Observations	14046	14845	14047	14847
R^2 adjusted	0.206	0.226	0.233	0.262
RMSE	4.213	8.096	34.693	57.802

Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including cohort fixed effects, and controlling for individual's demographic characteristics, for when we last observe an individual.

accrual. We add year effects based on the last year we observe an individual in the pension data to control for level differences due to different timing in observations. Because of the limitations of the data, we do not include the labour market state dummies.

Table 3.14: Probit regression – participation in 3rd pillar

	participates in 3rd pillar (in thousands)			
	women		men	
Pays into 2nd pillar	0.52***	0.48***	0.29***	0.29***
	(0.018)	(0.018)	(0.016)	(0.017)
LM state: pure SE	1.01***	0.87***	0.54***	0.51***
	(0.068)	(0.071)	(0.047)	(0.049)
hybrid SE	0.54***	0.52***	0.22***	0.18***
	(0.072)	(0.073)	(0.048)	(0.048)
freelancer	0.06	0.23**	0.13	0.15
	(0.080)	(0.084)	(0.091)	(0.089)
DGA	0.87***	0.62***	0.50***	0.26***
	(0.092)	(0.091)	(0.049)	(0.050)
cluster: weak LM attach. (SE)		-0.47***		-0.25*
		(0.062)		(0.112)
benefit recipient w/ SE		-0.38***		-0.19*
		(0.095)		(0.077)
short SE spells		-0.05		0.02
		(0.033)		(0.025)
employee to SE		0.02		0.02
		(0.048)		(0.034)
pure SE		0.09		0.02
		(0.051)		(0.032)
hybrid SE		-0.19*		0.06
		(0.082)		(0.051)
DGA		0.21**		0.26***
		(0.083)		(0.035)
is SSE	-0.34***	-0.32***	-0.25***	-0.24***
	(0.067)	(0.067)	(0.042)	(0.042)
unknown SSE status	-0.46***	-0.45***	-0.12***	-0.13***
	(0.057)	(0.057)	(0.036)	(0.036)
breadwinner	0.22***	0.21***	0.13***	0.12***
	(0.020)	(0.020)	(0.017)	(0.017)
main income source: profit	0.11***	0.10***	0.17***	0.16***

Table 3.14: Probit regression – participation in 3rd pillar (continued)

	participates in 3rd pillar (in thousands)			
	women		men	
	(0.029)	(0.030)	(0.025)	(0.025)
freelancer	0.04	0.03	0.29***	0.28***
	(0.071)	(0.073)	(0.047)	(0.047)
constant	-4.72***	-4.83***	-4.54***	-4.73***
	(0.376)	(0.377)	(0.272)	(0.272)
Non-SE clusters		✓		✓
Observations	467380	467380	548536	548536
Individuals	24017	24017	25411	25411
log likelihood	-87366.566	-86542.144	-208083.694	-207264.314

Clustered standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including year and cohort fixed effects, and controlling for individual's demographic characteristics, non-self-employed labour market states, and other main income sources.

Table 3.15: Tobit regression – contributions to 3rd pillar

	contributions to 3rd pillar (in thousands)			
	women		men	
Pays into 2nd pillar	4.77***	4.40***	2.35***	2.28***
	(0.320)	(0.303)	(0.164)	(0.163)
LM state: pure SE	9.92***	9.03***	5.53***	5.32***
	(0.796)	(0.828)	(0.455)	(0.477)
hybrid SE	5.72***	5.42***	3.15***	2.64***
	(0.759)	(0.779)	(0.458)	(0.457)
freelancer	-0.13	1.43	2.17**	2.06*
	(0.767)	(0.796)	(0.819)	(0.818)
DGA	7.70***	3.95***	4.74***	1.36**
	(0.923)	(0.894)	(0.443)	(0.451)
cluster: weak LM attach. (SE)		-4.05***		-2.13*
		(0.627)		(0.977)
benefit recipient (SE)		-3.65***		-1.31
		(0.917)		(0.726)
short SE spells		-0.45		0.51*
		(0.294)		(0.204)

Table 3.15: Tobit regression – contributions to 3rd pillar (continued)

	contributions to 3rd pillar (in thousands)			
	women		men	
employee to SE		0.04		0.11
		(0.442)		(0.284)
pure SE		0.22		-0.03
		(0.496)		(0.295)
hybrid SE		-0.65		0.78
		(0.783)		(0.460)
DGA		3.83***		3.79***
		(0.833)		(0.323)
is SSE	-3.10***	-2.92***	-2.97***	-2.86***
	(0.634)	(0.630)	(0.399)	(0.395)
unknown SSE status	-3.25***	-3.16***	-0.47	-0.64
	(0.534)	(0.534)	(0.346)	(0.346)
breadwinner	1.91***	1.84***	1.09***	1.02***
	(0.209)	(0.206)	(0.154)	(0.154)
main income source: profit	1.30***	1.22***	2.11***	2.02***
	(0.270)	(0.269)	(0.222)	(0.220)
freelancer	0.62	0.43	3.01***	2.94***
	(0.689)	(0.697)	(0.428)	(0.427)
constant	-46.87***	-47.57***	-43.86***	-45.61***
	(4.364)	(4.386)	(2.674)	(2.695)
Non-SE clusters		✓		✓
var(e.c_p3pen_defl)	102.08***	101.48***	94.56***	93.77***
	(12.240)	(12.172)	(6.949)	(6.919)
Observations	577333	577333	564127	564127
Individuals	25965	25965	25441	25441
log likelihood	-149235	-148545	-426205	-425207

Clustered standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Regression including year and cohort fixed effects, and controlling for individual's demographic characteristics, non-self-employed labour market states, and other main income sources.

Both regressions show the same picture and the cluster coefficients confirm what we have learned from Table 3.12. Self-employed who mostly receive benefits are worse off than their non-self-employed counterparts. Among the other self-employed clusters the pure self-employed are worst off and will on average receive substantially lower pension annuities after their retirement

– in the case of men a good 500 euro less per month. The cluster effect for career changers shows that the estimated average gap in pension capital for men is of similar magnitude to the corresponding coefficient estimated in the wealth regression, suggesting that this group may be particularly ill-prepared for retirement.

To investigate whether self-employed individuals use of the third pillar to counter-balance gaps in the second pillar, we analyse the probability of making a payment into the third pillar using a probit model, and estimate a tobit model to understand how much they invest. In both models we use the same controls as in the income regressions as well as a dummy that takes value one when a positive employee contribution to the second pillar is observed.

Table 3.14 shows the results of the probit model. The dependent variable is a dummy that takes value one if a payment to a tax-deductible third pillar pension plan is observed and zero otherwise. Since individuals with no income do not make contributions to the third pillar, these individuals are excluded. Four results stand out: first, individuals who make a contribution to the second pillar in the same year are more likely to save in the third pillar. Second, among the different self-employed clusters, only DGAs are more likely to make a third pillar contribution than employees, whereas we find significantly negative effects for the weak labour market attachment and benefit clusters (1 and 2). Third, current solo self-employed contribute less often than others. Finally, individuals that live in households deriving their main income from pure or hybrid self-employment (main income source: profit) are more likely than their other self-employed peers to make payments to the third pillar. The results of the tobit regressions are shown in Table 3.15. The results are largely in line with the probit results. Again, we find a positive association with contributing to the second pillar. As before, DGAs are the positive outliers. They are the only cluster with a large positive (and significant) coefficient, contributing much more than employees. The coefficients on labour market status tell us that once individuals are active as self-employed, their payment to the third pillar increases compared to employees. Furthermore, pure self-employed have the largest coefficients – they make the largest contributions to the third pillar. Again, we also find a negative effect for SSE.

3.6 Conclusion

To analyse the heterogeneity in labour market trajectories of the self-employed in the Netherlands, we have classified the self-employed into different clusters using sequence analysis. We have identified the following seven distinct clusters of self-employment across all cohorts considered: (1) self-employed with weak labour market attachment, (2) self-employed that spend a large portion of their trajectory as benefit recipients, (3) employees with short self-employment spells, (4) employees that switch to self-employment later in their career, (5) always pure self-employed, (6) always hybrid self-employed, and (7) DGAs (director and major shareholder).

Unlike cross-sectional studies, sequence analysis provides a holistic view that incorporates

the whole labour market trajectory of each individual. We find that half of the individuals that are ever self-employed, (those in clusters 1–3) remain so for only a few years. In any given calender year, approximately one third of the self-employment rate can be attributed to these short-term self-employed.

For wealth, we find no statistically significant cohort effects within clusters. This suggests that the finding in earlier studies that the current self-employed are less wealthy than those in older cohorts is due to a shift in the share of different clusters in the total self-employed population rather than to a shift within clusters. In line with this, we find that the share of individuals that become self-employed later in their career has increased over time, and this cluster has accumulated less wealth than the other self-employed.

In terms of income, wealth, and pension accumulation, we show that clusters 1 and 2 are worse off than employees and do not differ a lot from their counterparts amongst the never self-employed. Their average income is lower than that of employees, also in the periods when they are self-employed, they accumulate less wealth, and, particularly, do not invest in voluntary (or mandatory) pensions. Policies targeted at the self-employed only are unlikely to affect their post-retirement outcomes – These individuals would benefit more from policies targeted at everyone at the lower end of the labour market.

Individuals in cluster 3 have unexpectedly large gaps in their occupational pension savings. They accumulate more financial wealth than employees, but this will probably not be sufficient to cover these gaps. Similar results are found for those who change to self-employment halfway through their career (cluster 4). Compared to employees, they neither invest more in a voluntary pension nor in other (non-pension) wealth. On the other hand, their income is similar to that of employees. This suggests that further research into the adequacy or inadequacy of their pensions is necessary.

The long-term self-employed, clusters 5–7, hold larger amounts of wealth than employees or clusters 3 and 4. Clusters 5 and 7 have almost no accruals in the second pension pillar, but the majority of them may still be well prepared for their retirement.

There is substantial difference by gender: the wealth differential between self-employed clusters and employees is much larger for men than for women. Women in the self-employment clusters also have less pension accruals in the mandatory occupational pensions and participate less in voluntary private pensions. They may often be able to rely on a partner as main earner in the household, as long as the couple remains intact. Studying household composition dynamics (i.e. divorce or widowhood) and its consequences for pension adequacy is left for future research.

Our findings on DGAs replicate results found for incorporated self-employed in the US and Sweden. Given that we still find large variation in wealth *within* our clusters and that there are different unincorporated business structures in the Netherlands, it seems worthwhile to study if using more information on the legal form can help to further differentiate among clusters of self-employed. The same holds for the solo self-employed. We find that they are worse off

than the other self-employed but the short time period for which their information is available, did not allow to study them in detail. More work on this seems useful. Finally, better data on pension accruals is needed. Because of the current data limitations our results give us only a lower bound on accumulation of an occupational pension. Integral pension data is needed for a clearer understanding of how pension accruals differ across clusters.

Appendices

3.A Additional tables

Table 3.A.1: Number of self-employment spells and their length by cluster

	Number of spells			Spell length (years)			Sample
	1 spell	2 spells	3 or more	median	Q_1	Q_3	share
Women							
Weak LM attach.	65.82	23.72	10.46	2	1	5	27.30
Benefit recipient	72.99	21.55	5.46	2	1	4	5.34
Short SE spells	76.36	18.74	4.89	2	1	4	33.24
Employee to SE	76.37	19.02	4.61	7	3	11	10.65
Pure SE	76.25	19.19	4.56	15	7	21	14.47
Hybrid SE	72.17	22.33	5.50	11	5	19	4.74
DGA	55.76	29.14	15.11	8	3	13	4.27
Men							
Weak LM attach.	55.17	30.17	14.66	3	1	6	1.69
Benefit recipient	72.47	21.60	5.92	2	1	5	4.19
Short SE spells	74.41	19.56	6.03	3	1	5	29.06
Employee to SE	76.59	17.56	5.85	7	4	11	14.73
Pure SE	78.78	16.86	4.35	17	8	23	28.85
Hybrid SE	69.87	22.10	8.04	11	5	20	6.54
DGA	50.98	32.49	16.54	8	3	14	14.93

Note: Shares calculated by gender within the subsample that has at least one self-employment spell. We define any series of years spent consecutively in any type of self-employment (hybrid and pure self-employment, freelancer and DGA) as a self-employment spell. The spell ends once an individual spends at least one year not in self-employment. Hence individuals may have more than one self-employment spell.

Table 3.A.2: Number of self-employment spells and their length by cohort

	Number of spells			Spell length			Sample
	1 spell	2 spells	3 or more	median	Q_1	Q_3	share
Women							
Cohort 1	76.87	18.51	4.63	2	1	7	4.31
Cohort 2	68.22	24.89	6.89	3	1	9	6.91
Cohort 3	68.21	23.78	8.01	3	1	10	12.26
Cohort 4	68.10	21.31	10.59	3	1	9	11.74
Cohort 5	69.29	21.89	8.82	3	1	9	14.79
Cohort 6	71.39	21.72	6.89	3	1	9	16.03
Cohort 7	75.00	19.31	5.69	3	1	8	15.10
Cohort 8	75.65	19.43	4.92	3	2	7	11.22
Cohort 9	82.73	15.86	1.41	3	2	7	7.64
Men							
Cohort 1	69.09	25.68	5.23	7	2	13	6.43
Cohort 2	67.84	22.86	9.29	7	2	16	7.86
Cohort 3	65.87	22.44	11.69	7	2	17	11.00
Cohort 4	67.51	23.35	9.14	8	3	18	10.39
Cohort 5	68.15	21.98	9.88	6	2	15	11.83
Cohort 6	71.62	20.65	7.73	7	3	15	14.36
Cohort 7	74.95	19.17	5.87	7	2	14	15.92
Cohort 8	76.94	18.38	4.68	6	2	12	12.80
Cohort 9	81.06	16.30	2.64	5	2	10	9.41

Note: Shares calculated by gender within the subsample that has at least one self-employment spell. We define any series of years spent consecutively in any type of self-employment (hybrid and pure self-employment, freelancer and DGA) as a self-employment spell. The spell ends once an individual spends at least one year not in self-employment. Hence individuals may have more than one self-employment spell. Cohort 1: birth years 1936–1940, ..., cohort 9: birth years 1976–1980.

Table 3.A.3: Demographic characteristics across clusters; observations with pension participation

		in %						
Cluster	Age	Men	Bread-winner	No partner†	Dutch	Western	Non-Western	Sample share
with self-employment								
1	50.23	6.43	12.81	13.35	88.78	7.08	4.13	3.83
2	50.16	40.85	61.05	53.46	75.74	9.20	15.07	1.21
3	47.50	47.46	55.20	31.86	85.91	9.12	4.97	8.91
4	45.28	59.06	74.52	36.80	88.63	5.81	5.56	3.73
5	47.34	66.82	73.30	33.40	88.73	6.83	4.43	5.66
6	46.97	56.55	75.59	34.46	87.42	8.42	4.17	1.51
7	49.10	78.35	71.69	25.35	90.36	7.60	2.04	2.71
without self-employment								
8	53.70	2.77	8.20	12.48	85.37	7.94	6.68	5.91
9	49.03	36.71	63.48	59.31	72.70	9.59	17.71	7.42
10	50.84	39.09	53.97	40.33	82.00	8.37	9.63	9.04
11	46.48	57.94	64.99	36.34	87.47	7.68	4.85	49.65
12	62.76	44.24	71.73	43.24	85.96	10.97	3.07	0.42
Overall	47.92	49.30	58.80	35.48	85.71	7.93	6.36	

Note: The statistics are calculated over all individual-year observations for each labour market state.
† Includes singles, divorced and widowed individuals.

3.B Cluster correspondence across cohorts

The upper part of Table 4.C.1 gives an overview of how we sort the different groups based on Ward's algorithm into the seven clusters of self-employed across all cohorts as used in our analysis. The lower part lists the clusters for the individuals that are never self-employed based on their 6-group Ward solution.

Table 3.B.1: Correspondence between clusters used in analysis and Ward's solution

Individuals with at least one year of self-employment

	Cohorts								
Clusters in analysis	1	2	3	4	5	6	7	8	9
Weak labour market attachment:	1, 2	1, 2	1, 3	1, 2	1, 2	1, 2	1, 2	7	6, 5
Freelancer:	2	2	3	2	2		2		5
Benefit recipient:	9	5	5	4	5	5	3	6	4
Short SE spell(s):	3, 4	8, 9, 10	2, 7, 8, 11	6, 7, 8	9, 10	3, 10	8, 9, 10	4, 5, 12	11, 12
Employee to SE:	5	8, 11	9, 10	9, 10	7, 8	6, 7	6, 7	9, 10, 11	10
(Full-time) SE	10, 11, 12	3, 4, 12	12	11, 12	11, 12	11, 12	11, 12	1, 2	1, 2, 3
Hybrid SE:	6, 7	6	6	3	6	4	4	3	8, 9
DGA:	8	7	4	5	3, 4	8, 9	5	8	7

Individuals with no self-employment

	Cohorts								
	1	2	3	4	5	6	7	8	9
Weak labour market attachment:	1	1	1	1	1	1	1	5	1
Benefit recipient:	2	2	2	2	4	3, 4	3, 4	6	2, 3
Long spells as benefit recipient:	4	5	3	3	2	2	2	3, 4	6
Employee:	3, 6	4, 6	4†, 5, 6	5, 6	3, 5, 6	5, 6	, 5, 6	1, 2	4, 5
Pensioner:	5	3	4†	4					

Note: The cluster correspondence is for a 12-group partitioning for the self-employed and a 6-group partitioning for the non-self-employed. The partitioning is based on Ward's Method and implemented with Stata.

†In cohort 3 individuals that are classified as pensioners for almost all of the observed years in the non-self-employed data do not build a separate group in the six-group Ward solution. They are instead contained with other individuals who retire relatively early within cluster 4. We therefore manually separate them and sort those into the "pensioner" cluster that spend at least 10 of the 19 years in retirement. This affects approximately one third of the original group.

3.C Self-employed and the income tax

The Dutch system distinguishes between three types of self-employed: director and major shareholder (DGA), entrepreneur and freelancer. Both the chosen legal form and – in the case of the latter two – the recognition by the tax authority as entrepreneur determine how their profits are taxed.

The legal distinction is between incorporated and unincorporated businesses. DGAs are incorporated while entrepreneurs and freelancers are unincorporated. DGAs and their firm are two separate legal entities. DGAs pay income taxes on the salary they receive from their business, while the profits from their firm are subject to the lower corporate tax. Since it would be attractive for DGAs to pay themselves dividends instead of a salary, the law requires a minimum salary for DGAs ("gebruikelijkloonregeling") which was 45,000 euro in 2017. Exceptions are made for new enterprises, part-time workers or firms that make a structural loss.

Entrepreneurs and freelancers on the other hand are the same legal entity as their business. They do not receive a separate wage and the profit of their business is taxed under the income tax. If the tax authority recognises the self-employed as an entrepreneur ("IB ondernemer") they are allowed to make certain deductions from their profit before taxes. Their income is registered as *profit from company* ("winst uit onderneming"). Those not recognised as entrepreneurs will fill in their income in the tax returns as "income from other work" ("inkomen uit overige arbeid"). This is also how we distinguish between the two groups in the income data (based on tax records).

In order to qualify as an entrepreneur for the income tax, several conditions from a set of conditions need to be met. The set includes: the business needs to have a significant profit and cannot make a structural loss; independence of the business (i.e. no dependence on one client only, work can be chosen freely); the company should be advertised externally, and the owner carries the entrepreneurial risk. There is no specific set of rules that is decisive and the decision is made on a case-by-case basis.[36]

Freelancers are (almost) always solo self-employed while DGAs and the self-employed may also be employers.

3.D The Dutch pension pillars

The Dutch pension system essentially consists of three pillars: (1) the state pension as defined by the general old age pensions act ("Algemene Ouderdomswet", abbreviated AOW), (2) the occupational pension system which is regulated by the pension law, and (3) individual private pensions.

The first pillar is a pay-as-you go system, funded by government funds and payroll taxes. Everyone living and/or working in the Netherlands during the 50 years before they reach their

[36]Someone who is not an entrepreneur for the income tax may still have to pay value added taxes on their products or services.

statutory pension age is entitled to AOW. Individuals build up rights to 2% of the full pension for each of those 50 years. The full pension is tied to the minimum wage. Married and cohabiting couples each receive 50% of the minimum wage and those who live alone receive a pension slightly above 70% of the minimum wage. Until 2016 the statutory pension age was 65. In of 2012 the Dutch government announced that the pension age would rise to 67 and beyond if life expectancy would increase.

The second pillar consists of occupational pensions that are collective pension schemes. They are linked to either companies or specific industries and managed by pension funds or insurance companies. The government can enforce participation in a pension fund if a company decides to provide a pension scheme to its employees. Similarly, the government can mandate an industry-wide pension fund. The majority of second pillar pensions are built up via employer based pension plans. In total about 90% of all employees are required to contribute to a pension plan[37]. Examples of industry-specific pension funds are painters and independent professionals like e.g. general physicians. The majority of the self-employed, however, is not active in sectors with a mandatory pension scheme.

The third pillar is formed by voluntary saving schemes, usually taking the form of an annuity or life insurance and provided as commercial savings products. The third pillar is meant to bridge gaps in the first and second pillar – Payments in a third pillar product are tax-deductible for individuals that have low pension accruals (resulting in a projected pension that is less than 70% of their average income from work).

3.E Additional details on income data

IPO

The core of our analysis uses the Income Panel Study (IPO), which follows a representative sample of the Dutch population over time. When IPO was started as a panel in 1989 the sample was drawn from the population aged 10 years and older. From 1991 until 1993 the sample was refreshed with individuals aged 15 years and older, as well as with immigrants. As of 1993 the sample includes all ages and individuals from the pool of newborns and immigrants are added to the sample every year to account for births and immigration. To add younger individuals for the years before 1993, a number of household members of existing sample participants were added.

IPO records detailed information on the individuals' income from many different sources for all the years. Because tax records are the main source of information, a first break in the series occurs due to the major reform in the Dutch tax system in 2001, when a new tax law (Wet inkomestenbelasting 2001) was introduced. The new law mostly lead to minor changes in recorded income figures and to a different break down of income source categories. A second break occurs in 2011, when a revision in the income statistics took place. This break lead to a

[37] See https://www.rijksoverheid.nl/onderwerpen/pensioen/opbouw-pensioenstelsel

better measurement of income. For most income sources, these changes are small. Exceptions are the market rental value of real estate and income from substantial business interest. Because we do not use either of these two income sources to define an individual's labour market state, this break has no impact on our results.

To analyse the consequences of the break due to the 2001 tax reform for our labour market states, we exploit the fact that IPO provides datasets in both the pre- and post-reform format for the year 2000. We find that less than 2% of all individuals are categorised differently after the reform (6.6% of the self-employed and hybrid self-employed). The differences mostly result from better data sources underlying the more recent IPO, raising the number of individuals classified as self-employed.

IPI and INPATAB

Because the technological advances in computation and data storage facilities made the use of integral data much easier CBS started providing integral individual income (IPI) starting from the year 2003. With the income statistics revision in 2011 IPI was then replaced by a new integral income data set: "INPATAB". Both IPI and INPATAB are also mostly based on tax records and cover the same income variables as IPO with some minor differences. Unlike IPO the integral datasets do not contain any demographic and household characteristics. They therefore do not contain information on disposable or gross household income. One advantage of IPI and INPATAB is that they, starting from 2005, include information on whether self-employed individuals are solo self-employed or not, as well as the industry in which their business is active.

CHAPTER 4

SELF-EMPLOYMENT CAREERS AND FINANCIAL WELL-BEING IN OLD AGE IN EUROPE

This chapter is based on the identically entitled working paper

This paper uses data from the *Survey of Health, Ageing and Retirement in Europe* to identify different life-cycle patterns of self-employment in Europe for the cohorts born between 1931 and 1955. I explore how the different self-employment types are related to individuals' financial well-being in old age. I find in particular that individuals that spend a long part of their career in self-employment have lower household income and pension income once they are 60+ years old. This result is also reflected in the results from an analysis of indicators on financial well-being.

4.1 Introduction

A decade ago Choi (2009) came to the conclusion for most OECD countries' *pension provision for the self-employed is a matter of practical implementation of existing schemes rather than overhauling pension rules for these schemes*. Ten years on, sufficiency of the pensions of the self-employed is however in many countries still an issue. For example, the 2017 OECD report on pensions argues that in order to decrease income inequality in old-age countries should, among other things, *[increase] pension coverage, especially for the self-employed*. (OECD, 2017, p.29) Their newest report also points out that *In several OECD countries, all or some types of self-employed workers are exempt from enrolling in earnings-related pensions that are mandatory for dependent employees, increasing the risk of low old-age income.*(OECD, 2019, p.66) Similarly,

policy makers in Europe are also concerned with the self-employed: The European Commission's report on EU pension adequacy, for example, points out that the retired self-employed have a higher risk of income poverty and are more exposed to financial hardship. (European Commission, 2018, p.16)

Because pension entitlements are not just dependent on a country's institutions but also on an individual's decisions and actions throughout their working life, the analysis of individual outcomes in old-age should take account of the whole labour market history of the individual. One method that lends itself for such a task is sequence analysis, a method from sociology first introduced by Abbott (1983). With the help of sequence analysis I therefore construct different clusters of self-employment, based on individuals labour market trajectory between ages 15 and 60', using the retrospective Job Episodes Panel from the Survey of Health, Ageing and Retirement in Europe (SHARE).

There exist different studies that use sequence analysis to relate career patterns with measures of well-being after retirement. Most of these studies however focus on the later years in individuals' careers and do not look at the whole trajectory. Munnell et al. (2019) e.g. study individuals in the US from age 50 to 62 using the Health and Retirement Study (HRS) to asses how these individuals use non-traditional jobs and how their late career trajectories relate to income and depression. The authors find that those who consistently worked in non-traditional jobs end up with less retirement income and are also more likely to be depressed than other workers. McDonough et al. (2017) also use data from the HRS, to study trajectories between ages 52 and 69 for the cohort born 1931–1941. They study the association between these trajectories and self-rated health in the early 70s. They find, for example, that stable employment patterns are associated with better health outcomes for older women.

Other studies use sequence analysis to relate the early labour market years to later outcomes. Tophoven and Tisch (2016) e.g. use administrative data from Germany to study work trajectories and their implications for accrued statutory pension entitlement by the age of 42 for two baby boomer cohorts in comparison with two older cohorts. The authors find that individuals with late entry into employment as well as those with diversified and unstable employment patterns have lower levels of statutory pension entitlements. However, because the self-employed are not included in the German statutory pension system, this study does not offer any insights for the self-employed.

I am not the first to make use of the information contained in SHARE's retrospective interviews. Van Winkle and Fasang (2017) use an older version of the Job Episodes Panel and focus on the early careers of individuals and their complexity. They analyse employment trajectories from age 15 to 45 for men and women born between 1918 and 1963. They find that changes across cohorts are negligibly small, compared with a sizeable variation of complexity in employment trajectories across countries.

More similar to the study presented here, Ponomarenko (2016) also uses data from SHARE.

She studies how cumulative disadvantages of non-employment and non-standard work affect careers and subjective well-being of older Europeans, showing that labour market inactivity and unemployment have significant negative effects on subjective well-being in old-age for men only. While she also studies trajectories from ages 15 to 60, the author uses an older release of SHARE and therefore only includes the thirteen countries present in SHARE Wave 3. Because the focus of her study is on non-employment and non-standard work, she does not differentiate between employees and self-employed.

The study by Madero-Cabib and Fasang (2016) goes one dimension further. They take both careers and family life courses from ages 20 to 59 of cohorts born 1920–1950 in West Germany and Switzerland into account and do a multichannel sequence analysis. This allows them to examine how work and family life jointly affect financial well-being in retirement. They find that breadwinner policies (i.e. supporting a system where men do and women do not do paid work) in combination with liberal pension policies later in life, as in Switzerland, intensify pension penalties for typical female work-family life courses. However, for the same reason as the study by Tophoven and Tisch (2016), they do not offer any insights for the self-employed.

Most closely related to the current study is the one by Pettinicchi and Börsch-Supan (2019). They also use the newest release of SHARE and study the difference in outcomes for formerly traditionally employed and formerly self-employed individuals. They find that the formerly self-employed report a higher degree of financial distress and have lower incomes. The key difference between their study and mine is the categorisation of self-employed. While Pettinicchi and Börsch-Supan (2019) take individuals past into account their classification is made manually. They classify all individuals as self-employed who have worked longer as self-employed than in traditional employment.[1] In this respect, their definition is probably closest to the clusters I identify as long term self-employed. Furthermore, their analysis only includes descriptive statistics, while I also add perform regression analysis controlling for additional factors such as current labour market status.

Using sequence analysis, I define nine clusters of self-employed: (1) always self-employed individuals, (2) those that become self-employed in their 20s and (3) 30s, (4) late career self-employed, (5) self-employed that switch to employment halfway in their trajectory, (6) those with short self-employment spells, (7) individuals with weak labour market attachment, (8) pensioners, and (9) those with mostly missing information on their trajectories. Similar to the results by Pettinicchi and Börsch-Supan (2019), I find that the first three clusters, the long term self-employed, have lower income, both in terms of household income as well as pension income, and are also more likely than employees to report having difficulties making ends meet.

This paper continues as follows. Section 4.2 describes the different data sources and defines the labour market states. Section 4.3 describes the method and presents the self-employment clusters I find. Section 4.4 presents the results on differences of the clusters with respect to

[1] They also deal differently with individuals with missing job episode information.

income variables and subjective measures of financial well-being. Section 4.5 concludes.

4.2 Data

This paper uses data from the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a multidisciplinary and cross-national longitudinal micro data set of individuals aged 50 or older and their households, covering, among other topics, health and socio-economic status. The first wave of SHARE was collected in 2004/2005 and has since then been repeated every other year with the latest wave, number 7, taking place in 2017. As of Wave 7, SHARE covers all members of the European Union as well as Israel and Switzerland.[2] The first part of the analysis uses the Job Episodes Panel (JEP),[3] which is a generated retrospective panel based on the life history interviews conducted in Waves 3 and 7, to build a data driven taxonomy of labour market trajectories. The second part the analysis uses data from Waves 1, 2 and 4 to 7 (Börsch-Supan, 2019a,b,c,d,e,f) to examine the correlation between labour market trajectories and financial outcomes for individuals aged 60 and older.[4]

The sample consists of all individuals in the JEP born from 1931 until 1955, divided into five-year birth cohorts. Older cohorts are excluded because the early years of their labour market trajectories took place during the last years of the second world war. Younger cohorts on the other hand are excluded because they have not yet reached statutory retirement age in most countries by 2017, and only few among them have already gone into (early) retirement. In addition, I decide to focus the analysis on outcomes past age 60 which would exclude cohorts born after 1957 either way. Overall, the sample for the analysis of labour market trajectories includes 62,708 individuals.[5]

While the second part of the analysis focuses on individuals when they are at least 60 years old, the first part of the analysis includes all individuals belonging to the five cohorts, irrespective of whether they are included in the subsequent analysis or not. This guarantees that the sample contains a representative sample of trajectories.

While the JEP contains information for all individuals from their birth onwards, I compare their labour market trajectories during the years when they are between ages 15 and 60. These age thresholds coincide with the end of education and entry into retirement, respectively, for the many of the individuals in the sample. Looking at the distribution of the age of individuals at the end of their education, I find that half of the individuals in the sample finish their education between 15 and 20. (See Table 4.D.1 in Appendix 4.D.) Similarly, the distribution of the age at

[2] Ireland has however only participated in Wave 2 and Wave 3.

[3] DOI: 10.6103/SHARE.jep.700, see Brugiavini et al. (2013), Antonova et al. (2014) and Brugiavini et al. (2019) for methodological details.

[4] Wave 3 only consisted of the retrospective survey and did not collect the variables analyzed in the second part.

[5] Note that the sample selection does not take into account whether additional information other than the labour market trajectory is available for the individual. The main purpose of the sample is to have as many different trajectories as possible such that the clusters I find are indeed representative for the careers of individuals.

retirement shows a median age of 58 for women and 60 for men among those for whom retirement is already observed. (See Table 4.D.2 in Appendix 4.D.)

4.2.1 Definition of labour market states

I take the definition of an individual's labour market state at a given point in time directly from the JEP. Labour market states are therefore coded according to what individuals identify themselves as their main "situation" in a year. I consider the following labour market states: in education, employee, self-employed, unemployed or disabled, not in the labour force, retired, other, and unknown.[6] Two points should be noted regarding these labour market states: first, the definition of self-employment also includes individuals working for their family's business. The data does not allow to differentiate them from self-employed individuals who have their own enterprise.[7] Second, the "unknown" category not only includes individuals that refuse to disclose their labour market state in a given year or those that no longer remember what they did, but also all the labour market states that are coded as missing in the JEP.[8] The latter are the majority in this category.

Table 4.1 reports the distribution of labour market states by cohorts based on this definition, as well as the share of observations for the younger cohorts that we miss because the individuals have not yet reached age 60 at the time of the life history interview (category: "censored"). In addition, the lower part of Table 4.1 shows the share of individuals in each cohort that have spent at least one year between age 15 and 60 in self-employment. Overall there are 8,218 individuals in the sample that have been self-employed. Two observations can be made regarding the share of self-employed: first, their share is largest in the oldest cohort and comparable across the other cohorts. Second, similarly to the findings by Beusch and van Soest (2020) for different cohorts in the Netherlands, Table 4.1 also shows that a larger fraction of individuals is ever self-employed than the shares of self-employment observed at any given point in time. Last, Table 4.1 also shows that the share of periods spent in education is increasing across cohorts. The share of individual-year observations of individuals not participating in the labour market is decreasing, and the mirror image of this is the increasing share of observations of employees.

Self-employment rates across countries may be correlated with differences in institutions. For example, higher minimum wages and/or high employment protection could lead to more self-employment (out of necessity) if they lead to firms hiring more reluctantly. Figure 4.1 illustrates

[6]The following categories consist of several categories in the JEP: unemployed or disabled (unemployed and searching for a job; unemployed and not searching for a job; short term job (less than 6 months); sick or disabled), not in the labour force (looking after home or family; leisure, travelling or doing nothing), other (training; military services, war prisoner or equivalent; managing your assets; voluntary or community work; forced labour or in jail; exiled or banished; labour camp; concentration camp; other), and unknown (refusal; don't know).

[7]This should however not be seen as a drawback of the data as it is more likely to lead to the inclusion of women that help out in their partner's business among the self-employed.

[8]The labour market states coded as missing arise mostly because the timing of some spells is not sufficiently clear. The coding of the JEP furthermore means that one spell with unclear timing can lead to all the following spells being coded as missing as well, since their timing depends on knowing the exact timing of the earlier spell.

Table 4.1: Labour market states by cohorts (in %)

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5
	(1931-35)	(1936-40)	(1941-1945)	(1946-1950)	(1951-55)
Education	5.16	5.76	6.57	7.33	7.90
Employee	56.79	59.86	63.67	65.17	65.05
Self-empl. (SE)	8.39	7.12	6.03	5.80	5.81
Unempl./disabled	1.43	1.19	1.25	1.38	1.66
Inactive	10.54	9.10	7.55	6.13	4.95
Retired	6.13	6.50	6.42	5.53	4.02
Other	1.93	1.67	1.32	1.02	0.95
Unknown	9.64	8.79	7.19	7.43	6.93
Censored				0.21	2.72
Total observations	336,628	478,446	586,546	740,968	741,980
No. of individuals	7,318	10,401	12,751	16,108	16,130
...ever SE (in %)	14.39	13.21	12.45	12.84	13.24

Note: Total observation counts and shares are based on individual-year observations for all 62,708 individuals when they are 15–60 years old. "Unknown" denotes labour market states that are missing in the JEP. "Censored" denotes observations that we cannot make because individuals were younger than 60 at the time they were interviewed. "Ever self-employed" individuals spend at least one year in self-employment during the 46 years under consideration.

the differences in self-employment across countries in the JEP sample. The countries are grouped based on their percentile rank[9] in the Rule of Law indicator published in The World Bank's World Governance Indicators (Kaufman et al., 2010). The Rule of Law indicator captures perceptions regarding the extent to which individuals and firms have confidence in and abide by the rules of society. While it is not directly related to labour market policies, a stronger rule of law can ex-ante be expected to be either positively or negatively correlated with self-employment. The former, if individuals are more willing to start their own business because they are protected by the law, and the latter if individuals turn to self-employment because a weaker rule of law makes employment contracts unattractive. Christelis and Fonseca (2016) show for example that self-employment is negatively associated with the index.[10] The left column groups countries that are ranked below the 80th percentile rank, the middle column the countries ranked 80th and higher up to (and excluding) the 90th percentile rank, and the right column those countries that rank at the 90th percentile or higher in 1996. This is the earliest year for which the indicator is available and therefore the best proxy available for the Rule of Law during most individuals'

[9]The percentile rank of a country in a year is calculated over the sample of all countries for which the indicator is available in that year.

[10]Besides the Rule of Law indicator, I have also tried to group the countries based on other policy indicators that one would (ex-ante) expect to potentially be linked with self-employment. Neither the OECD's employment protection indicator (for temporary and/or regular employment contracts) or labour market programme expenditures (in % of GDP) lead to a grouping of countries with homogeneous self-employment patterns within groups.

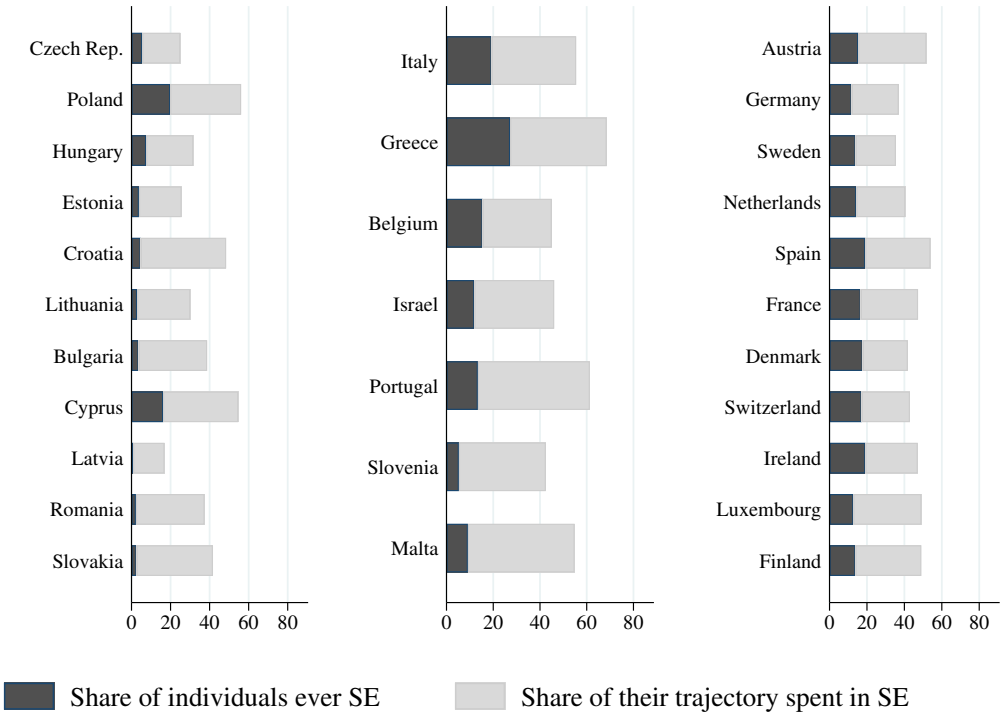


Figure 4.1: Average share of self-employed (SE) and share of years they spend as SE (in %)

active labour market years.

Two observations can be made: first, contrary to the findings by Christelis and Fonseca (2016) Figure 4.1 shows an increase in the share of (ever) self-employed as the average percentile rank increases across groups. It should however be noted here that Christelis and Fonseca (2016) not only included SHARE in their sample but also the US and UK, which in terms of sample size dominate over the SHARE sample and might thus drive their result. Second, while there is quite some heterogeneity within the two lower ranking groups, the average share of a trajectory spent in self-employment seems to be lowest in the lowest ranking countries, and highest in the middle group.

4.2.2 Descriptive statistics

As the second part of the analysis focusses on the individuals from the JEP once they are 60 or older, Table **??** shows the demographic characteristics of the sample conditional on that age restriction. Because the majority of observations pertain to already retired individuals, Table 4.2 groups individuals by the labour market state in which they spent the most years between age 15 and age 60. Exempted from this sorting rule are the individuals that are at least one year self-employed – They are sorted into the two self-employed groups instead.

Table 4.2: Demographic characteristics across labour market groups

	Average	in %							
	Age	Men	Has partner	Secondary education†	Tertiary education	Agriculture	Sample share: Men	Women	All
<10 years SE	68	56.03	46.06	35.61	25.67	6.95	4.14	2.66	3.33
≥10 years SE	70	63.18	47.41	30.40	16.41	29.96	15.01	7.16	10.70
Employee	69	50.56	53.07	39.82	24.41	7.85	77.21	61.87	68.78
Out of labour force	71	0.74	39.71	16.03	3.81	1.78	0.10	11.13	6.16
Retired	70	28.42	46.00	32.43	13.11	8.31	0.95	1.96	1.50
Other	70	19.71	45.60	22.33	12.84	2.80	0.80	2.69	1.84
Missing	70	10.46	46.71	28.38	9.02	5.29	1.79	12.53	7.68
Overall	69	45.05	50.65	35.90	20.78	9.54			

Note: Labour market groups are defined by the years spent in self-employment (SE) and for individuals with no SE the labour market state in which they spent most time between age 15–60. The statistics are calculated for the period when individuals are 60 years or older, over all individual-year observations for each group over a total of 181,847 observations for 60,993 individuals.

† Includes upper secondary and post-secondary non-tertiary education.

Table 4.2 shows that the groups vary a lot in terms of gender shares. The group of individuals that spend most of their trajectory as employees consists of equal parts of men and women. The self-employed group, however, has more men than women, and an even larger share of men is found in the long term self-employed group (≥10 years SE). The remaining groups on the other hand consist of more women than men. There are more than twice as many women than men among the retired, and even more women in the other groups. In particular the group that spends the majority of their prime age years out of the labour force consists almost entirely of women.

Both long- and short-term self-employed have a partner less frequently compared to employees. The long-term self-employed also differ in terms of educational attainment from the other self-employed and employees. They are on average less well educated. The share of individuals with higher secondary or tertiary education is lower than in the other two groups. Table 4.2 also shows that almost a third of the long term self-employed had been working in the agriculture, forestry, and fishery sector, whereas this share is less than 10% in any of the other groups.

Lastly, Table 4.2 also shows that a larger sample-share among men is self-employed. Furthermore, the share of long term self-employed is larger for men than for women. For both genders though, the share of long term self-employed is much larger than the share of short term self-employed. Among women, the group in which periods with missing information dominate the labour market trajectory makes up more than a tenth of the sample. For men on the other hand, the missing time periods are relatively short, as only a small part of the sample is sorted in that

group.

Two variables that I will use to measure financial well-being are reported in Table 4.3. The first variable is total household income. SHARE asks the designated household respondent to report the household's average monthly overall income, after taxes and contributions, in the past year. In order to make the values comparable across waves and countries they are (if necessary) converted into euros, and deflated and PPP adjusted. I use the rates SHARE provides, which use Germany and the year 2015 as the base. I further adjust the income figures for household size using the square root of household size as the equivalence scale. Because the question is asked to household respondents only, the sample is restricted to one person per household. In addition, I also drop outliers from the sample.[11]

The second variable is individual total pension income. It is the sum of state and occupational pension income, and annuities from a private pension/insurance plan in the past year. Individuals who report that they receive no pension income are coded as zeros. Because retirement in SHARE is only known for individuals that self-identify as retired, and individuals who are not in the labour force in particular will frequently not identify themselves as retired despite of being above statutory retirement age and receiving a pension, we cannot know with certainty whether those with zero pensions have no pension or have either not yet retired. Total pensions are also converted in euros, and deflated and PPP adjusted using the SHARE rates.[12] The question is asked to all individuals in the regular survey. Because the question is only asked in the regular survey, no information is available for the countries that only start participating in SHARE with Wave 7[13] or from individuals in refreshment samples of other countries that are recruited in Wave 7, as well as from individuals that participate only in Wave 3. Furthermore, I drop outliers from the pension income sample.[14]

Table 4.3 shows the first (Q_1) and third quartile (Q_3) of the two income variables introduced above. Comparing the sample shares to those in Table 4.2, we can see that the sub-sample for those who answered the pension questions seems to be rather representative. On the other hand, in the income sample for women, those with missing periods in the trajectory, as well as those out of the labour force, are under-represented, whereas employees are over-represented.

A comparison across genders shows that adjusted total household income is similar for the two self-employed groups. Male employees on the other hand report higher household income than their female counterparts. The opposite holds for all other categories. This suggests that

[11] Outliers are defined as values outside the range $[Q_1 - 1.5 * IQR, Q_3 + 1.5 * IQR]$. The quantiles ($Q_1$ and Q_3) and inter quantile range (IQR) are calculated within countries. There are no observations that fall below the lower bound. Because the majority of the excluded observations report unrealistically high values of monthly income, this is unlikely to generate selection bias.

[12] I would like to thank Andrea Bonfatti for providing me with the PPP rates for the Netherlands for the years 2014/15.

[13] Bulgaria, Cyprus, Finland, Latvia, Lithuania, Malta, Romania, and Slovakia started participating in SHARE with Wave 7.

[14] Outliers are defined in the same way as for household income but are calculated without taking the zeros into account. Again, only the upper bound is binding for some observations.

Table 4.3: Total household and pension income, and sample share (%) by gender and labour market group

	Men				Women			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Total household income (monthly)								
<10 years SE	876	232	1539	4.23	871	223	1366	2.73
≥10 years SE	829	302	1319	13.35	846	233	1281	6.34
Employee	920	232	1571	79.03	578	150	1250	67.32
Out of labour force	133	108	458	0.12	768	477	1094	9.16
Retired	604	128	1171	0.99	770	176	1361	1.79
Other	481	139	1016	0.92	817	281	1144	2.48
Missing	482	150	995	1.36	932	577	1318	10.18
Total pension income (annual)								
<10 years SE	2811	0	14280	4.42	1247	0	8243	2.91
≥10 years SE	5609	0	10911	16.28	1362	0	7648	7.54
Employee	6307	319	16390	75.91	2213	258	9349	59.82
Out of labour force	1611	0	3010	0.10	0	0	2619	11.88
Retired	1592	0	10825	0.92	1129	0	6612	1.89
Other	1003	0	6917	0.81	333	0	6215	2.75
Missing	1293	0	8841	1.56	0	0	5533	13.21

Note: All values are in euro, and PPP as well as inflation adjusted. Household income is also adjusted for household size with its square root. The statistics for total household and pension income are calculated over all individual-year observations for each labour market group for all individuals in the sample aged 60 or older. (Total observation count household income: 33,870 (m), 48,329 (w), pension income: 48,070 (m), 62,444 (w).) Observation counts differ across measurements because of sample differences. See Section 4.2.2 for more information on the adjustments and sample selection.

most women in these labour market groups could probably rely on a partner's income whereas their male counterparts could not, corresponding to traditional gender roles in these generations. Pension income reflects these findings but also shows that self-employed and employed men receive more pension income than their female counterparts.[15] Both genders however have more individuals that receive no pensions among the self-employed than among employees. For men, Table 4.3 also shows that short term self-employed have a similar median and third quartile of total pension income as employees have, whereas the long term self-employed have lower values. In contrast though, the first quartile of short-term self-employed is much lower than employees' whereas the spread of pensions of long-term self-employed is smaller. Among women, the two self-employed groups have slightly lower pension incomes than employees.

[15]This should however not entirely be taken at face value as some pension systems (e.g. Switzerland) do not divide pensions equally between partners while the second person's (usually the wife's) pension will generally be caped.

4.3 Labour market trajectories over the life cycle

This section first describes the analysis of the labour market trajectories. Each individual's trajectory is defined by the labour market states the individual occupies between ages 15 and 60. I use sequence analysis to identify various career patterns and to group the trajectories into clusters. In the second part of this section I describe the differences between the clusters that I find.

4.3.1 Sequence analysis: comparing trajectories

The idea to analyse sequences of social events as a whole rather than the events in a given year (e.g. as cross sectional analysis, event studies, or transition probabilities) was introduced into the field of social sciences by Abbott (1983). The idea was further developed by, e.g., Abbott and Forrest (1986), Abbott and Hrycak (1990) and Abbott (1995). The goal of all sequence analysis is to group trajectories that display similar patterns. That is, to group individuals that are in the same states at similar times and in a similar order. In technical terms this means that the sequence of each individual gets compared pairwise to all other individuals' sequences and a matrix of how different each sequence is from the others is computed. The summary by Studer and Ritschard (2016) provides a good overview of the dissimilarity measures used in the field. The dissimilarity measure can then be used to group the individuals with e.g. a clustering algorithm.

To calculate the dissimilarity measure I use optimal matching (OM), which is the measure that is most commonly used.[16] OM computes an "edit distance" – the least costly way in which a sequence can be transformed into another sequence. The operations OM uses for this are substitution, insertion and deletion, and each of them is associated with a cost.[17] Choosing these costs is left to the researcher and generally depends on the question the researcher wants to answer. It is common to choose the same costs for insertion and deletion, the so-called "indel" costs.[18] This leaves the choice limited to the relative size of indel costs compared to substitution costs. The cheaper indel costs are, the less importance OM places on the timing of events and the more on their sequencing, i.e, the order in which they occur. Conversely, relatively high indel costs will place more importance on the order of events. If indel costs are half of substitution costs, then OM is indifferent between the two. Hence, if one were only to care about self-employment spells taking place in a sequence, lower indel costs might be preferred. However, since the timing of self-employment spells may affect later outcomes – e.g. because mandatory participation in the pension system differs for employees and the self-employed, as is the case in many European countries, an individual that is self-employed at the beginning of their career has their whole working life left to compensate for potential pension gaps, whereas an individual that enters self-

[16] I use the package SADI in Stata Halpin (2017) for OM.

[17] See, for example, Beusch and van Soest (2020, Section 3.2) on how edit distances are calculated.

[18] Hollister (2009) is one of the few to question the methodology of fixed indel and substitution costs. Her proposed solution has its own problems though, as pointed out by Studer and Ritschard (2016).

employment later in life has less time during which they can make arrangements to compensate for these gaps. In such a case, lower substitution costs are preferred such that individuals with a similar sequence of events are closer to each other.

In this study, I choose symmetric indel costs of 1.5 and take substitution costs of 2 for all labour market states, with two exceptions: Education and missing years get higher substitution costs of 3.[19] In the case of education I choose the higher costs to differentiate individuals that remain longer in education. With higher substitution costs for educational spells, trajectories of individuals that in any form participate in the labour market will be closer to each other than to those individuals that remain longer in education. For years with missing information on the other hand, the higher cost is chosen such that substitutions in a sequence with missing information years are not too easily made.

The choice of substitution costs has little impact on the clustering of never self-employed trajectories. For this sample, results with uniform substitution costs for all labour market states would give almost the same solution (not shown in this paper). For the trajectories involving self-employment, the choice of the costs matters more, in particular the higher cost for missing years. Without these higher costs, individuals with missing years would tend to get sorted in the same clusters as individuals who display a similar overall pattern but with a different labour market state instead of a missing observation.

The discussion of substitution costs for missing years directly raises another discussion point regarding the sample of trajectories on which OM is performed. As is clear from the discussion of costs, I also include trajectories that are incomplete – because individuals do not remember what they did in every year, or because of censoring due to age at the time of the interview. By simply treating these missing years as a separate state, OM can include the trajectories with missing years in the analysis. I therefore do not need to rely only on the much smaller sub-sample of individuals that recall their labour market trajectories perfectly. In fact, without this trick I would lose around 45% of the sample, of which two thirds of the trajectories miss less than ten years out of the forty-six under consideration. This would be particularly problematic if the subsample of complete trajectories is different from those with missing years, since only then the analysis built on OM uses a population representative set of sequences. As shown in Appendix 4.A the sub-sample of complete trajectories differs from the full sample, so that it is indeed useful to include the incomplete trajectories.

OM is applied to the full sub-sample of individuals that are at least once self-employed and separately to the complementing sample of individuals that are never self-employed. The latter sample is further split by five-year birth cohorts because of computational limitations. After the OM algorithm has determined the edit distances, I apply Ward's Method (Ward, 1963), a hierarchical agglomerative method, to group the trajectories. Ward's Method produces a tree of potential groupings. In practice the researcher has to choose where to prune the tree, i.e. how

[19] While this makes the two labour market states in fact equivalent to two indel operations instead of a substitution, the more important point is that they are more expensive to substitute with than other states.

many groups they want to use. While there are some criteria that help with this decision, there exists no "hard" statistical criterium. In this application I am less concerned with finding the right number of clusters and more with using the algorithm as a tool to separate individuals automatically in groups. Hence, the number of clusters is chosen such that the groups match with ex-ante expectations of self-employment and labour market trajectory types, but also such that the solution is as parsimonious as possible.

Figure 4.2 shows the Ward's 12 cluster solution for all individuals in our sample that spend at least one year in self-employment. I combine the twelve groups from Ward's solution (denoted in "[ ]") into the following nine clusters:

1. *Always self-employed* [12] – These individuals are typically self-employed during their whole labour market career.

2. *Self-employed from their twenties* [9, 10, 11] – These individuals turned self-employed in their twenties (after spending longer years in education or a few years in employment) and remain self-employed until they retire.

3. *Self-employed from their thirties* [1, 2] – These individuals turn self-employed in their thirties or early forties and remain self-employed until they retire.

4. *Late career self-employed* [5] – This cluster predominantly consists of individuals who spend the last 10 to 15 years of their trajectory in self-employment.

5. *Self-employed to employee* [7] – These individuals are self-employed in the first half of their career and then switch to employee status.

6. *Short self-employment spells* [8] – The individuals in these two clusters spend only a few years in self-employment and spend the majority of their career as employees.

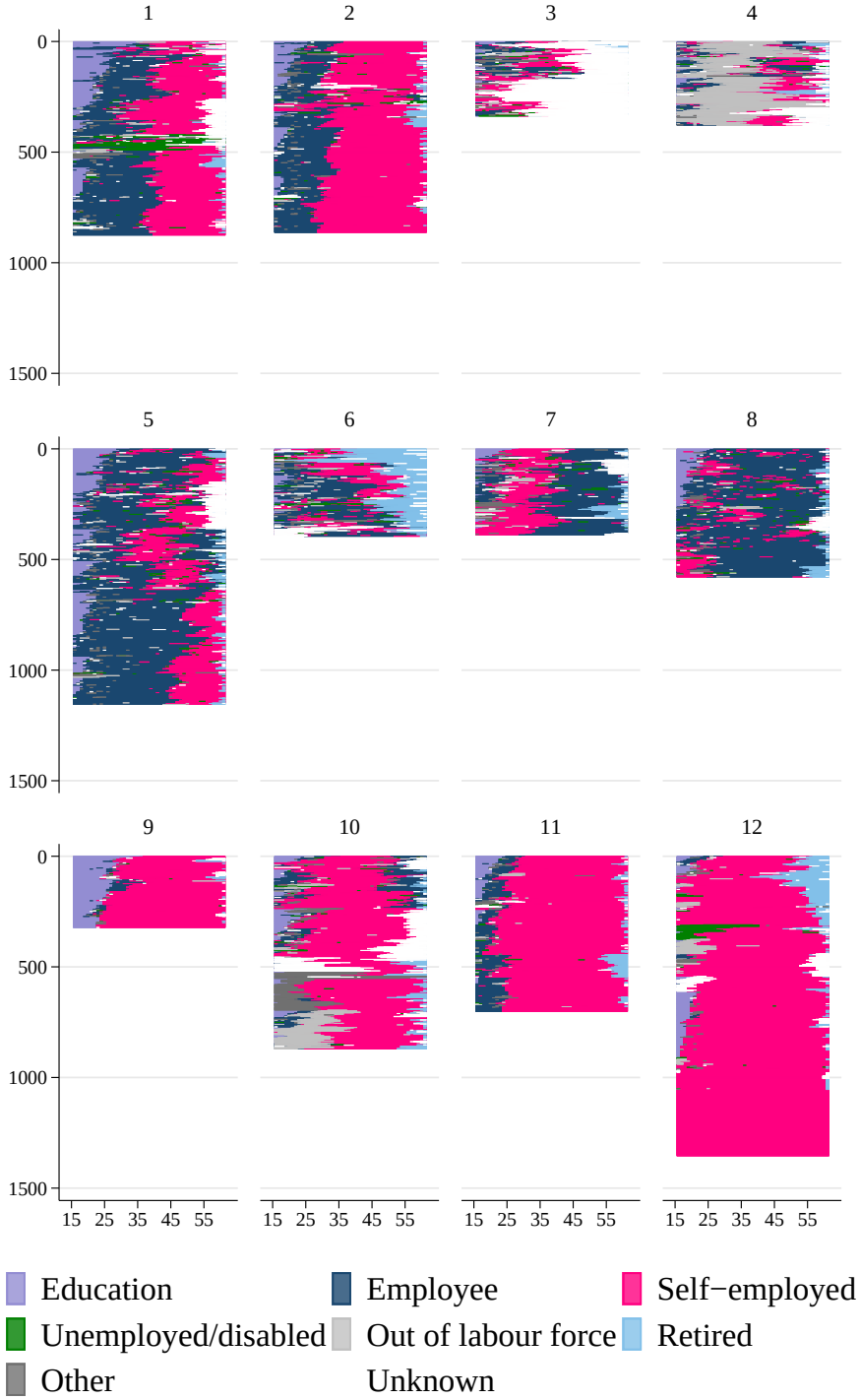
7. *Weak labour market attachment* [4] – The individuals in this cluster spend the majority of their trajectory out of the labour market.

8. *Pensioners* [6] – This cluster is defined by individuals that spend a large part of their trajectory as (early-)retirees.

9. *Unknown trajectories* [3] – Typically, a large part of the labour market history of these individuals is unknown.

The individuals with no self-employment (SE) spells are combined into seven different clusters:[20]

[20] See Appendix 4.C for the index plots, as well as the correspondence table of Ward's solution for all five cohorts and the six clusters.



Graphs by Ward's 12 cluster solution

Figure 4.2: Indexplot of all cohorts

11. *Employees* – Individuals in this cluster spend almost their complete career as employees. The cluster also includes those that leave education in their twenties.

12. *Early retirees* – Individuals in this cluster retire from employment already in their fifties.

13. *Weak labour market attachment (no SE)* – This cluster includes individuals that spend some time of their labour market trajectories as employees and the remainder out of labour force or as benefit recipients, etc.

14. *Out of labour force* – The individuals in this cluster spend almost their entire trajectory out of the labour force.

15. *Pensioners (no SE)* – Similar to the cluster with self-employment these individuals spend almost the whole trajectory as pensioners.

16. *Benefit recipient or "other" income* – This cluster combines the two rather small groups of individuals that receive benefits or have "other" income throughout their trajectories.

17. *Unknown trajectories (no SE)* – Like their counterparts among the self-employed, individuals in this cluster miss most information of their trajectories.

4.3.2 Differences across clusters involving self-employment

How does the distribution of clusters differ across cohorts? Table 4.4 shows their shares by cohort for the SE sub-sample, while Table 4.5 shows the overall distribution merging all self-employed clusters into one category. Several observations can be made. Among the self-employed, the share of individuals that spend their full labour market trajectory in self-employment is decreasing across cohorts. While a quarter in the oldest cohort is always self-employed, this holds for only about one tenth of the youngest cohort. There are changes over time in the size of the other clusters too. While the changes are relatively small in most, there are two clusters that stand out. The share of the cluster of individuals that become self-employed in their thirties increases from cohort 3 to 6, and the cluster of late career self-employed increases in particular between cohorts 1 and 3. For the non-self-employed individuals, Table 4.5 shows that the share of employees increases over time while the shares of individuals with weak labour market attachment and those that remain out of the labour force for their whole career decrease over time. The early retirees on the other hand peak for cohorts 2 and 3 and decrease again thereafter. Similar patterns are also visible when we look at the shares across cohorts by groups. The exception are the self-employed that start in their thirties in the lowest ranking countries. There, the increase is much steeper and likely driving the overall trend.[21]

[21] See Table 4.D.4 in Appendix 4.D for the shares of the self-employed clusters across regions and Table 4.D.5 for the non-self-employed.

Table 4.4: Self-employed clusters by cohorts (in % of self-employed sub-sample)

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Overall
Always SE	26.10	22.96	16.02	12.47	10.80	16.26
SE from 20s	25.33	23.69	23.64	21.69	22.31	23.04
SE from 30s	18.29	17.71	19.48	21.59	25.40	21.08
Late career SE	6.29	11.08	15.45	16.46	16.70	14.10
SE to employee	6.10	5.10	3.85	4.29	5.04	4.77
Short SE spells	5.81	7.00	6.87	7.89	7.28	7.11
Weak LM attachment	4.29	4.30	4.92	5.72	3.81	4.64
Pensioners	3.52	4.52	7.00	5.52	3.47	4.85
Unknown trajectories	4.29	3.64	2.77	4.39	5.19	4.14
No. of individuals	1,050	1,372	1,586	2,029	2,102	8,139

Table 4.5: Clusters by cohorts (in %)

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Overall
All SE clusters	14.35	13.19	12.44	12.63	13.06	13.00
Employees	35.15	34.56	39.56	42.36	46.66	40.76
Early retirees	16.23	28.06	24.96	15.97	10.95	18.55
Weak LM attach. (no SE)	13.15	3.79	7.11	9.28	7.32	7.88
Out of labour force	7.86	6.37	5.27	5.86	4.83	5.79
Pensioners (no SE)	1.89	1.39	2.13	3.31	5.72	3.20
Benefits or other	1.74	1.66	1.67	0.57	1.39	1.32
Unknown traject. (no SE)	9.64	10.97	6.87	10.01	10.07	9.50
No. of individuals	7,315	10,399	12,750	16,069	16,096	62,629

Note: All SE clusters includes all self-employed clusters as shown in Table 4.4.

The upper graph in Figure 4.3 shows the distribution of clusters in the self-employed sub-sample for the countries in SHARE that rank below the 80th percentile rank by Rule of Law indicators. Compared to the countries ranking between the 80th and 90th percentile in the middle, and those ranking above the 90th percentile shown at the bottom of Figure 4.3, one can see that there are some general differences between the groups despite of some heterogeneity in the two lower ranking groups. In the first group we see that for most countries the self-employed consist mostly of individuals that turn to self-employment after their thirties or late in their career. Exceptions are Poland, Croatia and Cyprus which all show a pattern more similar to the second group of countries where always self-employed and those turning to self-employment in their twenties dominate. Last, the self-employed individuals in most of the countries that rank highest in the rule of law indicator are made up relatively evenly between individuals that turn self-employed in their twenties and thirties while the share of late career self-employed

is between the shares of the first and second group. The high ranking countries also have a higher share of individuals with weak labour market attachment that spend some years in self-employment.

The differences are less strong for the shares of total self-employment and the non-self-employed clusters across countries and regions. (See Figure 4.E.1 in Appendix 4.E.) The main characteristics are the following: In the lowest ranking countries the share of self-employed is almost negligible except for Poland and Cyprus while the share of early retirees is relatively large. In the group ranking between the 80th and 90th percentile rank, the self-employment shares are comparable to the shares in the countries ranking above the 90th percentile rank. In the latter group most countries however have a larger share of employees than the countries in the other two groups.[22]

Table 4.6 shows the demographic characteristics from Table 4.2 but now by cluster.[23] The average age in the sample is approximately the same for all clusters, except for the never self-employed pensioner cluster where individuals are slightly younger on average. In terms of gender, Table 4.6 shows large differences across the clusters. These differences are stronger than with the sorting by labour market group. There are about 50% more men than women in the employee cluster, and some of the self-employment clusters have an even higher share of men. The clusters of individuals turning self-employed in their 30s, late career self-employed, and the cluster with short self-employment spells all consist for at least 70% of men. Clusters with weak labour market attachment, out of the labour force, and unknown trajectory parts are mostly or almost entirely female. And while the self-employed pensioners cluster contains approximately equal numbers of men and women, the non-self-employment clusters with both early retirees and pensioners have a share of approximately 60% women.

Compared with the remainder of the sample, individuals in many of the self-employed clusters have less frequently a partner. They also differ in terms of their attained education level. These differences correspond with those already seen in Table 4.2. The clusters that spend more years in self-employment have fewer indiviudals that have attained higher secondary or even tertiary education. Among the always self-employed 75% have at most finished lower secondary education and less than 5% have attained tertiary education. This stands in contrast with e.g. late career self-employed who are the best educated among all self-employed, with the largest share of individuals with higher secondary education and a share with tertiary education comparable to that of employees. While the other clusters involving self-employment all have comparable secondary education shares, they have fewer individuals with tertiary education compared to employees. Finally, Table 4.6 also shows that the majority of the always self-employed have been active in the sector agriculture, fishery and forestry. Among the self-employed that start in their 20s, only around one third is working in that sector, and for those that start in their 30s, less

[22]The share of employees and early retirees/pensioners of course also depends on the sampling of cohorts in SHARE and countries' statutory retirement age.

[23]The sample is slightly smaller because OM is not calculating an edit distance for some individuals.

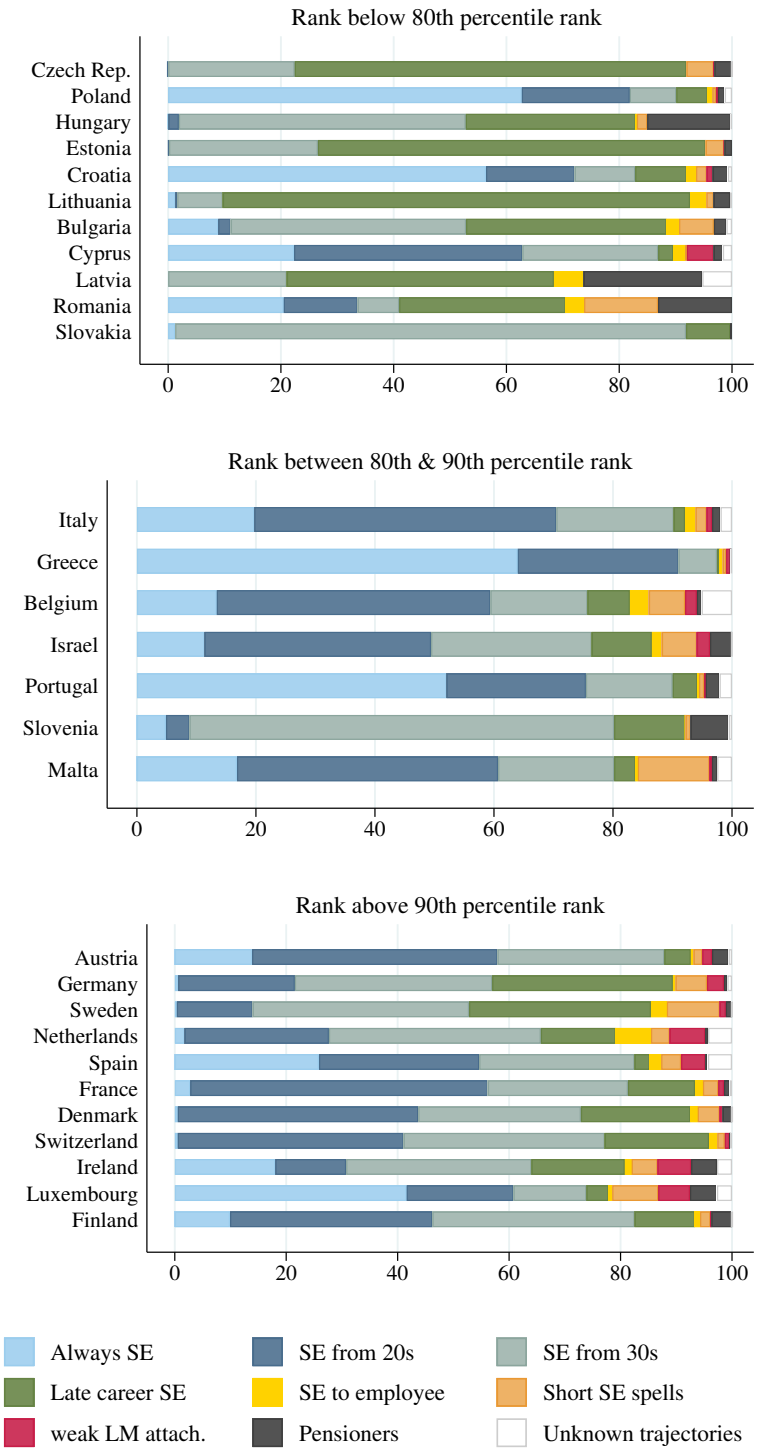


Figure 4.3: Share (in %) of clusters in countries

Table 4.6: Demographic characteristics across clusters

	Average	in %							
	Age	Men	Lives with partner	Secondary education†	Tertiary education	Agriculture	Sample share: Men	Women	All
SE clusters									
Always SE	71	62.06	45.66	20.23	4.13	59.62	3.12	1.56	2.26
SE from 20s	70	62.44	47.65	27.28	19.85	31.86	4.57	2.25	3.29
SE from 30s	69	73.52	50.42	36.02	22.88	18.31	4.70	1.39	2.88
Late career SE	68	70.63	49.32	45.16	30.48	7.77	2.99	1.02	1.90
SE to employee	70	59.74	43.23	29.20	18.48	8.58	0.88	0.49	0.67
Short SE spells	68	73.39	46.67	33.25	22.93	7.94	1.74	0.52	1.07
weak LM attach. (SE)	69	2.16	38.03	33.18	13.51	11.04	0.03	1.27	0.71
Pensioners (SE)	69	54.36	49.60	40.52	14.62	13.00	0.82	0.57	0.68
Unknown traj.(SE)	69	23.86	38.61	21.45	10.32	9.81	0.29	0.77	0.55
Never SE clusters									
Employee	69	61.60	54.82	38.51	32.25	7.33	54.44	27.81	39.80
Early retiree	70	41.71	51.74	42.93	14.41	8.74	17.58	20.12	18.98
weak LM att. (no SE)	70	17.01	45.65	35.60	11.93	7.47	3.00	12.00	7.95
Out of LF (no SE)	70	1.21	40.93	16.54	4.34	1.23	0.16	10.57	5.88
Pensioner (no SE)	67	36.38	53.39	40.05	15.77	7.52	2.09	2.99	2.58
Benefits or other	70	19.23	44.13	20.44	5.65	2.33	0.56	1.91	1.30
Unknown traj. (no SE)	69	14.42	46.85	30.36	9.53	6.62	3.04	14.77	9.49
Overall	69	45.05	50.65	35.90	20.78	9.54			

Note: The statistics are calculated for the period when individuals are 60 years or older, over all individual-year observations for each group over a total of 181,847 observations of 60,993 individuals.

† Includes upper secondary and post-secondary non-tertiary education.

than one in five is in that sector. The self-employed pensioners as well as those with weak labour market attachment also have slightly higher shares of individuals in agriculture, whereas the other self-employment clusters are all comparable to employees.

Table 4.7 reports the same income variables as Table 4.3, but broken down by clusters instead of labor market status. For brevity, the non-self-employment clusters, except for employees, are left out because the results for these clusters are similar to those for the groups with corresponding main labour market states shown in Table 4.3. The sub-samples are also relatively representative with respect to the clusters.[24] While the different income measures were rather similar across the two labour market groups, this no longer holds for the clusters.

Three observation holds across genders: First, always self-employed report lower adjusted

[24]See Table 4.D.3 in Appendix 4.D for the corresponding sample shares of clusters by gender in the full sample.

Table 4.7: Total household and pension income, and sample share (%) by gender and cluster

	Men				Women			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Total household income (monthly)								
Always SE	787	499	1160	13.40	689	147	1003	14.82
SE from 20s	877	419	1367	23.76	925	386	1316	21.52
SE from 30s	833	243	1399	24.74	945	266	1469	14.30
Late career SE	580	187	1359	16.99	780	191	1275	12.47
SE to employee	976	462	1455	4.80	884	191	1544	5.65
Short SE spells	1056	330	1602	9.72	868	164	1500	6.16
Weak LM attach.	954	474	1119	0.20	1001	505	1397	11.40
Pensioners	896	185	1314	4.69	645	135	1155	6.52
Unknown trajectories	829	225	1321	1.70	940	654	1206	7.16
Employee	873	220	1617	55.70	620	193	1312	29.70
Total pension income (annual)								
Always SE	6526	0	9566	15.93	1263	0	6235	15.17
SE from 20s	6827	0	11160	24.57	2367	0	8198	22.65
SE from 30s	3145	0	10911	24.72	1341	0	8092	13.99
Late career SE	1638	0	10800	15.13	1440	0	8177	10.48
SE to employee	9980	0	15630	4.82	2527	0	12092	5.07
Short SE spells	9513	339	15712	8.94	1576	0	12053	5.38
weak LM attach.	7232	0	12408	0.15	1160	0	6946	13.29
Pensioners	4084	561	13148	4.19	1744	23	9588	5.59
Unknown trajectories	141	0	8243	1.56	0	0	5077	8.37
Employee	5007	243	15911	53.78	1947	32	9600	26.96

Note: All values are in euro, and PPP as well as inflation adjusted. Household income is also adjusted for household size with its square root. The statistics for total household and pension income are calculated over all individual-year observations for each cluster for all individuals in the sample aged 60 or older. Sample shares of the SE clusters are in % of the SE sub-sample. Employee sample shares are in % of the full sample of each variable. (Total observation count household income: 33,870 and 5,955 SE (m), 48,329 and 4,386 SE (w), pension income: 48,070 and 9,948 SE (m), 62,444 and 6,525 SE (w).) Observation counts differ across measurements because of sample differences. See Section 4.2.2 for more information on sample selection.

household income compared with the other self-employed clusters or employees. Second, for late career self-employed the median household and total pension income is low but the third quartile is as high as that of self-employed that start earlier in their career. Third, we don't see large differences for household income within both genders. For pension income however, among those individuals who only spend a short time in self-employment as well as those that start as self-employed and then switch to employment a larger share among them seem to do better in financial terms. Both of these groups report the highest third quartile vaues..[25]

4.4 Regression results

To better understand whether there are differences with respect to financial outcomes in old age between the self-employment clusters and the individuals that spend most of their career as employees, I estimate regression models for several measures of financial well-being. The general regression model is given by

$$y_{it} = \beta_0 + X'_{it}\beta_x + W'_i\beta_c + C'_i\gamma + T'_t\delta + \epsilon_{it} \tag{4.1}$$

where y_{it} is the dependent variable of interest. X_{it} contains the demographic characteristics which are controlled for using different dummy variables. The first set controls for educational attainment. The reference category is less than lower secondary education, and I control for upper secondary and tertiary education. I also control for marital status at the time of the interview. The reference category here are individuals without a partner, and two dummy variables control for individuals who have a partner who's alive (irrespective whether they are married or not), or who are widowed. Next, I control for the sector in which an individual was longest active during their labour market trajectory. One dummy variable controls for individuals that were active in the agricultural, forestry, and fishery sector. Another dummy controls for individuals whose sector is unknown. The reference category are all other sectors. A set of variables control for individual's labour market states at the time of the interview. The reference category are retired individuals, and the controls that enter the regression are a dummy for individuals who are working and one for all other labour market states. Last, I also include a dummy to control for individuals being above statutory retirement age.[26] Finally, W_i controls for the clusters, taking employees as the reference category, and C_i and T_t capture cohort and wave fixed effects, respectively.

Table 4.8 shows a summary of the regression results for the income variables discussed earlier. All models are estimated separately by country group as panel regressions with random effects.[27]

[25]The values in Table 4.7 suggest that men with weak labour market attachment do well too, but their sample size is too small to make strong statements for their cluster. The cluster is only included for the sake of completeness as it is an important cluster among the female respondents.

[26]The JEP includes information on statutory retirment ages. I take the maximum age observed for an individual as the threshold. There are is no statutory retirement age imputed for individuals born before 1939 (almost all countries), 1941 (Bulgaria and Romania) and 1947 (Croatia). For these I take the next available cohort's threshold as a proxy.

[27]Because of the many time invariant controls the models cannot be estimated with fixed effects.

The regressions for household income are estimated using linear regression. Pension income models are estimated with Tobit to account for the fact that some individuals that report zero pensions may not yet be retired rather than not have any pension.

With respect to the variables not shown in Table 4.8 the effects are the same across genders. I find, as expected, that a higher attained education level is associated with higher income. Individuals that work in the agricultural sector have higher household and pension incomes in group 1 but not in group 2 and 3. The effect of widowhood also depends on the country group. Both in group 1 and 3 the effect is positive whereas it is negative in group 2.

With respect to the cluster coefficients the regressions show different patterns across both genders and country groups. I find that self-employed men in countries in group 1 on average have statistically significantly lower income than employees if they are always self-employed, or late in their career, or during short spells. In group 2 and 3 on the other hand most of the long-term self-employed are worse off than employees. The effect is however insignificant/of weak statistical significance for men that are always self-employed or become self-employed in their 20s in group 3 countries.

The picture is different if we look at men's self-reported pensions. Here I find for all men that are always self-employed that their pension income is lower than employees'. In group 1 individuals that become self-employed in their 20s are found to have comparably lower pension income too whereas the other clusters have insignificant or only weakly statistically significant effects (at the 10% confidence level). Similarly, I find, apart from the always self-employed, only those who become self-employed in their 30s to also have a statistically significant lower income in group 2. Group 1 and 2 therefore stand in contrast with group 3 where all long-term en late career self-employed are found to have statistically signficantly lower pension incomes compared to employees. The effect is estimated to be larger the later individuals start with self-employment.

Opposite to the findings for men, are the regression results for the women. First, I find for group 1 that not just women with short self-employment spells live in households with statistically significantly lower household incomes but also all long-term self-employed women. They stand in contrast to women in group 2 and 3. In group 2 only those that are always self-employed have lower household incomes and in group 3 I even only find positive income effects for self-employment. This is also reflected in the pension income results where all self-employment clusters in Table 4.8 in group 1 have negative coefficients whereas only few clusters in group 2 and 3 are found to have negative coefficients. This suggests that women in group 2 and 3 might not have the same incentives to turn self-employed as women in group 1. That is, it might be that women in countries in group 1 are rather driven to self-employment out of necessity whereas women in countries in group 2 and 3 might rather be additional income earners.

While the income variables show that most long-term self-employment clusters have less income than comparable individuals that were mostly employed throughout their labour market trajectory, income alone cannot tell us whether individuals that used to be self-employed are

Table 4.8: Regression summary: income variables

		Counry Group 1	*Country Group 2*	*Country Group 3*
Rule of Law indicator:		*(pct rank < 80)*	*(80 ≤ pct rank < 90)*	*(90 ≤ pct rank)*
Monthly household income (PPP and household size adjusted, deflated; RE panel)				
Men:	Always SE	-115.2***	-92.38***	52.60
		(39.11)	(34.88)	(54.58)
	SE from 20s	-77.07	-125.5***	-86.69*
		(51.94)	(34.80)	(47.55)
	SE from 30s	-20.82	-98.33**	-147.6***
		(40.79)	(42.56)	(44.15)
	Late career SE	-74.21***	-122.8*	-210.6***
		(28.67)	(67.51)	(57.10)
	SE to employee	-41.43	-111.3	-74.08
		(77.69)	(72.14)	(101.1)
	Short SE spells	-128.2***	73.36	-100.9
		(45.53)	(64.99)	(68.86)
Women:	Always SE	-233.9***	-183.4***	165.5***
		(22.14)	(47.40)	(61.46)
	SE from 20s	-179.3***	-69.79	218.1***
		(26.36)	(48.78)	(54.26)
	SE from 30s	-162.2***	-95.17	114.6**
		(56.88)	(76.80)	(57.47)
	Late career SE	-60.94	-94.14	-16.33
		(39.84)	(132.7)	(72.56)
	SE to employee	-94.64	95.91	31.87
		(73.91)	(104.4)	(127.1)
	Short SE spells	-157.6***	402.7***	-98.57
		(56.43)	(113.2)	(105.7)
Annual pension income (PPP adjusted and deflated; RE Tobit)				
Men:	Always SE	-1,471***	-3,372***	-1,783**
		(294.1)	(1,055)	(844.5)
	SE from 20s	-1,467***	-961.5	-2,275***
		(475.0)	(962.2)	(532.0)
	SE from 30s	-518.0*	-2,196**	-3,039***
		(277.7)	(1,071)	(495.8)
	Late career SE	-310.3	-3,112*	-3,748***
		(225.9)	(1,629)	(624.7)
	SE to employee	-1,269*	-335.2	998.7
		(726.2)	(2,005)	(1,109)
	Short SE spells	47.61	1,692	17.38
		(429.4)	(1,629)	(801.1)
Women:	Always SE	-2,308***	-3,221**	-803.5
		(261.4)	(1,307)	(886.2)
	SE from 20s	-2,292***	-1,769	-769.9
		(349.5)	(1,144)	(545.9)
	SE from 30s	-999.0**	-5,656***	-1,317**
		(397.3)	(1,763)	(630.9)
	Late career SE	-843.3***	-920.8	-979.0
		(278.6)	(2,536)	(753.4)
	SE to employee	-1,879**	4,196*	-637.0
		(940.8)	(2,329)	(1,073)
	Short SE spells	-1,256**	4,096*	915.9
		(507.2)	(2,463)	(1,087)

(Robust) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: See Tables 4.D.7–4.D.10 in Appendix 4.D for more detailed results.

really worse off because of this. The data can for example not account for the self-employed living to a higher extent off their savings than former employees, or for individual preferences, i.e. if self-employed have accounted for this beforehand and made a conscious choice to live with less means in their old age. Furthermore, state pension income is frequently split between partners with some countries not assigning equal parts to both individuals but a lower share to the one that turns retired later in time, which are frequently the women. This might therefore bias the effects estimated for pension income of women downwards. I therefore also look at other variables to complete the picture. One further advantage of all the variables discussed hereafter is that the sample needs no treatment for outliers and there is therefore no sample selection based upon the dependent variable in the following.

One frequent argument made is that self-employed individuals are more frequently house owners[28] and that they may therefore potentially have lower costs of living if they pay down their mortgage. As a robustness check I therefore also estimate a panel probit regression for the probability that the household respondent's family lives in a dwelling that they own. (Results not reported here.) Except for the first two self-employment clusters in the second country group, I however find no statistically significant results that indicate that any of the self-employed clusters are more likely that employees to own their dwelling. Hence, it seems unlikely that the self-employed have a potential for reduced living costs through the channel of owning their dwelling.

In a next step, I look at individuals subjective financial well being. The SHARE survey asks all household respondents whether their household is able to make ends meet, with answers on a scale from 1 to 4, where 1 indicates "with great difficulty" and 4 indicates "easily". The upper half of Table 4.9 presents the regression results from an ordered probit panel regression explaining this survey item. Once more I find that most coefficients of the long term self-employed men are statistically significant and negative. In the case of the always self-employed this also holds across both genders and all country groups. Overall, this offers some support for the findings for the income variables and also confirms the findings of Pettinicchi and Börsch-Supan (2019). The results are also in line with the estimated effects of demographic variables earlier. That is, higher educated individuals report less difficulties with making ends meet. More notably, widowers report less difficulties while the effect is the opposite for widows.

As a robustness check I also compute a material deprivation score following Bertoni et al. (2015). This material deprivation score takes the sum of different survey items that indicate whether the respondent's household put up with cold to save on heating costs, did not go to the dentist because of costs, could not pay unexpected expenses, was at least two months behind with the rent, or avoided going to the doctor because of the costs. The weights of these survey items in the deprivation index are based on a regression of the respondents answer on how satisfied they are with life on a scale from 0 to 10, where 10 indicates complete satisfaction on the survey

[28] E.g. Zwinkels et al. (2017) show for households with current working solo self-employed in the Netherlands that they are more frequently houseowners than households with only employees.

Table 4.9: Regression summary: financial well-being indicators

		Counry Group 1	*Country Group 2*	*Country Group 3*
Rule of Law indicator:		*(pct rank < 80)*	*(80 ≤ pct rank < 90)*	*(90 ≤ pct rank)*
Making ends meet: 1 – "with great difficulty"; 4 – "easily" (RE ordered probit)				
Men:	Always SE	-0.426***	-0.467***	-0.195*
		(0.154)	(0.104)	(0.110)
	SE from 20s	-0.625***	-0.291***	-0.136*
		(0.174)	(0.0898)	(0.0762)
	SE from 30s	-0.00869	-0.259**	-0.236***
		(0.123)	(0.104)	(0.0728)
	Late career SE	0.167*	-0.114	-0.157*
		(0.100)	(0.162)	(0.0878)
	SE to employee	-0.0640	-0.112	-0.0936
		(0.316)	(0.168)	(0.182)
	Short SE spells	-0.201	0.225	-0.157
		(0.224)	(0.160)	(0.119)
Women:	Always SE	-0.436***	-0.577***	-0.271**
		(0.127)	(0.142)	(0.133)
	SE from 20s	-0.0513	-0.190	-0.0619
		(0.143)	(0.118)	(0.0960)
	SE from 30s	0.108	-0.292	-0.367***
		(0.170)	(0.185)	(0.0987)
	Late career SE	0.256*	-0.363	-0.555***
		(0.140)	(0.273)	(0.129)
	SE to employee	-0.123	0.283	-0.332*
		(0.364)	(0.248)	(0.193)
	Short SE spells	-0.236	0.558**	-0.222
		(0.275)	(0.248)	(0.164)
Material deprivation score: higher values – more material deprivation (RE panel)				
Men:	Always SE	0.0173	0.0509***	-0.00512
		(0.0287)	(0.0156)	(0.00924)
	SE from 20s	0.0936**	0.0406***	0.00632
		(0.0373)	(0.0137)	(0.00643)
	SE from 30s	0.0179	0.0289**	0.0101
		(0.0185)	(0.0138)	(0.00628)
	Late career SE	-0.0160	0.000146	0.00668
		(0.0149)	(0.0187)	(0.00661)
	SE to employee	0.0560	-0.00156	0.0152
		(0.0749)	(0.0213)	(0.0158)
	Short SE spells	-0.0437*	-0.0207	0.0115
		(0.0248)	(0.0173)	(0.0108)
Women:	Always SE	0.0825***	0.0659***	-0.0110
		(0.0219)	(0.0232)	(0.0122)
	SE from 20s	0.0345	0.0150	0.00810
		(0.0245)	(0.0171)	(0.00921)
	SE from 30s	-0.00714	0.000133	0.0168*
		(0.0276)	(0.0228)	(0.0102)
	Late career SE	-0.0289*	0.0301	0.0233**
		(0.0172)	(0.0342)	(0.0111)
	SE to employee	0.0494	-0.0114	0.00635
		(0.0428)	(0.0344)	(0.0143)
	Short SE spells	0.0386	-0.0512*	0.00670
		(0.0362)	(0.0266)	(0.0150)

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: See Tables 4.D.11–4.D.14 for more detailed results.

items; see Appendix 4.B for details. The score is constructed such that its value lies between 0 and 1, where 0 indicates no material deprivation and 1 very high material deprivation.

The bottom of Table 4.9 reports the results from a random effects linear panel regression. The results are less strong than for the make ends meet question but overall I find some indication that long-term self-employed are on average reporting more material deprivation, in line with the results for making ends meet. That is, individuals with careers with a long period of self-employment face more frequently financial hardship in old age. Interestingly, there are no statistically significant result for men in countries in group 3. Since material deprivation can be considered to be an even stronger measure of financial difficulties than the making ends meet question, observing less difficulties in this country group might also be related to the relatively strong social welfare systems in many of these countries.

The results therefore underline that the exclusion of self-employed in many countries' mandatory pension systems for employees is on average not sufficiently compensated through own financial provisions by these individuals. Those with long-term self-employment careers are more frequently observed to have financial difficulties at age 60 or older than their peers that spend the majority of their career in employment and are never self-employed. This is worrying if we think of the current self-employed who, due to the change in the economy towards more service industries, on average need less capital to start their business than older generations did.[29] As such they are therefore even less likely than older generations to have wealth accumulated in their company, an observation that e.g. Mastrogiacomo et al. (2016) make for the Netherlands. Unless their savings behaviour differs greatly from the older generations, their financial situation in old age might therefore be even worse if policies are not adapted.

4.5 Conclusion

In order to find out whether self-employment careers are correlated with different outcomes in individuals' financial well-being in old age in Europe I first analyse individuals' labour market trajectories between ages 15 and 60. Using information from the Job Episodes Panel from SHARE, I classify the self-employed into different clusters using sequence analysis. I identify nine clusters of careers involving self-employment across all cohorts and countries considered: (1) always self-employed individuals, (2) those that become self-employed in their 20s and (3) 30s, (4) late career self-employed, (5) self-employed that switch to employment halfway in their trajectory, (6) those with short self-employment spells, (7) individuals with weak labour market attachment, (8) pensioners, and (9) those with mostly missing information on their trajectories.

I find differences in the shares of the self-employment clusters across country groups by their percentile rank in terms of the rule of law indicator. Self-employed in countries that rank lower are mostly individuals that turn self-employed after their thirties or late in their career. In the

[29] E.g. an individual working as a graphic designer vs. a carpenter.

countries that rank between the 80th and 90th percentile rank, the majority of self-employed are either always self-employed or turn self-employed relatively early into their careers. Last, the self-employed in the countries ranking above the 90th percentile rank most start in their twenties or thirties. In terms of demographic characteristics I show that the long term self-employed are on average less well educated than employees and also more likely to have worked in the agricultural sector.

The results from the regression analysis of two different income measures – household income and total (individual) pension income – show that even when controlling for education and sectoral activity the long term self-employed on average receive less income than employees when they are 60 years and older no matter the gender. The same holds also for late career self-employed men but only those in group 3 countries are found to have a strongly statistically significant lower pension income compared to employees. Over all, only the individuals in the cluster that switch to employment halfway in their career and those with short self-employment spells are found to have on average no statistical difference in income from employees.

More striking though is the difference between women where the results suggest that most women in the third group turn to self-employment to earn an extra income that improves the households financial situation and puts them above the average household where women are employees. This stands in contrast to women in group 1 countries where the negative coefficients in the income regression suggest that these women were rather self-employed out of necessity and not because they would earn a higher income.

In order to better understand whether clusters with lower incomes are also doing less well financially otherwise I also analyse other measures of financial well-being. First, I look at probability of households owning the dwelling they live in. I find overall no statistical evidence for self-employed being more likely to live in a self-owned dwelling. In the following analysis I also look at the household's ability to make ends meet. The results show that the long term self-employed have more difficulties making ends meet than employees. This result is also confirmed when analysing a variant of a material deprivation score.

The results found in this study therefore suggest that an inclusion of the self-employed in the pension system comparable to that of employees is needed. This is even more a concern if we think of current self-employed careers in the service economy, where individuals are unlikely to accumulate much capital in their company, in comparison with the older generations who worked more frequently in more capital intense enterprises. If older generation long-term self-employed are already more frequently having financial difficulties in old age, the risk for the younger generation may be even higher.

Table 4.A.1: Share (%) of (in)complete labour market trajectories by different categories

		Wave 3			Wave 7	
		Incomplete		Complete	Incomplete	Complete
		missing	censored			
Gender:	men	14.65	22.57	62.78	16.43	83.57
	women	30.92	20.89	48.18	24.53	75.47
Cohort:	1	22.85		77.15	22.18	77.82
	2	25.09		74.91	20.45	79.55
	3	22.95		77.05	19.70	80.30
	4	24.95	17.52	57.54	20.93	79.07
	5	21.11	78.87	0.02	21.60	78.40
Self-employed:	never	24.79	21.37	53.84	21.27	78.73
	at least 1 year	17.30	22.97	59.72	17.82	82.18
Overall		23.41	21.67	54.92	20.92	79.08

Note: The total JEP sample consists of 75,431 individuals of which 24,006 were interviewed in Wave 3 and 51,425 in Wave 7. For the two youngest cohorts interviewed in Wave 3 there are (almost) no complete trajectories because the interviews took place in all countries except for Ireland in 2009. Wave 7 interviews took place in 2017 and hence few individuals in the youngest cohort have reached age 60.

Appendices

4.A On missing observations and their inclusion

Because the JEP is a retrospective panel its usefulness relies on there being no recall bias in the life history interviews. Havari and Mazzonna (2015) have analysed the first version of the JEP, which only included Wave 3, and conclude that in general the respondents are able to remember their health and socio-economic status in their youth (ages 0–15) well. Hence one can assume that the respondents should also remember events that took place later in their life, which are more recent after all, relatively well. As Brugiavini et al. (2013) point out however, special attention should be paid to the missing data and inconsistencies regarding job spell dates. There are in fact quite a lot of individuals that have gaps in their labour market history (see share of "Unknown" in Table 4.1). Some of these missing observations are by construction (the JEP codes all following years' labour market states as missing if there is a date conflict that cannot be resolved) and because wave 3 interviews were less "thorough" in covering all episodes in individuals' lifes. Unlike the interviews of wave 7 they did not fill in information for individuals that never worked, and did not ask for information on gaps in employment longer than 6 months. Hence, excluding those individuals with missing years may likely lead to selection bias against those with weaker ties to the labour market.

Table 4.A.1 shows the shares of incomplete trajectories by different variables across the two waves of SHARELIFE. I make a difference here too between trajectories that are censored – the

individual is interviewed before they have turned 60 – and those that have missing years – the interview did not cover the period or the individual could not remember, or periods are missing because a previous job spell had missing data or date inconsistencies. The share of censored trajectories suggests that their exclusion would likely lead to little selection bias as numbers are relatively equal for both genders or across the two sub-samples of the self-employed and those that were never self-employed. However, looking at the same variables we see that the trajectories are unequally distributed between these variables. Women have a larger share of trajectories with missing information than men and so do the self-employed compared to their counterpart. Because of this it is prudent to deal with the missing years in the trajectories in a manner that allows to keep all individuals in the sample.

One last observation can be made regarding the differences between the Wave 3 and 7 interviews. While the latter are more thorough in covering all episodes in an individual's life Table 4.A.1 shows that the share of trajectories with missing years still remains relatively large among individuals interviewed in Wave 7. The change in interview structure mostly improves the trajectories of women. There is a 20% decrease among their missing trajectories but the shares still remain large overall. The same can be seen in Table 4.A.2 which shows the share of incomplete labour market trajectories by country and wave. It shows that the share of incomplete trajectories varies quite between countries but also that the change in the interview structure actually does not greatly improve the share of trajectories that miss at least one year.[30]

4.B Material deprivation score

Table 4.B.1 shows the regression results on which the hedonic weights for the material deprivation score following Bertoni et al. (2015) are based. The dependent variable in all regressions is individual life satisfaction where 0 indicated complete dissatisfaction and 10 complete satisfaction. Since the questions for the material deprivation indicators were only asked to household respondents the sample is limited to them. The first column shows the benchmark regression using the full set of indicators as used by Bertoni et al. (2015). All indicators take value one if an individual indicates that their household could not afford to buy meat, fruit, or the necessary groceries regularly, have continued to wear worn out clothes or shoes longer to save money, put up with cold to save on heating costs, not gone to the dentist because of costs, not bought (new) glasses despite of needing them, could not afford to go on a week long holiday or pay unexpected expenses, or were at least two months behind with their rent.

Because the full set of material deprivation indicators were only asked in Wave 5 of SHARE the resulting sample based on that regression is relatively small. As a robustness check I therefore also construct hedonic weights for a smaller set of indicators. The second column shows the regression result for the reduced set of indicators that are available beyond Wave 5 of SHARE

[30]The shares in both tables however do not take into account if the gaps in the trajectories are reduced in their length.

Table 4.A.2: Share (%) of (in)complete labour market trajectories by country

	Wave 3			Wave 7	
	Incomplete		Complete	Incomplete	Complete
	missing	censored			
Austria	23.58	13.11	63.30	20.35	79.65
Belgium	31.19	21.91	46.90	26.02	73.98
Bulgaria				15.89	84.11
Croatia				20.16	79.84
Cyprus				22.37	77.63
Czech Republic	8.56	24.15	67.29	12.96	87.04
Denmark	13.09	29.24	57.68	17.05	82.95
Estonia				18.33	81.67
Finland				14.59	85.41
France	26.29	22.93	50.79	24.62	75.38
Germany	31.02	20.68	48.31	29.27	70.73
Greece	13.15	27.34	59.51	20.97	79.03
Hungary				11.95	88.05
Ireland	27.63	16.12	56.26		
Israel				29.34	70.66
Italy	22.67	14.91	62.42	22.50	77.50
Latvia				17.99	82.01
Lithuania				20.21	79.79
Luxembourg				29.04	70.96
Malta				29.96	70.04
Netherlands	34.93	19.46	45.61		
Poland	29.84	25.79	44.37	25.49	74.51
Portugal				23.71	76.29
Romania				21.46	78.54
Slovakia				8.99	91.01
Slovenia				17.64	82.36
Spain	27.35	19.21	53.44	28.02	71.98
Sweden	14.28	17.35	68.38	14.87	85.13
Switzerland	26.99	24.04	48.97	27.54	72.46

Note: The total JEP sample consists of 75,431 individuals of which 24,006 were interviewed in Wave 3 and 51,425 in Wave 7.

but using the same sample. Column 3 expands the reduced model with an indicator that takes value one if an individual answered that they did not go to the doctor in the past year because of the costs of a doctor visit. Column 4 then shows the same regression but for a larger sample adding Wave 6 and 7. Last, column 5 does not include the indicator for meat consumption as it was not asked to individuals that did the retrospective interview in Wave 7 and thereby increases the sample further by all other individuals that participated in that wave.

All indicators are highly statistically significant in all regressions in which they enter. In general, as less regressors are included in the model the absolute size of the estimated coefficients increases. Overall, the coefficients are however relatively stable across the reduced models with the only exception being the indicator for rent payment arrears that decreases substantially in absolute size once Wave 6 and 7 are included in the sample too. It should also be noted that in the reduced models the hedonic weights based on the coefficients are relatively close to constant weights.

4.C Non-self-employed cluster solution

Table 4.C.1 gives an overview of how I sort the different groups based on Ward's algorithm into the seven clusters of non-self-employed across all cohorts. Figures 4.C.1–4.C.5 show Ward's solution for a 12-group partitioning for each cohort.

Table 4.B.1: Ordered probit regressions for hedonic weights for material deprivation score

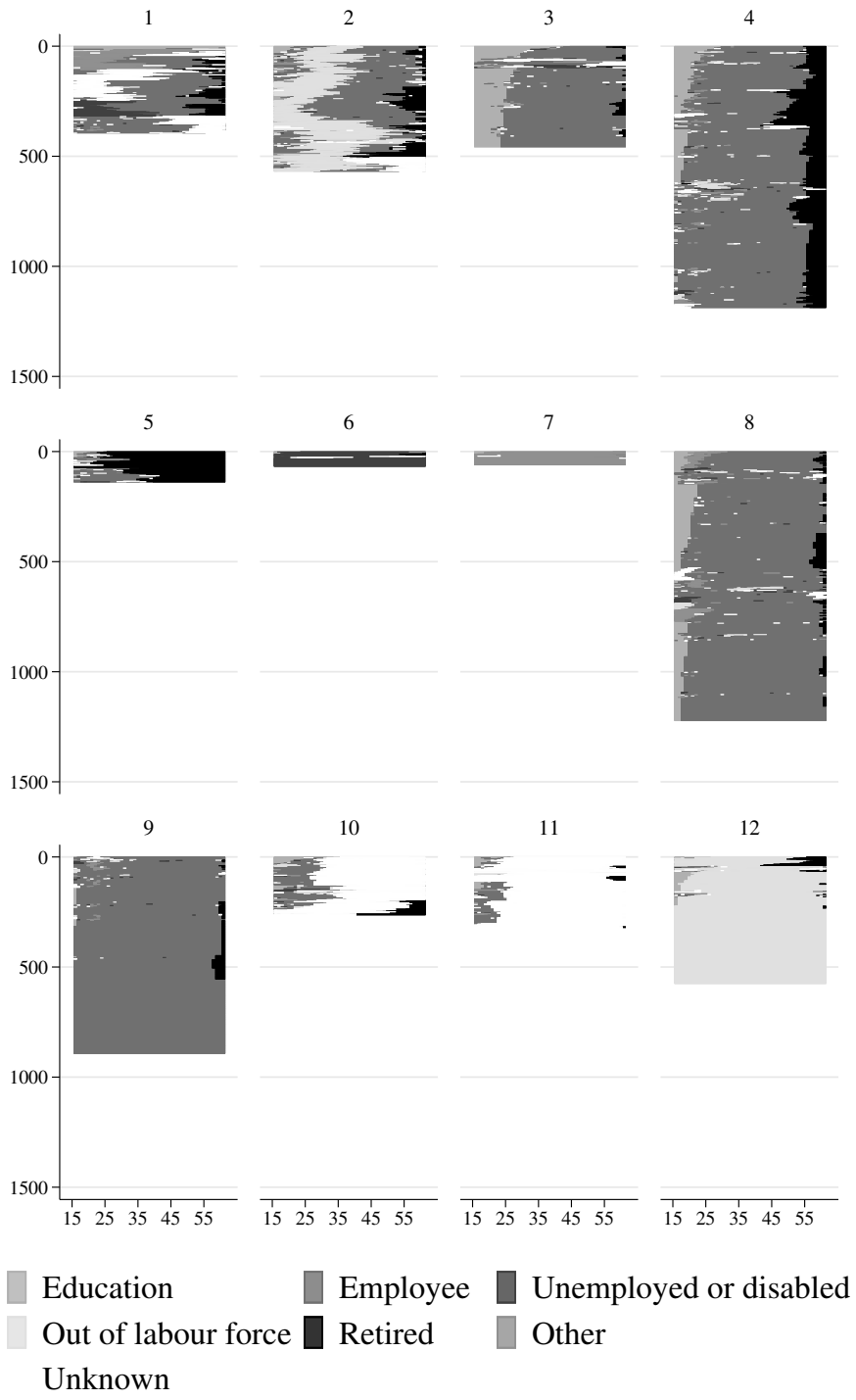
Dependent variable *Life satisfaction*	(1) Benchmark	(2) Reduced	(3) Reduced+	(4) Reduced+ all	(5) Max. sample
Meat	-0.224***	-0.382***	-0.339***	-0.318***	
	(0.0536)	(0.0508)	(0.0511)	(0.0399)	
Fruit	-0.272***				
	(0.0843)				
Groceries	-0.136***				
	(0.0253)				
Clothing	-0.164***				
	(0.0288)				
Shoes	-0.0666**				
	(0.0319)				
Heating	-0.163***	-0.312***	-0.307***	-0.301***	-0.314***
	(0.0260)	(0.0248)	(0.0248)	(0.0197)	(0.0153)
Dentist	-0.192***	-0.366***	-0.300***	-0.277***	-0.296***
	(0.0285)	(0.0264)	(0.0277)	(0.0235)	(0.0181)
Glasses	-0.149***				
	(0.0307)				
Holidays	-0.411***				
	(0.0203)				
Unexpected expenses	-0.147***	-0.430***	-0.420***	-0.411***	-0.399***
	(0.0213)	(0.0186)	(0.0187)	(0.0148)	(0.0110)
Rent	-0.444***	-0.531***	-0.507***	-0.333***	-0.353***
	(0.121)	(0.121)	(0.121)	(0.0918)	(0.0716)
Doctor			-0.283***	-0.334***	-0.337***
			(0.0365)	(0.0300)	(0.0233)
Observations	22,120	22,120	22,120	32,802	56,489
Wald Chi-squared	5415	4655	4715	6573	10508
logLikelihood	-38639	-39019	-38989	-57891	-98977

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Panel regression including country (all) and wave (column 4 and 5) fixed effects.

Table 4.C.1: Correspondence between clusters used in analysis and Ward's solution

	Cohort				
Clusters in analysis	1	2	3	4	5
Employee:	3, 8, 9	1, 4, 5	1, 2, 4	1, 9, 10	1, 2, 4
Employees with early retirement:	4	2, 3	3, 9	3, 4	3
Weak labour market attachment:	1, 2	9	8	2	6
Not active in labour force:	12	12	10	8, 12	7, 9
Pensioners:	5	8	5	5	5
Benefit recipients and other:	6, 7	6, 7	6, 7	7	8
Unknown trajectories:	10, 11	10, 11	11, 2	6, 11	10, 11, 12

Note: The cluster correspondence is for a 12-group partitioning by cohort. The solution is based on Ward's Method and implemented with Stata.



Graphs by Ward's 12 cluster solution

Figure 4.C.1: Indexplot of cohort 1 (never self-employed)

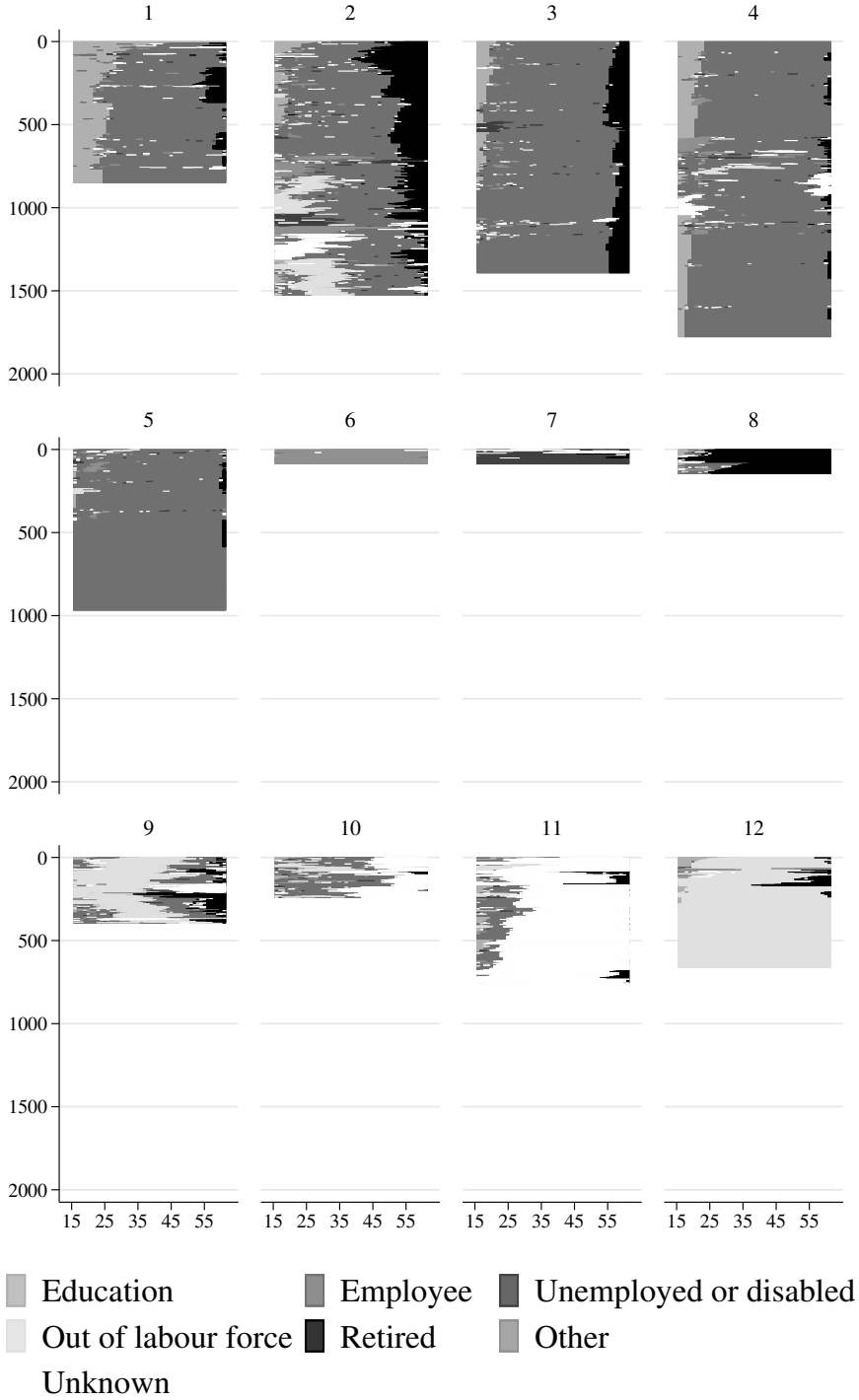
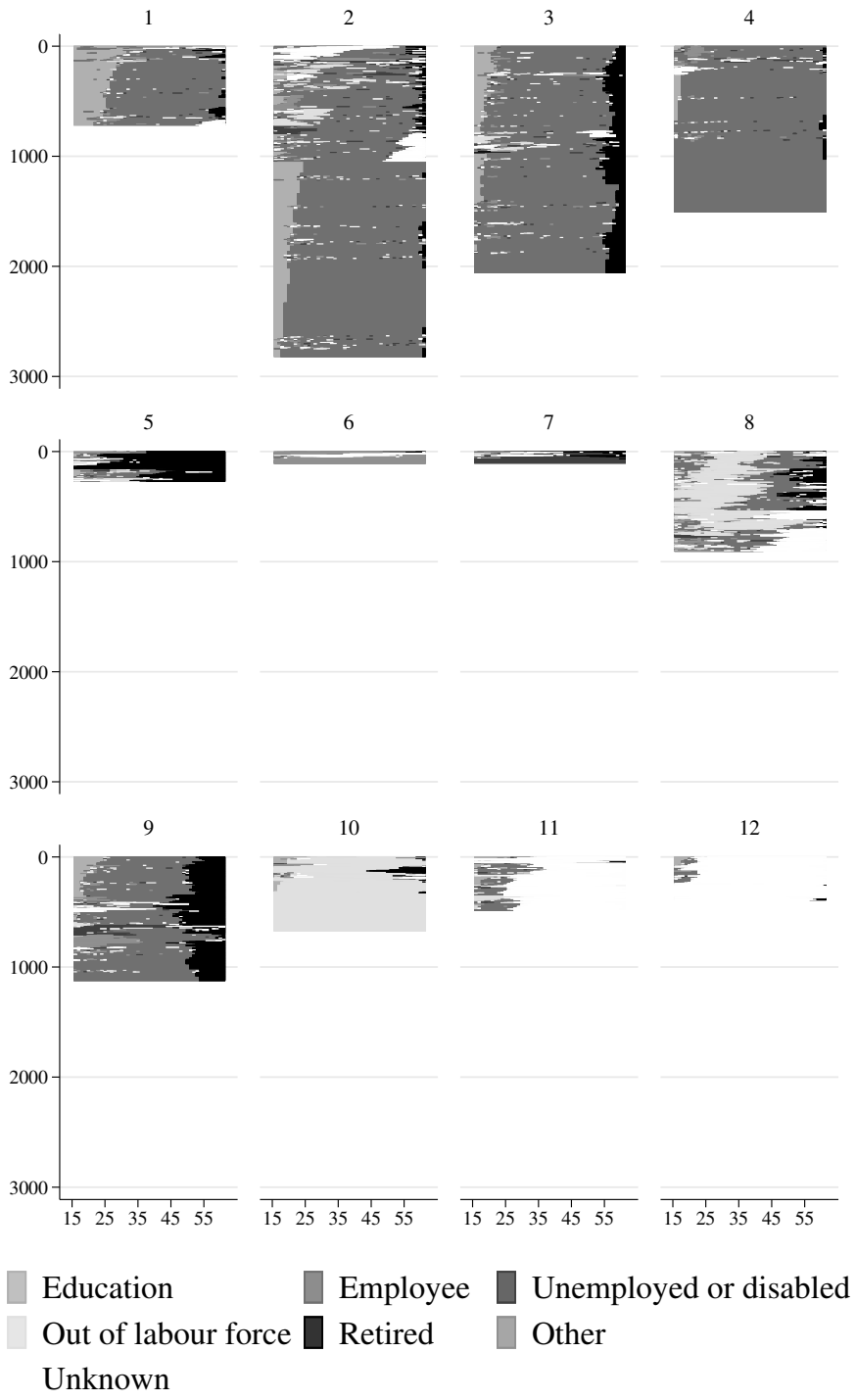
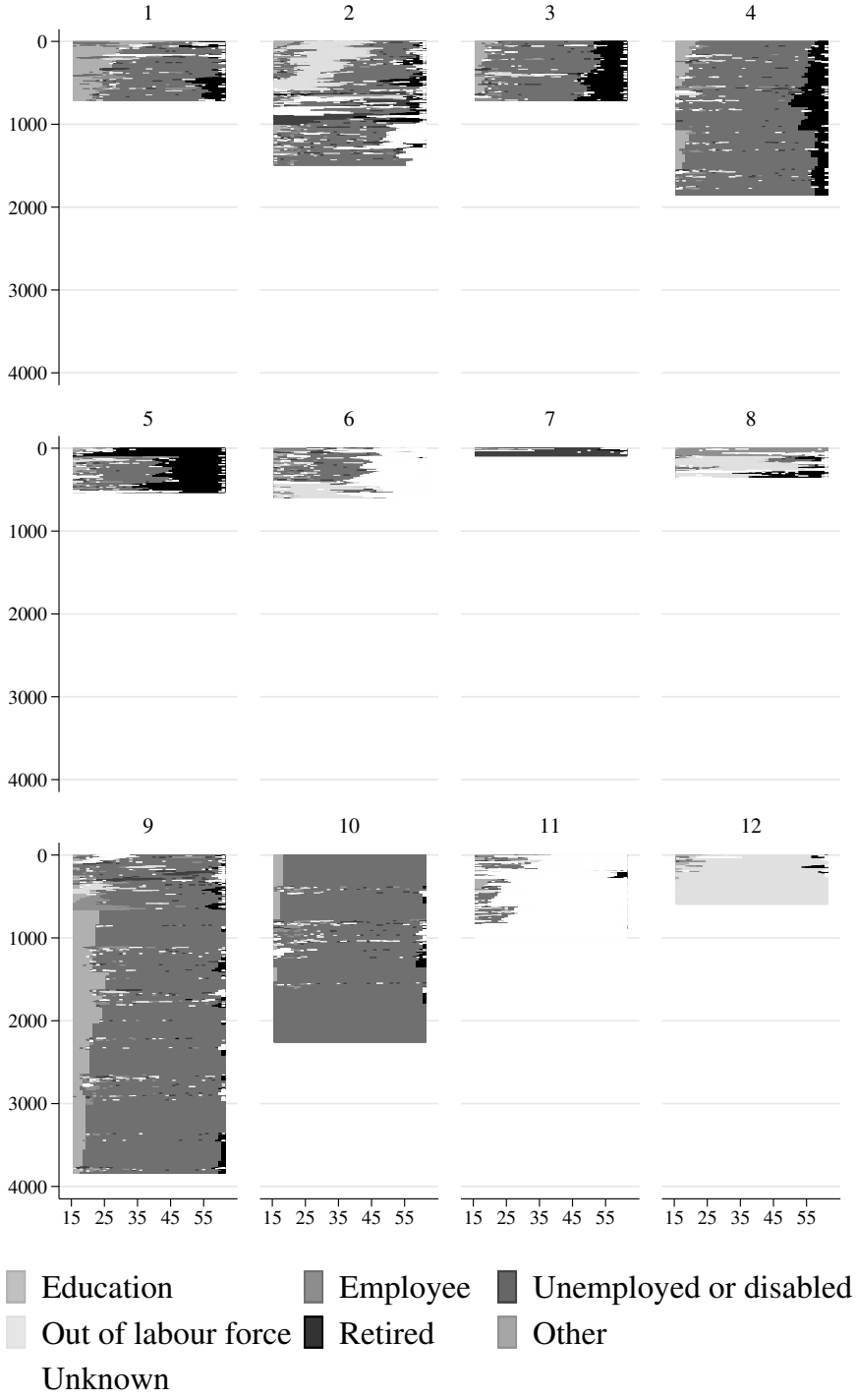


Figure 4.C.2: Indexplot of cohort 2 (never self-employed)



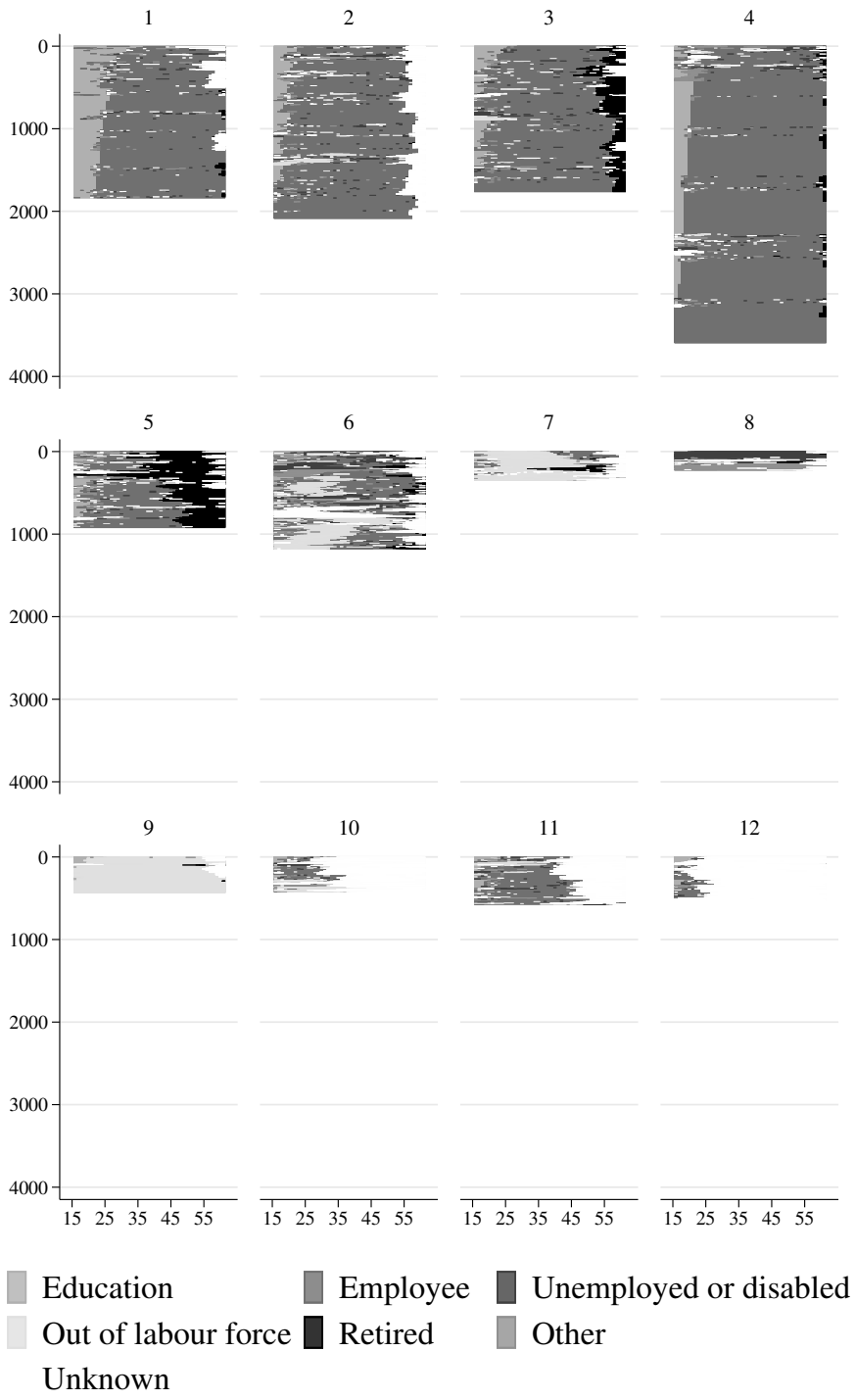
Graphs by Ward's 12 cluster solution

Figure 4.C.3: Indexplot of cohort 3 (never self-employed)



Graphs by Ward's 12 cluster solution

Figure 4.C.4: Indexplot of cohort 4 (never self-employed)



Graphs by Ward's 12 cluster solution

Figure 4.C.5: Indexplot of cohort 5 (never self-employed)

4.D Additional tables and regressions

Table 4.D.1: Age at end of education by gender in JEP sample

	1st Quartile	Median	3rd Quartile	Share (%) with missing information
Men	15	18	20	3.10
Women	14	17	19	3.87
Overall	15	17	20	3.52

Note: The total JEP sample consists of 62,708 individuals (28,313 men and 34,395 women).

Table 4.D.2: Age at retirement by gender in JEP sample

	1st Quartile	Median	3rd Quartile	Share (%) with no retirement
Men	57	60	63	24.68
Women	55	59	62	34.89
Overall	55	60	63	30.28

Note: The total JEP sample consists of 62,708 individuals (28,313 men and 34,395 women).

Table 4.D.3: Sample shares (%) of clusters by gender

	Men	Women
Always SE	16.29	15.90
SE from 20s	23.86	22.91
SE from 30s	24.55	14.12
Late career SE	15.60	10.36
SE to employee	4.62	4.97
Short SE spells	9.08	5.25
weak LM attach.	0.18	12.93
Pensioners	4.29	5.75
Unknown trajectories	1.53	7.80

Note: Sample shares are calculated over all individual-year observations for all individuals in the sample aged 60 or older that are at least one year self-employed. Total observation counts: 15,680 (m), 9821 (w).

Table 4.D.4: Self-employed clusters by cohorts & groups (in % of self-employed sub-sample)

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Overall
Percentile rank < 80						
Always SE	48.00	37.86	23.55	12.21	9.41	19.99
SE from 20s	16.80	19.42	13.13	11.45	12.76	13.76
SE from 30s	4.00	7.77	12.74	23.16	32.64	20.60
Late career SE	5.60	13.59	27.03	26.21	24.48	22.25
SE to employee	6.40	6.31	3.47	3.05	2.09	3.56
Short SE spells	9.60	6.31	3.86	7.89	5.23	6.23
weak LM attach.	3.20	1.94	3.47	3.05	1.88	2.60
Pensioners	4.00	3.88	11.97	9.41	5.65	7.39
Unknown trajectories	2.40	2.91	0.77	3.56	5.86	3.63
No. of individuals	125	206	259	393	478	1,461
80 ≤ *percentile rank* < 90						
Always SE	34.95	32.42	24.45	21.13	15.05	24.23
SE from 20s	25.54	22.46	26.84	26.61	29.03	26.36
SE from 30s	11.29	13.56	18.29	18.71	22.64	17.64
Late career SE	4.84	7.42	6.16	7.74	9.27	7.35
SE to employee	6.18	4.45	3.38	3.87	6.23	4.80
Short SE spells	4.57	7.63	5.77	6.77	6.99	6.48
weak LM attach.	3.76	3.39	5.57	4.35	3.65	4.15
Pensioners	3.23	3.18	7.16	4.84	2.28	4.11
Unknown trajectories	5.65	5.51	2.39	5.97	4.86	4.88
No. of individuals	372	472	503	620	658	2,625
Percentile rank ≥ 90						
Always SE	15.19	12.10	8.50	7.28	8.59	9.75
SE from 20s	27.12	25.79	25.00	22.64	22.46	24.23
SE from 30s	26.22	23.49	22.33	22.74	23.71	23.49
Late career SE	7.41	12.82	17.48	18.01	17.91	15.54
SE to employee	5.97	5.19	4.25	5.02	5.69	5.18
Short SE spells	5.79	6.77	8.50	8.56	8.49	7.85
weak LM attach.	4.88	5.62	4.98	7.58	4.87	5.70
Pensioners	3.62	5.62	5.34	4.43	3.21	4.42
Unknown trajectories	3.80	2.59	3.64	3.74	5.07	3.85
No. of individuals	553	694	824	1,016	966	4,053

Table 4.D.5: Clusters by cohorts and groups (in %)

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Overall
Percentile rank < 80						
All SE clusters	6.01	6.42	6.46	7.32	8.57	7.22
Employees	37.87	37.34	38.82	44.82	48.48	42.74
Early retirees	31.72	42.07	38.82	22.85	13.32	27.35
Weak LM attach. (no SE)	13.26	2.96	6.73	9.02	6.87	7.45
Out of labour force	4.47	3.64	3.32	3.47	2.33	3.26
Pensioners (no SE)	1.35	1.12	2.07	4.77	9.38	4.57
Benefits or other	1.15	0.97	1.12	0.41	1.27	0.95
Unknown traject. (no SE)	4.18	5.48	2.67	7.34	9.77	6.47
No. of individuals	2,081	3,211	4,011	5,366	5,576	20,245
80 ≤ *percentile rank* < 90						
All SE clusters	18.47	17.26	15.77	14.64	15.52	15.99
Employees	24.98	23.80	26.75	31.48	36.75	29.84
Early retirees	14.75	26.14	27.06	18.35	14.06	19.79
Weak LM attach. (no SE)	12.07	2.56	5.39	8.15	6.18	6.65
Out of labour force	13.41	12.29	10.57	10.72	8.66	10.75
Pensioners (no SE)	1.44	1.35	2.79	3.47	5.00	3.13
Benefits or other	2.93	3.18	2.82	0.99	2.17	2.25
Unknown traject. (no SE)	11.97	13.42	8.84	12.19	11.67	11.58
No. of individuals	2,014	2,735	3,189	4,234	4,240	16,412
Percentile rank ≥ 90						
All SE clusters	17.17	15.58	14.85	15.71	15.38	15.61
Employees	39.75	39.16	47.46	47.44	51.74	46.11
Early retirees	7.14	19.13	13.73	8.72	6.74	10.90
Weak LM attach. (no SE)	13.76	5.14	8.36	10.25	8.50	8.98
Out of labour force	6.58	4.69	3.64	4.65	4.46	4.64
Pensioners (no SE)	2.52	1.62	1.78	1.99	2.96	2.18
Benefits or other	1.37	1.24	1.41	0.43	0.97	1.02
Unknown traject. (no SE)	11.71	13.43	8.77	10.81	9.25	10.56
No. of individuals	3,220	4,453	5,550	6,469	6,280	25,972

Note: All SE clusters includes all self-employed clusters as shown in Table 4.D.4.

Table 4.D.6: Total household and pension income, and sample share (%) by gender and cluster

	Men				Women			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Total household income (monthly)								
All self-employed	844	277	1367	17.58	855	229	1316	9.08
Employee	873	220	1617	55.70	620	193	1312	29.70
Early retiree	1069	451	1556	17.73	486	41	1081	23.15
Weak LM att. (no SE)	705	213	1210	3.07	783	172	1362	12.07
Out of LF (no SE)	366	117	626	0.18	773	493	1101	8.73
Pensioner (no SE)	634	143	1177	2.33	579	139	1241	2.99
Benefits or other	450	110	985	0.57	843	449	1130	1.75
Unknown traj. (no SE)	525	157	1030	2.84	890	501	1286	12.55
Total pension income (annual)								
All self-employed	5217	0	11439	20.69	1341	0	7808	10.45
Employee	5007	243	15911	53.78	1947	32	9600	26.96
Early retiree	12636	2135	18173	17.23	4927	285	10239	19.08
Weak LM att. (no SE)	3818	0	12373	3.01	2160	91	8547	12.17
Out of LF (no SE)	589	0	3831	0.17	0	0	1514	11.26
Pensioner (no SE)	1608	0	11974	1.78	1078	0	6913	2.68
Benefits or other	714	0	6772	0.56	0	0	5812	1.95
Unknown traj. (no SE)	1376	0	9136	2.79	124	0	5753	15.46

Note: All values are in euro, and PPP as well as inflation adjusted. (Total household and pension income use the constant PPP scale from SHARE with Germany as the base. The base year is 2015.) Household adjusted for household size with its square root. The statistics for total household and pension income are calculated over all individual-year observations for each labour market group for all individuals in the sample aged 60 or older. (Total observation count household income: 33,870 (m), 48,329 (w), pension income: 48,070 (m), 62,444 (w).) Observation counts differ across measurements because of sample differences. See Section 4.2.2 for more information on sample selection.

Table 4.D.7: Regression results: household income (monthly), men

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	86.94***	141.3***	299.3***
	(10.61)	(18.08)	(20.76)
Tertiary education	127.9***	451.8***	581.0***
	(16.62)	(26.00)	(27.46)
Has partner	74.89***	-104.0***	37.27***
	(8.840)	(13.41)	(13.19)
Widowed	25.78**	-77.33***	133.6***
	(12.54)	(27.48)	(25.88)
LM status: working	113.0***	71.79**	47.86***
	(18.40)	(28.94)	(18.53)
other	16.94	-134.7***	-68.06***
	(15.81)	(20.92)	(22.35)
Above statutory ret. age	52.70***	4.986	43.40***
	(11.11)	(14.00)	(13.85)
Cluster: Always SE	-115.2***	-92.38***	52.60
	(39.11)	(34.88)	(54.58)
SE from 20s	-77.07	-125.5***	-86.69*
	(51.94)	(34.80)	(47.55)
SE from 30s	-20.82	-98.33**	-147.6***
	(40.79)	(42.56)	(44.15)
Late career SE	-74.21***	-122.8*	-210.6***
	(28.67)	(67.51)	(57.10)
SE to employee	-41.43	-111.3	-74.08
	(77.69)	(72.14)	(101.1)
Short SE spells	-128.2***	73.36	-100.9
	(45.53)	(64.99)	(68.86)
Weak LM attach.	-134.2***	-167.2	-278.2**
	(13.27)	(176.0)	(123.3)
Pensioners	-193.3***	1.696	-198.1**
	(48.42)	(83.87)	(97.65)
Unknown trajectories	-225.7***	-294.3***	128.5
	(23.73)	(89.56)	(167.4)
Constant	66.89***	1,062***	838.6***
	(21.82)	(30.85)	(33.34)
Observations	7,167	9,292	17,083
Nr. of individuals	4,616	4,641	7,994
σ_u	321.5	511.2	839.5
σ_e	129.1	258.5	342.2
R^2 within	0.0584	0.0128	0.0076
R^2 between	0.115	0.170	0.0962
R^2 overall	0.111	0.163	0.0849

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects panel regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Income is PPP adjusted, deflated (2015 as base year) and adjusted for household size. The sample is limited to household respondents only.

Table 4.D.8: Regression results: household income (monthly), women

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	89.72***	115.4***	261.6***
	(7.109)	(16.10)	(18.55)
Tertiary education	149.2***	327.8***	315.1***
	(11.76)	(26.51)	(26.32)
Has partner	74.00***	-39.40***	65.61***
	(6.645)	(12.20)	(11.31)
Widowed	4.271	-89.27***	49.47***
	(6.071)	(14.10)	(14.23)
LM status: working	152.7***	-29.02	81.70***
	(15.76)	(26.48)	(16.43)
other	-0.280	-66.46***	-27.04**
	(11.43)	(10.97)	(12.88)
Above statutory ret. age	-13.47	-44.41***	67.12***
	(10.09)	(13.12)	(11.72)
Cluster: Always SE	-233.9***	-183.4***	165.5***
	(22.14)	(47.40)	(61.46)
SE from 20s	-179.3***	-69.79	218.1***
	(26.36)	(48.78)	(54.26)
SE from 30s	-162.2***	-95.17	114.6**
	(56.88)	(76.80)	(57.47)
Late career SE	-60.94	-94.14	-16.33
	(39.84)	(132.7)	(72.56)
SE to employee	-94.64	95.91	31.87
	(73.91)	(104.4)	(127.1)
Short SE spells	-157.6***	402.7***	-98.57
	(56.43)	(113.2)	(105.7)
Weak LM attach.	-8.864	-178.4***	213.3***
	(84.55)	(49.43)	(61.24)
Pensioners	-164.8***	-172.6*	-144.6
	(44.51)	(90.14)	(98.04)
Unknown trajectories	-56.49	-63.53	219.6***
	(79.55)	(55.18)	(72.14)
Constant	268.9***	1,127***	607.9***
	(16.63)	(31.90)	(30.22)
Observations	15,452	11,950	20,432
Nr. of individuals	8,595	5,837	9,221
σ_u	269.7	469.8	759.4
σ_e	132.7	239.4	305.5
R^2 within	0.051	0.001	0.005
R^2 between	0.190	0.182	0.0745
R^2 overall	0.201	0.165	0.0701

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects panel regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Income is PPP adjusted, deflated (2015 as base year) and adjusted for household size. The sample is limited to household respondents only.

Table 4.D.9: Regression results: pension income (annual), men

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	426.1***	1,158**	3,691***
	(88.34)	(525.0)	(268.8)
Tertiary education	777.4***	4,858***	4,512***
	(115.1)	(601.0)	(305.7)
Has partner	366.1***	-1,496***	582.9***
	(73.80)	(454.1)	(182.0)
Widowed	547.3***	-685.8	1,140***
	(125.8)	(991.3)	(371.2)
LM status: working	-1,121***	-36,579***	-17,279***
	(89.07)	(890.2)	(280.2)
other	-2,275***	-27,057***	-14,275***
	(139.9)	(1,068)	(407.7)
Above statutory ret. age	1,300***	3,213***	4,711***
	(72.96)	(593.3)	(209.1)
Cluster: Always SE	-1,471***	-3,372***	-1,783**
	(294.1)	(1,055)	(844.5)
SE from 20s	-1,467***	-961.5	-2,275***
	(475.0)	(962.2)	(532.0)
SE from 30s	-518.0*	-2,196**	-3,039***
	(277.7)	(1,071)	(495.8)
Late career SE	-310.3	-3,112*	-3,748***
	(225.9)	(1,629)	(624.7)
SE to employee	-1,269*	-335.2	998.7
	(726.2)	(2,005)	(1,109)
Short SE spells	47.61	1,692	17.38
	(429.4)	(1,629)	(801.1)
Weak LM attach.	-374.1	20,475**	-5,815
	(2,331)	(10,043)	(8,570)
Pensioners	-911.0*	-5,276**	-2,760**
	(535.3)	(2,146)	(1,235)
Unknown trajectories	-382.7	-9,047**	-4,966**
	(874.4)	(4,292)	(2,080)
Constant	-184.1	12,139***	6,922***
	(181.2)	(1,042)	(467.2)
Observations	8,805	14,138	24,267
Censored obs.	1,736	3,486	5,479
Nr. of individuals	4,212	6,067	9,715
σ_u	2,193***	9,793***	9,218***
	(33.35)	(257.3)	(98.73)
σ_e	1,339***	17,492***	7,165***
	(15.61)	(149.8)	(49.39)
Wald Ξ^2	2525	3027	8694
logLikelihood	-64944	-121921	-202045

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects Tobit regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Pension ncome is PPP adjusted and deflated (2015 as base year).

Table 4.D.10: Regression results: pension income (annual), women

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	698.1***	648.7	1,964***
	(69.18)	(451.3)	(204.4)
Tertiary education	899.4***	3,789***	2,849***
	(99.88)	(547.0)	(245.4)
Has partner	225.4***	-902.7**	-13.94
	(57.78)	(394.2)	(151.3)
Widowed	217.9***	-1,875***	38.04
	(64.10)	(501.6)	(212.5)
LM status: working	-612.9***	-27,179***	-13,322***
	(80.15)	(884.5)	(275.7)
other	-2,629***	-23,488***	-9,400***
	(121.9)	(491.9)	(195.8)
Above statutory ret. age	903.5***	2,821***	5,631***
	(70.56)	(534.0)	(184.1)
Cluster: Always SE	-2,308***	-3,221**	-803.5
	(261.4)	(1,307)	(886.2)
SE from 20s	-2,292***	-1,769	-769.9
	(349.5)	(1,144)	(545.9)
SE from 30s	-999.0**	-5,656***	-1,317**
	(397.3)	(1,763)	(630.9)
Late career SE	-843.3***	-920.8	-979.0
	(278.6)	(2,536)	(753.4)
SE to employee	-1,879**	4,196*	-637.0
	(940.8)	(2,329)	(1,073)
Short SE spells	-1,256**	4,096*	915.9
	(507.2)	(2,463)	(1,087)
Weak LM attach.	-1,432**	-4,807***	-2,040***
	(575.6)	(1,609)	(669.1)
Pensioners	-1,573***	-7,057***	-1,991*
	(386.2)	(2,293)	(1,172)
Unknown trajectories	-1,584	-8,117***	-5,690***
	(969.1)	(1,805)	(1,095)
Constant	1,243***	7,330***	4,857***
	(157.1)	(1,019)	(395.2)
Observations	13,222	17,879	30,150
Censored obs.	1,499	7,824	9,995
Nr. of individuals	6,032	7,288	11,498
σ_u	2,059***	8,446***	7,054***
	(24.84)	(224.1)	(81.16)
σ_e	1,280***	15,369***	6,974***
	(11.32)	(131.0)	(44.53)
Wald Ξ^2	3185	4856	8838
logLikelihood	-106343	-115063	-215880

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects Tobit regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Pension income is PPP adjusted, and deflated (2015 as base year).

Table 4.D.11: Regression results: making ends meet, men

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	0.306***	0.455***	0.681***
	(0.0441)	(0.0496)	(0.0381)
Tertiary education	0.734***	1.142***	1.061***
	(0.0595)	(0.0582)	(0.0439)
Has partner	-0.0536	0.169***	-0.00894
	(0.0459)	(0.0409)	(0.0301)
Widowed	-0.0573	0.204***	0.179***
	(0.0670)	(0.0750)	(0.0565)
LM status: working	0.527***	0.196***	0.387***
	(0.0590)	(0.0577)	(0.0408)
other	-0.376***	-0.609***	-0.412***
	(0.0784)	(0.0735)	(0.0574)
Above statutory ret. age	0.290***	0.00784	-0.0135
	(0.0543)	(0.0433)	(0.0353)
Cluster: Always SE	-0.426***	-0.467***	-0.195*
	(0.154)	(0.104)	(0.110)
SE from 20s	-0.625***	-0.291***	-0.136*
	(0.174)	(0.0898)	(0.0762)
SE from 30s	-0.00869	-0.259**	-0.236***
	(0.123)	(0.104)	(0.0728)
Late career SE	0.167*	-0.114	-0.157*
	(0.100)	(0.162)	(0.0878)
SE to employee	-0.0640	-0.112	-0.0936
	(0.316)	(0.168)	(0.182)
Short SE spells	-0.201	0.225	-0.157
	(0.224)	(0.160)	(0.119)
Weak LM attach.	-0.636	-0.265	-1.567***
	(0.471)	(0.647)	(0.371)
Pensioners	-0.429	-0.0699	-0.286*
	(0.288)	(0.222)	(0.155)
Unknown trajectories	-1.129***	-0.227	-0.699**
	(0.368)	(0.409)	(0.336)
Cut 1	-1.024***	-1.396***	-2.044***
	(0.108)	(0.0954)	(0.0727)
Cut 2	0.650***	0.272***	-0.616***
	(0.107)	(0.0944)	(0.0688)
Cut 3	2.020***	1.549***	0.881***
	(0.112)	(0.0963)	(0.0694)
σ_u^2	0.936***	1.601***	1.383***
	(0.0639)	(0.0711)	(0.0508)
Observations	9,068	13,346	23,090
Nr. of individuals	5,364	5,568	9,118
logLikelihood	-10878	-15589	-22260

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects ordered probit regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Scale from 1–4 where 1 indicates making ends meet "with great difficulty" and 4 "easily". The sample is limited to household respondents only.

Table 4.D.12: Regression results: making ends meet, women

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	0.312***	0.610***	0.548***
	(0.0335)	(0.0428)	(0.0341)
Tertiary education	0.574***	1.176***	0.892***
	(0.0492)	(0.0581)	(0.0422)
Has partner	0.246***	0.118***	0.128***
	(0.0349)	(0.0373)	(0.0273)
Widowed	-0.210***	-0.0923**	-0.0773**
	(0.0377)	(0.0432)	(0.0338)
LM status: working	0.548***	0.0685	0.312***
	(0.0498)	(0.0629)	(0.0416)
other	-0.263***	-0.246***	-0.191***
	(0.0597)	(0.0368)	(0.0305)
Above statutory ret. age	0.276***	-0.112***	-0.103***
	(0.0441)	(0.0408)	(0.0308)
Cluster: Always SE	-0.436***	-0.577***	-0.271**
	(0.127)	(0.142)	(0.133)
SE from 20s	-0.0513	-0.190	-0.0619
	(0.143)	(0.118)	(0.0960)
SE from 30s	0.108	-0.292	-0.367***
	(0.170)	(0.185)	(0.0987)
Late career SE	0.256*	-0.363	-0.555***
	(0.140)	(0.273)	(0.129)
SE to employee	-0.123	0.283	-0.332*
	(0.364)	(0.248)	(0.193)
Short SE spells	-0.236	0.558**	-0.222
	(0.275)	(0.248)	(0.164)
Weak LM attach.	-0.402	-0.248	-0.209*
	(0.276)	(0.171)	(0.110)
Pensioners	-0.160	-0.210	-0.584***
	(0.213)	(0.232)	(0.163)
Unknown trajectories	-0.161	-0.121	-0.588***
	(0.446)	(0.173)	(0.150)
Cut 1	-0.664***	-1.624***	-2.037***
	(0.0916)	(0.0955)	(0.0671)
Cut 2	1.063***	0.0329	-0.541***
	(0.0920)	(0.0941)	(0.0649)
Cut 3	2.447***	1.276***	0.881***
	(0.0943)	(0.0955)	(0.0648)
σ_u^2	1.104***	1.401***	1.394***
	(0.0451)	(0.0539)	(0.0432)
Observations	19,094	17,214	28,512
Nr. of individuals	9,662	6,911	10,741
logLikelihood	-22592	-20011	-29921

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects ordered probit regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Scale from 1–4 where 1 indicates making ends meet "with great difficulty" and 4 "easily". The sample is limited to household respondents only.

Table 4.D.13: Regression results: material deprivation score, men

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	-0.0162**	-0.0335***	-0.0306***
	(0.00699)	(0.00644)	(0.00347)
Tertiary education	-0.0425***	-0.0717***	-0.0426***
	(0.00812)	(0.00685)	(0.00362)
Has partner	-0.00328	-0.0163***	-0.00219
	(0.00691)	(0.00557)	(0.00265)
Widowed	0.0320***	-0.0163	0.000116
	(0.0112)	(0.0101)	(0.00471)
LM status: working	-0.0320***	0.00385	-0.0102***
	(0.00906)	(0.00873)	(0.00357)
other	0.0705***	0.0959***	0.0464***
	(0.0158)	(0.0144)	(0.00809)
Above statutory ret. Age	-0.0425***	0.00657	0.00423
	(0.0122)	(0.00743)	(0.00406)
Cluster: Always SE	0.0173	0.0509***	-0.00512
	(0.0287)	(0.0156)	(0.00924)
SE from 20s	0.0936**	0.0406***	0.00632
	(0.0373)	(0.0137)	(0.00643)
SE from 30s	0.0179	0.0289**	0.0101
	(0.0185)	(0.0138)	(0.00628)
Late career SE	-0.0160	0.000146	0.00668
	(0.0149)	(0.0187)	(0.00661)
SE to employee	0.0560	-0.00156	0.0152
	(0.0749)	(0.0213)	(0.0158)
Short SE spells	-0.0437*	-0.0207	0.0115
	(0.0248)	(0.0173)	(0.0108)
Weak LM attach.	-0.0109	0.0403***	0.161
	(0.0762)	(0.00883)	(0.141)
Pensioners	0.114	0.00116	0.0321**
	(0.0852)	(0.0321)	(0.0156)
Unknown trajectories	0.164***	0.0230	0.0567*
	(0.0597)	(0.0501)	(0.0312)
Constant	0.146***	0.132***	0.0604***
	(0.0168)	(0.0119)	(0.00596)
Observations	3,268	6,665	11,765
Number of id	2,273	4,199	6,768
σ_u	0.101	0.140	0.0824
σ_e	0.115	0.115	0.0828
R^2 within	0.0283	0.0041	0.0017
R^2 between	0.119	0.0868	0.0688
R^2 overall	0.108	0.0723	0.0570

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects panel regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Lower values indicate more material deprivation. The sample is limited to household respondents only.

Table 4.D.14: Regression results: material deprivation score, women

	Counry Group 1 *Rule of Law indicator: (pct rank < 80)*	*Country Group 2* *(80 ≤ pct rank < 90)*	*Country Group 3* *(90 ≤ pct rank)*
Education: Upper/post-secondary	-0.0173***	-0.0500***	-0.0275***
	(0.00567)	(0.00629)	(0.00333)
Tertiary education	-0.0408***	-0.0758***	-0.0392***
	(0.00755)	(0.00703)	(0.00369)
Has partner	-0.0221***	-0.00676	-0.0131***
	(0.00577)	(0.00592)	(0.00288)
Widowed	0.0149**	0.0152**	0.00291
	(0.00618)	(0.00695)	(0.00374)
LM status: working	-0.0223***	-0.00611	-0.0156***
	(0.00819)	(0.00982)	(0.00405)
other	0.0484***	0.0445***	0.0216***
	(0.0122)	(0.00668)	(0.00401)
Above statutory ret. Age	-0.0569***	0.0277***	0.00276
	(0.0104)	(0.00775)	(0.00400)
Cluster: Always SE	0.0825***	0.0659***	-0.0110
	(0.0219)	(0.0232)	(0.0122)
SE from 20s	0.0345	0.0150	0.00810
	(0.0245)	(0.0171)	(0.00921)
SE from 30s	-0.00714	0.000133	0.0168*
	(0.0276)	(0.0228)	(0.0102)
Late career SE	-0.0289*	0.0301	0.0233**
	(0.0172)	(0.0342)	(0.0111)
SE to employee	0.0494	-0.0114	0.00635
	(0.0428)	(0.0344)	(0.0143)
Short SE spells	0.0386	-0.0512*	0.00670
	(0.0362)	(0.0266)	(0.0150)
Weak LM attach.	0.0276	0.0285	-0.00317
	(0.0450)	(0.0238)	(0.00911)
Pensioners	-0.0154	0.0360	0.0300
	(0.0282)	(0.0343)	(0.0193)
Unknown trajectories	0.0107	0.000850	0.0225
	(0.0868)	(0.0270)	(0.0164)
Constant	0.208***	0.114***	0.0748***
	(0.0142)	(0.0133)	(0.00610)
Observations	7,195	8,894	15,195
Number of id	4,465	5,411	8,409
σ_u	0.116	0.154	0.0927
σ_e	0.130	0.132	0.0995
R^2 within	0.0149	0.0049	0.0024
R^2 between	0.0712	0.117	0.0503
R^2 overall	0.0570	0.102	0.0396

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Random effects panel regression including cohort and wave fixed effects, and controlling for non-SE clusters and primary sector. Higher values indicate more material deprivation. The sample is limited to household respondents only.

4.E Additional figures

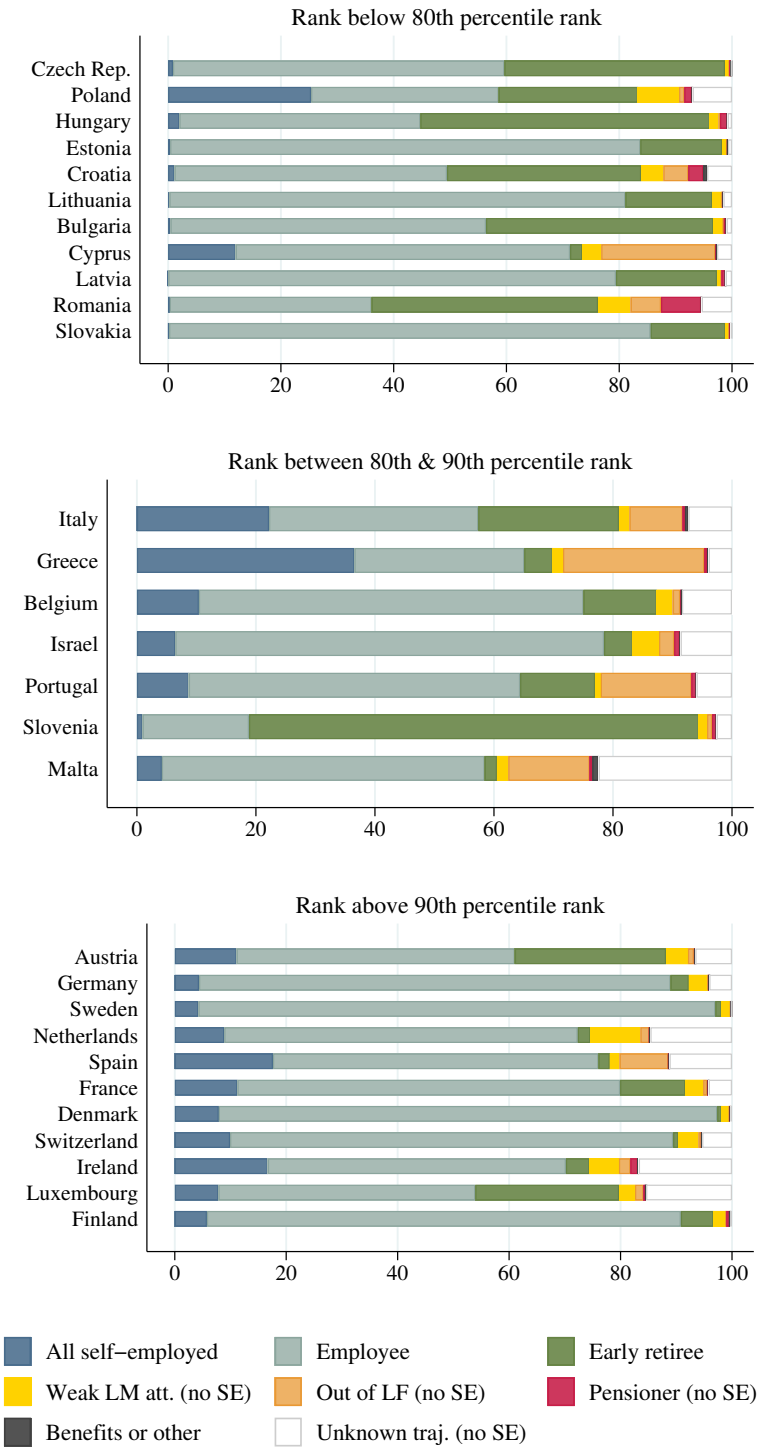


Figure 4.E.1: Share (in %) of clusters in countries

Bibliography

Abbott, A. (1983): "Sequences of social events: Concepts and methods for the analysis of order in social processes," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 16, 129–147.

——— (1995): "Sequence analysis: new methods for old ideas," *Annual Review of Sociology*, 21, 93–113.

Abbott, A. and J. Forrest (1986): "Optimal matching methods for historical sequences," *The Journal of Interdisciplinary History*, 16, 471–494.

Abbott, A. and A. Hrycak (1990): "Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers," *American Journal of Sociology*, 96, 144–185.

Albarrán, P., R. Carrasco, and J. M. Carro (2019): "Estimation of Dynamic Nonlinear Random Effects Models with Unbalanced Panels," *Oxford Bulletin of Economics and Statistics*, 81, 1424–1441.

Aldenderfer, M. and R. Blashfield (1984): *Cluster Analysis*, no. 44 in Quantitative Applications in the Social Sciences, SAGE Publications, Inc.

Antonova, L., L. Aranda, G. Pasini, and E. Trevisan (2014): "Migration, family history and pension: the second release of the SHARE Job Episodes Panel," Working Paper Series 18, SHARE-ERIC.

Been, J. and M. Knoef (2017): "Job-Search Requirements for Unemployed at the End of Working Life Effects on Unemployment Dynamics and Self-Employment Probabilities," *Journal of Human Resources*, 52, 491–530.

Bertoni, M., D. Cavapozzi, M. Celidoni, and E. Trevisan (2015): "Assessing the material deprivation of older Europeans," *Ageing in Europe – supporting policies for an inclusive society*.

Beugelsdijk, S. and N. Noorderhaven (2005): "Personality characteristics of self-employed; an empirical study," *Small Business Economics*, 24, 159–167.

Beusch, E. and A. van Soest (2020): "Labour Market Trajectories of the Self-employed in the Netherlands," *De Economist*, 168, 109–146.

BHAT, C. R. (2001): "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research Part B*, 35, 677–693.

——— (2003): "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences," *Transportation Research Part B*, 37, 837–855.

BLANCHFLOWER, D. G. (2000): "Self-employment in OECD countries," *Labour Economics*, 7, 471–505.

BOLHAAR, J., A. BROUWERS, AND B. SCHEER (2016): "De flexibele schil van de Nederlandse arbeidsmarkt: een analyse op basis van microdata," CPB Achtergronddocument, CPB.

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND B. TER WEEL (2008): "The economics and psychology of personality traits," *Journal of Human Resources*, 43, 972–1059.

BÖRSCH-SUPAN, A. (2019a): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.0.0," .

——— (2019b): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.0.0," .

——— (2019c): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.0.0," .

——— (2019d): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.0.0," .

——— (2019e): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.0.0," .

——— (2019f): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.0.0," .

BOSCH, N. (2014): "Succes als startende zelfstandige," CPB Achtergronddocument, CPB.

BOSCH, N., G. ROELOFS, D. VAN VUUREN, AND M. WILKENS (2012): "De huidige en toekomstige groei van het aandeel zzp'ers in de werkzame beroepsbevolking," CPB Achtergronddocument, CPB.

BOSCH, N. AND D. VAN VUUREN (2010): "De heterogeniteit van zzp'ers," *Economisch Statistische Berichten*, 95, 682–684.

BRUGIAVINI, A., D. CAVAPOZZI, G. PASINI, AND E. TREVISAN (2013): "Working life histories from SHARELIFE: a retrospective panel," Working Paper Series 11, SHARE-ERIC.

BRUGIAVINI, A., C. E. ORSO, M. G. GENIE, R. NACI, AND G. PASINI (2019): "Combining the retrospective interviews of wave 3 and wave 7: the third release of the SHARE Job Episodes Panel," Working Paper Series 36, SHARE-ERIC.

BRZINSKY-FAY, C., U. KOHLER, AND M. LUNIAK (2006): "Sequence analysis with Stata," *The Stata Journal*, 6, 435–460.

BUDDELMEYER, H. AND M. WOODEN (2011): "Transitions out of casual employment: the Australian experience," *Industrial Relations: A Journal of Economy and Society*, 50, 109–130.

CBS (2014): "Achtergrondkenmerken en ontwikkelingen van zzp'ers in Nederland," cbs.nl.

CHENG, T. C. AND P. K. TRIVEDI (2015): "Attrition bias in panel data: a sheep in wolf's clothing? A case study based on the Mabel survey," *Health Economics*, 24, 1101–1117.

CHOI, J. (2009): "Pension schemes for the self-employed in OECD countries," OECD Social, Employment and Migration Working Papers 84, OECD.

CHRISTELIS, D. AND R. FONSECA (2016): "Labor Market Policies and Self-Employment Transitions of Older Workers," .

COBB-CLARK, D. A. AND S. SCHURER (2012): "The stability of Big-Five personality traits," *Economics Letters*, 115, 11–15.

DE BRESSER, J. AND M. KNOEF (2015): "Can the Dutch meet their own retirement expenditure goals?" *Labour Economics*, 34, 100–117.

EUROPEAN COMMISSION (2018): "Pension adequacy report 2018 – Current and future income adequacy in old age in the EU (Volume 1)," Tech. rep.

GOLDBERG, L. R. (1992): "The development of markers for the Big-Five factor structure," *Psychological Assessment*, 4, 26–42.

GONG, X., A. VAN SOEST, AND E. VILLAGOMEZ (2004): "Mobility in the urban labor market: a panel data analysis for Mexico," *Economic Development and Cultural Change*, 53, 1–36.

HALPIN, B. (2017): "SADI: Sequence analysis tools for Stata," *The Stata Journal*, 17, 546–572.

HAVARI, E. AND F. MAZZONNA (2015): "Can we trust older people's statements on their childhood circumstances? Evidence from SHARELIFE," *European Journal of Population*, 31, 233–257.

HECKMAN, J. J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5, 475–492.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

——— (1981a): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski and D. L. McFadden, MIT Press, Cambridge, MA, 179–195.

——— (1981b): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski and D. L. McFadden, MIT Press, Cambridge, MA, 114–178.

HOLLISTER, M. (2009): "Is optimal matching suboptimal?" *Sociological Methods & Research*, 38, 235–264.

HUMPHRIES, J. E. (2016): "The Causes and Consequences of Self-Employment over the Life Cycle," Mimeo.

JÜRGES, H. (2007): "True health vs response styles: exploring cross-country differences in self-reported health," *Health Economics*, 16, 163–178.

KABÁTEK, J. (2020): "Happy birthday, you're fired! The effects of an age-dependent minimum wage on youth employment flows in the Netherlands," *ILR Review, forthcoming*.

KAUFMAN, D., A. KRAAY, AND M. MASTRUZZI (2010): "The Worldwide Governance Indicators: Methodology and Analytical Issues," Tech. rep., world Bank Policy Research Working Paper No. 5430.

KNOEF, M., J. BEEN, R. ALESSIE, K. CAMINADA, K. GOUDSWAARD, AND A. KALWIJ (2016): "Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands," *Journal of Pension Economics & Finance*, 15, 55–89.

LEVINE, R. AND Y. RUBINSTEIN (2017): "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics*, 132, 963–1018.

MADERO-CABIB, I. AND A. E. FASANG (2016): "Gendered work–family life courses and financial well-being in retirement," *Advances in Life Course Research*, 27, 43–60.

MASTROGIACOMO, M. (2016): "De pensioenpuzzel van zelfstandigen," Netspar Brief 7, Netspar.

MASTROGIACOMO, M. AND R. J. ALESSIE (2015): "Where are the retirement savings of self-employed? An analysis of 'unconventional' retirement accounts," Working Paper 454, De Nederlandsche Bank.

MASTROGIACOMO, M., Y. LI, AND R. DILLINGH (2016): "Entrepreneurs without wealth? An overview of their portfolio using different data sources for the Netherlands," Netspar Design Paper 37, Netspar.

McDonough, P., D. Worts, L. M. Corna, A. McMunn, and A. Sacker (2017): "Later-life employment trajectories and health," *Advances in Life Course Research*, 34, 22–33.

Munnell, A., G. Sanzenbacher, and A. Walters (2019): "How Do Older Workers Use Nontraditional Jobs?" Forthcoming from Journal of Pension Economics and Finance (Cambridge University Press).

Obschonka, M., E. Schmitt-Rodermund, R. K. Silbereisen, S. D. Gosling, and J. Potter (2013): "The regional distribution and correlates of an entrepreneurship-prone personality profile in the United States, Germany, and the United Kingdom: A socioecological perspective." *Journal of Personality and Social Psychology*, 105, 104.

OECD (2017): *Pensions at a Glance 2017: OECD and G20 Indicators*, OECD publishing.

——— (2019): *Pensions at a Glance 2019: OECD and G20 Indicators*, OECD publishing.

Oguzoglu, U. (2016): "Disability and Multi-State Labour Force Choices with State Dependence," *Economic Record*, 92, 28–46.

Parker, S. C. (2004): *The Economics of Self-employment and Entrepreneurship*, Cambridge University Press.

Pettinicchi, Y. and A. Börsch-Supan (2019): "Retirement Income Adequacy of Traditionally Employed and Self-Employed Workers: Analyses with SHARE Data," MEA Discussion Paper 02-2019, MEA Munich Center for the Economics of Aging.

Ponomarenko, V. (2016): "Cumulative disadvantages of non-employment and non-standard work for career patterns and subjective well-being in retirement," *Advances in Life Course Research*, 30, 133–148.

Prowse, V. (2012): "Modeling employment dynamics with state dependence and unobserved heterogeneity," *Journal of Business & Economic Statistics*, 30, 411–431.

Rietveld, C. A., H. van Kippersluis, and A. R. Thurik (2015): "Self-employment and health: Barriers or benefits?" *Health Economics*, 24, 1302–1313.

Scherer, S. (2001): "Early career patterns: A comparison of Great Britain and West Germany," *European Sociological Review*, 17, 119–144.

SER (2010): "ZZP'ers in beeld," SER Advies 10/04, Sociaal-Economische Raad.

Studer, M. (2013): "WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R," LIVES Working Papers 24, University of Geneva Institute for Demographic and Life Course Studies.

STUDER, M. AND G. RITSCHARD (2016): "What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179, 481–511.

SZW (2016): "Perspectiefnota toekomst pensionstelsel," Appendix to the letter to the house of representatives from 8 july 2016, Ministerie van Sociale Zaken en Werkgelegenheid.

THE WORLD BANK (2018): "Consumer Price Index," Data retrieved from World Development Indicators, `http://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?locations=NL`.

TOPHOVEN, S. AND A. TISCH (2016): "Employment trajectories of German baby boomers and their effect on statutory pension entitlements," *Advances in Life Course Research*, 30, 90–110.

TRAIN, K. E. (2009): *Discrete Choice Methods with Simulation*, Cambridge University Press.

VAN WINKLE, Z. AND A. FASANG (2017): "Complexity in employment life courses in Europe in the twentieth century – Large cross-national differences but little change across birth cohorts," *Social Forces*, 96, 1–30.

VERBEEK, M. AND T. NIJMAN (1992): "Testing for selectivity bias in panel data models," *International Economic Review*, 33, 681–703.

VISSER, M. (2018): "Toenemende complexiteit van late beroepsloopbanen?" *Tijdschrift voor Arbeidsvraagstukken*, 34, 511–525.

WARD, J. H. J. (1963): "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, 58, 236–244.

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

——— (2007): "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141, 1281–1301.

ZACHER, H., T. BIEMANN, M. M. GIELNIK, AND M. FRESE (2012): "Patterns of entrepreneurial career development: An optimal matching analysis approach," *International Journal of Developmental Science*, 6, 177–187.

ZUCCHELLI, E., M. N. HARRIS, AND X. ZHAO (2012): "Ill-health and transitions to part-time work and self-employment among older workers," Tech. rep., Lancaster University WP 12/04, available at SSRN 2269449.

ZWINKELS, W., M. KNOEF, J. BEEN, C. CAMINADA, AND K. GOUDSWAARD (2017): "Zicht op zzp-pensioen," Netspar Design Paper 91, Netspar.

CENTER DISSERTATION SERIES

CentER for Economic Research, Tilburg University, the Netherlands

No.	Author	Title	ISBN	Published
579	Julius Rüschenpöhler	Behavioural Perspectives on Subsistence Entrepreneurship in Emerging Markets	978 90 5668 580 5	January 2019
580	Khulan Altangerel	Essays on Immigration Policy	978 90 5668 581 2	January 2019
581	Kun Zheng	Essays on Duration Analysis and Labour Economics	978 90 5668 582 9	January 2019
582	Tatiana Zabara	Evolution of Entrepreneurial Teams in Technology-Based New Ventures	978 90 5668 583 6	February 2019
583	Yifan Yu	Essays on Mixed Hitting-Time Models	978 90 5668 584 3	April 2019
584	Daniel Martinez Martin	Unpacking Product Modularity, Innovation in Distributed Innovation Teams	978 90 5668 585 0	April 2019
585	Katalin Katona	Managed Competition in Practice Lessons for Healthcare Policy	978 90 5668 586 7	April 2019
586	Serhan Sadikoglu	Essays in Econometric Theory	978 90 5668 587 4	May 2019
587	Hoang Yen Nguyen	Emotions and Strategic Interactions	978 90 5668 588 1	May 2019
588	Ties de Kok	Essays on reporting and information processing	978 90 5668 589 8	May 2019
589	Yusiyu Wang	Regulation, Protest, and Spatial Economics	978 90 5668 590 4	June 2019
590	Ekaterina Neretina	Essays in Corporate Finance, Political Economy, and Competition	978 90 5668 591 1	June 2019
591	Ruth Wandhöfer	Technology innovation in Financial Markets: Implications for Money, Payments and Settlement Finality	978 90 5668 592 8	June 2019

No.	Author	Title	ISBN	Published
592	Andinet Worku Gebreselassie	On communicating about taboo social issues in least developed countries: The case of Ethiopia	978 90 5668 593 5	June 2019
593	Filip Bekjarovski	Active Investing	978 90 5668 594 2	June 2019
594	Miguel Sarmiento	Essays on Banking, Financial Intermediation and Financial Markets	978 90 5668 595 9	June 2019
595	Xiaoyin Ma	Essays on Alternative Investements	978 90 5668 596 6	June 2019
596	Victor van Pelt	A Dynamic View of Management Accounting Systems	978 90 5668 597 3	June 2019
597	Shuai Chen	Marriage, Minorities, and Mass Movements	978 90 5668 598 0	July 2019
598	Ben Gans	Stabilisation operations as complex systems: order and chaos in the interoperability continuum	978 90 5668 599 7	July 2019
599	Mulu Hundera	Role Conflict, Coping Strategies and Female Entrepreneurial Success in Sub-Saharan Africa	978 90 5668 600 0	August 2019
600	Hao Hu	The Quadratic Shortest Path Problem – Theory and Computations	978 90 5668 601 7	September 2019
601	Emerson Erik Schmitz	Essays on Banking and International Trade	978 90 5668 602 4	September 2019
602	Olga Kuryatnikova	The many faces of positivity to approximate structured optimization problems	978 90 5668 603 1	September 2019
603	Sander Gribling	Applications of optimization to factorization ranks and quantum information theory	978 90 5668 604 8	September 2019
604	Camille Hebert	Essays on Corporate Ownership and Human Capital	978 90 5668 605 5	October 2019
605	Gabor Neszveda	Essays on Behavioral Finance	978 90 5668 606 2	October 2019

No.	Author	Title	ISBN	Published
606	Ad van Geesbergen	Duurzame schaarste - Een kritische analyse van twee economische duurzaamheids-paradigma's geïnspireerd door de filosofie van Dooyeweerd	978 90 5668 607 9	October 2019
607	Richard T. Mason	Digital Enrollment Architecture and Retirement Savings Decisions: Evidence from the Field	978 90 5668 608 6	November 2019
608	Ron Triepels	Anomaly Detection in the Shipping and Banking Industry	978 90 5668 609 3	November 2019
609	Feng Fang	When performance shortfall arises, contract or trust? A multi-method study of the impact of contractual and relation governances on performance in Public-Private Partnerships	978 90 5668 610 9	November 2019
610	Yasir Dewan	Corporate Crime and Punishment: The Role of Status and Ideology	978 90 5668 611 6	November 2019
611	Mart van Hulten	Aiming for Well-Being through Taxation: A Framework of Caution and Restraint for States	978 90 5668 612 3	December 2019
612	Carlos Sandoval Moreno	Three essays on poverty measurement and risk protection	978 90 5668 613 0	December 2019
613	Harmke de Groot	Core strength or Achilles' heel: Organizational competencies and the performance of R&D collaborations	978 90 5668 614 7	December 2019
614	Peter Brok	Essays in Corporate Finance and Corporate Taxation	978 90 5668 615 4	December 2019
615	Pascal Böni	On the Pricing, Wealth Effects and Return of Private Market Debt	978 90 5668 616 1	December 2019
616	Ana Martinovici	Revealing Attention: How Eye Movements Predict Brand Choice and Moment of Choice	978 90 5668 617 8	December 2019
617	Matjaz Maletic	Essays on international finance and empirical asset pricing	978 90 5668 618 5	January 2020
618	Zilong Niu	Essays on Asset Pricing and International Finance	978 90 5668 619 2	January 2020

No.	Author	Title	ISBN	Published
619	Bjorn Lous	On free markets, income inequality, happiness and trust	978 90 5668 620 8	January 2020
620	Clemens Fiedler	Innovation in the Digital Age: Competition, Cooperation, and Standardization	978 90 5668 621 5	Oktober 2020
621	Andreea Popescu	Essays in Asset Pricing and Auctions	978 90 5668 622 2	June 2020
622	Miranda Stienstra	The Determinants and Performance Implications of Alliance Partner Acquisition	978 90 5668 623 9	June 2020
623	Lei Lei	Essays on Labor and Family Economics in China	978 90 5668 624 6	May 2020
624	Farah Arshad	Performance Management Systems in Modern Organizations	978 90 5668 625 3	June 2020
625	Yi Zhang	Topics in Economics of Labor, Health, and Education	978 90 5668 626 0	June 2020
626	Emiel Jerphanion	Essays in Economic and Financial decisions of Households	978 90 5668 627 7	July 2020
627	Richard Heuver	Applications of liquidity risk discovery using financial market infrastructures transaction archives	978 90 5668 628 4	September 2020
628	Mohammad Nasir Nasiri	Essays on the Impact of Different Forms of Collaborative R&D on Innovation and Technological Change	978 90 5668 629 1	August 2020
629	Dorothee Hillrichs	On inequality and international trade	978 90 5668 630 7	September 2020
630	Roland van de Kerkhof	It’s about time: Managing implementation dynamics of condition-based maintenance	978 90 5668 631 4	October 2020
631	Constant Pieters	Process Analysis for Marketing Research	978 90 5668 632 1	December 2020
632	Richard Jaimes	Essays in Macroeconomic Theory and Natural Resources	978 90 5668 633 8	November 2020

No.	Author	Title	ISBN	Published
633	Olivier David Armand Zerbib	Asset pricing and impact investing with pro-environmental preferences	978 90 5668 634 5	November 2020
634	Laura Capera Romero	Essays on Competition, Regulation and Innovation in the Banking Industry	978 90 5668 635 2	December 2020
635	Elisabeth Beusch	Essays on the Self-Employed in the Netherlands and Europe	978 90 5668 636 9	December 2020

This dissertation consists of three essays, each aiming to better understand the heterogeneity among self-employed individuals, and the potential implications of self-employment over the life-cycle for individuals' pension preparedness. The first essay models the transitions into and out of self-employment in the Netherlands. The results show that simulations that project future pension incomes and pension adequacy should account for labour market dynamics because the paths that "static" projections assume are not representative for many individuals. The second essay studies realised labour market trajectories of Dutch individuals using the technique of sequence analysis. Seven distinct clusters of self-employment experiences over the life-cycle can be identified. These are associated with different financial outcomes, implying that targeted policies are needed to address retirement income adequacy of the self-employed. The last essay studies self-employment careers and financial well-being in old age across several European cohorts. Overall, the results show that individuals that were mostly self-employed during their career are on average more frequently observed to have financial difficulties after retirement.

Elisabeth Beusch (Zurich, Switzerland, 1987) obtained her Bachelor and Master degree in Business Administration and Economics from Zurich University. She worked as a research assistant at the Swiss National Bank before joining the Research Master program in Economics at the CentER Graduate School of Tilburg University in 2013. After obtaining her M.Sc. degree in 2015 she joined the Department of Econometrics and Operations Research as a PhD candidate where she was supervised by prof. dr. Arthur van Soest and prof. dr. Theo Nijman. Since June 2020 she works as a Risk Modeling & Analytics Specialist at UBS, Switzerland.

ISBN: 978 90 5668 636 9
DOI: 10.26116/center-lis-2016