

Labour Market Trajectories of the Self-employed in the Netherlands

Elisabeth Beusch¹ · Arthur van Soest¹

Published online: 7 February 2020 © The Author(s) 2020

Abstract

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.

Keywords Sequence analysis · Self-employment · Pensions · Life-cycle

JEL Classification $~J62\cdot J32\cdot J23\cdot L26\cdot D15$

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 accumulate.² Policy

Elisabeth Beusch e.beusch@uvt.nl

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

 $^{^2}$ See Online "Appendix D" for more information on the institutional background of the Dutch pension system.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s1064 5-020-09358-x) contains supplementary material, which is available to authorized users.

¹ Netspar, Tilburg University, Tilburg, Netherlands

makers are therefore concerned that the self-employed are not building up enough pension wealth until their retirement (See SZW 2016, Sect. 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 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 1930s. Munnell et al. (2019) follow individuals in the US from age 50–62 to assess how they use nontraditional 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 wellbeing 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 5-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 Sect. 2 we describe the different data sources and give our definition of self-employment. Section 3 introduces the methods and presents the self-employment clusters that we find. Section 4 discusses the differences between the clusters with respect to demographic characteristics, income and wealth, while Sect. 5 investigates pension accumulation across clusters. Section 6 concludes.

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 1989⁴ 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 the individuals to data on household wealth (VEHTAB) and (for a subsample) to accrued pension rights in the second pillar ("Pensioendeelnemingen").

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 E" for more details on the income data and the breaks in IPO.

We divide the IPO core individuals into 5-year birth cohorts and limit our analysis to those for whom we have at least 14 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 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 1–4, 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.

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 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 DGA¹⁰), *benefit recipients* (based on the

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.

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 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).

 Table 1 Overview of cohorts in sample

Cohort	Birth years	Age	Years sequenced	Т	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					50,837	13,362 (26)

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 1 shows that the shares of individuals in each cohort who have spent at least 1 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.

2.2 Descriptive Statistics

Table 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

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.

	Average age	in %	in %								
		Men	Bread- winner	Single ^a	Dutch	Western	Non-west- ern	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				

Table 2 Demographic characteristics across labour market states

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

^aIncludes singles, divorced and widowed individuals

and are also more frequently 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 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% in 1989 to 16.5% in 2017.

Figure 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 fades out over 5 years. Similarly, the trough in the early 2000s is due to the youngest cohort, which enters the sample

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

¹⁴ As in Fig. 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.



Fig. 1 Distribution of labour market states over time



Fig. 2 Self-employment share and contributions of cohorts

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 selfemployment 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 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 45,000 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 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

¹⁵ The income data presented in Table 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.

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

^a 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	Taxable	individual	and	disposable	household	income,	wealth	and	sample	share	by	gender	and
labour s	state												

	Women				Men				
	Median	Q_1	Q_3	Share (%)	Median	Q_1	Q_3	Share (%)	
Individual tax	xable incom	e							
No income ^a	0	0	0	17.91	0	0	0	0.71	
Employee	20,457	12,925	28,754	59.03	33,142	26,253	43,312	74.55	
Pure SE	10,701	2195	24,074	4.03	19,996	8193	38,220	8.07	
Hybrid SE	17,847	8466	30,015	1.91	28,833	15,672	46,116	3.26	
Freelancer	3513	1642	8840	2.12	23,925	9342	43,091	0.32	
DGA	31,161	18,152	49,364	0.69	56,338	40,593	80,695	2.52	
Benefits	13,772	9557	16,906	12.35	16,570	13,338	22,803	9.13	
Pensioner	20,260	9272	31,605	1.49	32,393	24,600	44,071	1.11	
Other	0	0	10,909	0.47	0	0	4346	0.33	
Disposable h	ousehold in	come							
No income	17,815	14,083	23,348	18.71	11,302	1395	17,606	0.68	
Employee	23,457	18,525	29,715	58.61	22,944	18,084	28,998	75.06	
Pure SE	25,255	17,444	36,209	3.84	23,950	16,306	33,759	7.83	
Hybrid SE	25,336	18,061	34,730	1.78	25,066	17,977	33,551	3.23	
Freelancer	20,073	15,792	26,480	2.13	19,408	13,414	27,244	0.31	
DGA	35,564	25,788	48,590	0.67	31,219	23,236	42,497	2.36	
Benefits	14,358	11,468	20,234	12.25	14,275	11,281	19,606	9.02	
Pensioner	23,127	17,958	30,108	1.50	21,859	17,001	28,517	1.13	
Other	12,684	5443	18,966	0.51	8643	3579	16,922	0.38	
Wealth (in the	ousands)								
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	

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 reports household wealth adjusted for household size. Wealth includes both financial wealth and net value of real estate. 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 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 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 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 selfemployed, as expected. Still, compared to employees, a larger share of hybrid selfemployed 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.

	Women			Men			
	2nd Pillar	3rd Pillar	N	2nd Pillar	3rd Pillar	Ν	
Employee	67.74	6.51	332,831	78.28	16.81	409,571	
Pure SE	0.00	12.68	22,808	0.00	26.12	44,584	
Hybrid SE	51.08	9.57	10,756	57.04	20.88	17,971	
Freelancer	2.97	1.41	11,939	4.90	15.27	1775	
DGA	3.88	10.01	3895	2.58	25.32	13,845	
Benefit recipient	5.62	1.58	69,764	7.93	3.87	50,196	
Pensioner	12.83	3.94	8433	20.73	9.20	6096	
Other	0.00	3.19	2791	0.00	3.94	1955	

Table 4 Share of individuals (in %) contributing to pension pillars

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

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.

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 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.

3.1 Visualisation of Employment Trajectories

An employment trajectory is an individual's sequences of labour market states over time. Figure 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 Figs. 1 and 2. In addition, the share of benefit recipients in this subsample is smaller than in the overall population shown in Fig. 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 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 4 essentially does that and shows a so-called index plot¹⁸ of the trajectories of the same individuals as in Fig. 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

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.



Fig. 3 Distribution of labour market states over time of cohort born in 1961–1965 (self-employed)

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

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

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 selfemployment are not always accounted for.



Fig. 4 Index plot of cohort born in 1961–1965 (self-employed)

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

Table 5 Hypothetical labour market trajectories		Yea	r						
		1	2	3	4	5	6	7	8
	Trajectory 1	Е	Е	Е	S	S	s	S	Е
	Trajectory 2	Е	Е	Ν	S	S	S	S	Е
	Trajectory 3	S	Ν	Е	Ν	S	S	S	Е
	Trajectory 4	S	S	S	S	Е	Е	Е	Е
	Trajectory 4	S	S	S	S	Е	Е	Е	

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 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 al 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 4 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.

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.

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

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 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 1 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

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

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 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.

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 6–8 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 5 presents the same trajectories as shown in Fig. 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 short self-employment spells.

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

- 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 selfemployment 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.

²⁵ The mapping for each cohort from Ward's twelve groups solution to our seven clusters is given in Online "Appendix B".



Graphs by Ward's 12 cluster solution

Fig. 5 Index plot by clusters of cohort born in 1961–1965 (self-employed)

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

4 Differences Among Self-employment Clusters

In this section we compare demographic characteristics, income and wealth of individuals across the clusters identified in Sect. 3. Table 6 shows the demographic characteristics across the twelve clusters involving (1-7) and not involving (8-12) selfemployment. The patterns are largely similar to those for the labour market states shown in Table 2. 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 2. 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 6 than in Table 2. This can be explained by the distribution of the length of self-employment spells.²⁷ Men generally spend more time in selfemployment than women and hence make up a larger share of the self-employed observations in Table 2. 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.

 $^{^{26}}$ 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.

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

Cluster	Age	in %								
		Men	Bread-winner	Single ^a	Dutch	Western	Non-western	Sample share		
With self	employ	nent								
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-emp	loyment								
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			

 Table 6
 Demographic characteristics across clusters in sample

The statistics are calculated over all 1,228,203 individual-year observations for each cluster

^a Includes singles, divorced and widowed individuals

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.

4.1 Self-employment over time

Figure 6 shows the same overall self-employment shares as Fig. 2 and adds the contributions that different clusters make to the self-employment rate. Only about half of the share is 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 selfemployment (clusters 1–3). These findings are similar if we look at the cohorts separately (details available upon request).

Table 1 in "Appendix A" shows that in most clusters, around three out of four individuals have only one self-employment spell, while roughly one in five have



Fig. 6 Self-employment share and contributions of clusters

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 2-3 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

4.2 Income

The upper and middle sections of Table 7 show individual taxable income and disposable household income (adjusted for household size) by gender and cluster. Comparing the results to Table 3 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.

²⁸ See Table 2 in "Appendix A" for spell lengths across cohorts.

	Women				Men			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Taxable individual income	e							
Weak LM attach. (SE)	1141	0	8584	7.39	16,631	939	31,880	0.51
Benefit recipient (SE)	12,483	5077	17,226	1.44	16,729	10,824	26,467	1.07
Short SE spells	15,677	6320	25,832	9.13	32,622	23,061	47,380	8.49
Employee to SE	19,032	7847	30,509	2.84	29,333	19,381	42,109	4.31
Pure SE	11,219	1890	23,401	3.66	20,731	9249	34,621	7.64
Hybrid SE	16,450	6338	26,802	1.26	28,899	16,177	45,410	1.72
DGA	26,444	12,857	44,002	1.16	50,669	33,345	74,614	4.28
Weak LM attach.	0	0	0	13.07	759	0	18,402	0.34
Benefit recipient	14,045	9142	18,232	9.91	18,376	13,729	25,069	6.17
Long benefit rec.	14,945	7575	22,412	10.97	27,625	20,199	35,272	7.68
Employee	21,582	14,253	29,846	38.58	33,405	26,522	43,299	57.31
Pensioner	21,092	10,535	31,060	0.59	33,879	26,184	41,556	0.47
Disposable household inc	ome							
Weak LM attach. (SE)	19,033	14,737	25,157	7.45	18,004	11,953	24,937	0.50
Benefit recipient (SE)	16,560	12,295	22,428	1.45	15,488	11,485	22,242	1.08
Short SE spells	22,496	16,947	29,572	9.08	23,251	17,446	30,770	8.42
Employee to SE	24,604	17,667	33,326	2.78	22,791	16,929	30,515	4.25
Pure SE	23,559	16,195	33,522	3.67	23,074	15,848	31,791	7.65
Hybrid SE	24,132	17,468	32,340	1.25	24,325	17,298	32,523	1.72
DGA	33,470	23,234	46,698	1.17	30,053	22,136	41,097	4.28
Weak LM attach.	19,114	14,930	24,874	13.58	15,221	10,733	21,973	0.34
Benefit recipient	15,943	12,112	22,024	9.93	15,037	11,676	20,211	6.23
Long benefit rec.	20,039	15,035	26,552	11.00	19,714	15,130	25,199	7.85
Employee	23,500	18,599	29,695	37.99	23,009	18,199	28,928	57.16
Pensioner	23,193	18,389	30,735	0.64	21,420	17,354	26,831	0.52
Household wealth (in tho	usands)							
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

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 adjusted for household size with the CBS equivalence scale. The statistics are calculated over all individual-year observations for each cluster for all individuals in the sample aged 24–60. Taxable income is available for all years. Disposable household income is only available 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

0.12

89

270

7

158

0.09

Pensioner

147

20

$$y_{it} = \beta_0 + X'_{it}\beta_x + L'_{it}\beta_l + W'_i\beta_c + C'_i\gamma + T'\delta + \epsilon_{it}$$
(1)

Here y_{it} is the income variable of interest.²⁹ X_{it} is a vector with characteristics of individual *i* 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 Finally, we include dummy variables indicating the household's main income source in each period.

 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 Eq. (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 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 9.

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.

³⁰ We take singles as the base category and have separate dummy variables 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.

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.

	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***				
	(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		1		\checkmark				
Observations	576,376	576,376	563,348	563,348				
Individuals	25,965	25,965	25,441	25,441				
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 charac-teristics, non-self-employed labour market states, and other main income sources

	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)				
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		\checkmark		1				
Observations	577,212	577,212	564,003	564,003				
Individuals	25,965	25,965	25,441	25,441				
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				

 Table 9
 Panel regression—disposable household income

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 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 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.

4.3 Wealth

The lower panel of Table 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-employed 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 Eq. (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 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.

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.

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

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 nonhousing wealth than for housing wealth (details available upon request).

	Household wealth (i	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)
Non-SE clusters		1		1
Observations	168,055	168,055	162,981	162,981
Individuals	21,186	21,186	20,621	20,621
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

 Table 10
 Panel regression—household wealth

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

	Women			Men			
	2nd Pillar	3rd Pillar	N	2nd Pillar	3rd Pillar	N	
Weak LM attach. (SE)	13.14	0.95	41,698	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	51,543	67.09	17.53	46,710	
Employee to SE	44.91	8.69	16,007	52.41	18.32	23,745	
Pure SE	8.59	11.24	20,739	10.55	22.85	42,192	
Hybrid SE	39.13	6.47	7136	48.82	20.23	9496	
DGA	21.89	9.94	6561	25.46	25.02	23,525	
Low LM attach.	5.13	0.25	73,707	12.25	3.01	1861	
Benefit recipient	18.18	1.76	55,894	20.86	3.63	33,943	
Long benefit rec.	42.14	3.91	61,851	53.57	10.97	42,238	
Employee	69.32	7.10	217,481	77.96	17.07	314,903	
Pensioner	9.88	5.60	3340	17.77	9.52	2605	

Table 11 Percentage contributing to second and third pension pillar

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

Table 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 4. The cluster of DGAs contributes more often than what Table 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 4), but they do not pay occupational pension premiums when they are DGA (Table 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 selfemployment 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 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.

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

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 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 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 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 accrual. We add year effects based on the last year we observe an individual in the pension data to control for level differences

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

	Women	Women			Men			
	Median	Q_1	Q_3	(%)	Median	Q_1	Q_3	(%)
Pension contribution year	s							
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 thouse	ands)							
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 the	ousands)							
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

Table 12 Occupational pension entitlements

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

	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	14,046	14,845	14,047	14,847	
R^2 adjusted	0.206	0.226	0.233	0.262	
RMSE	4.213	8.096	34.693	57.802	

 Table 13 Regression—accruals in 2nd pillar

Standard errors in parentheses; *** p < 0.001, **p < 0.01, *p < 0.05Regression including cohort fixed effects, and controlling for individual's demographic characteristics, for when we last observe an individual

due to different timing in observations. Because of the limitations of the data, we do not include the labour market state dummies.

Both regressions show the same picture and the cluster coefficients confirm what we have learned from Table 12. Self-employed who mostly receive benefits

	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***	
	(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		1		1	
Observations	467,380	467,380	548,536	548,536	
Individuals	24,017	24,017	25,411	25,411	
log likelihood	- 87,366.566	- 86,542.144	- 208,083.694	- 207,264.314	

Table 14 Probit regression—participation in 3rd pillar

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

	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)		
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***		
1	(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	. /	1	. /	1		
Var(e.c p3pen defl)	102.08***	101.48***	94.56***	- 93.77***		
· · · - r · r · · - · · · · · · · · · · · · · · · ·	(12.240)	(12,172)	(6.949)	(6.919)		
Observations	577.333	577,333	564.127	564,127		
Individuals	25.965	25.965	25.441	25.441		
Log likelihood	- 149,235	- 148,545	- 426,205	- 425,207		

 Table 15
 Tobit regression—contributions to 3rd pillar

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 are worse off than their non-self-employed counterparts. Among the other selfemployed 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 the third pillar to counterbalance 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 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 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.

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 selfemployment 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 (nonpension) 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 selfemployed 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.

Acknowledgements This research was funded by Instituut Gak through Netspar. We would like to thank Mauro Mastrogiacomo, Theo Nijman and the participants of the Structural Econometrics Group at Tilburg University for valuable comments and discussions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- 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(4), 129–147.
- Abbott, A. (1995). Sequence analysis: New methods for old ideas. Annual Review of Sociology, 21(1), 93–113.
- Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *The Journal of Interdisciplinary History*, 16(3), 471–494.
- Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology*, 96(1), 144–185.
- Aldenderfer, M., & Blashfield, R. (1984). Cluster analysis. In: Number 44 in quantitative applications in the social sciences. SAGE Publications Inc.
- Blanchflower, D. G. (2000). Self-employment in OECD countries. Labour Economics, 7(5), 471-505.
- Bolhaar, J., Brouwers, A., & Scheer, B. (2016). De flexibele schil van de Nederlandse arbeidsmarkt: een analyse op basis van microdata. Technical report, CPB.
- Bosch, N., & Van Vuuren, D. (2010). De heterogeniteit van zzp'ers. Economisch Statistische Berichten, 95(4597), 682–684.
- Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal*, 6(4), 435–460.
- de Bresser, J., & Knoef, M. (2015). Can the Dutch meet their own retirement expenditure goals? *Labour Economics*, 34, 100–117.
- Halpin, B. (2017). Sadi: Sequence analysis tools for stata. The Stata Journal, 17(3), 546–572.
- Hollister, M. (2009). Is optimal matching suboptimal? Sociological Methods and Research, 38(2), 235–264.
- Humphries, J. E. (2016). The causes and consequences of self-employment over the life cycle. New York: Mimeo.
- Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., & Kalwij, A. (2016). Measuring retirement savings adequacy: Developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance*, 15(1), 55–89.
- Levine, R., & Rubinstein, Y. (2017). Smart and illicit: Who becomes an entrepreneur and do they earn more? *The Quarterly Journal of Economics*, 132(2), 963–1018.
- Madero-Cabib, I., & Fasang, A. E. (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. In: Netspar Brief 7. Netspar.

- Mastrogiacomo, M., Li, Y., & Dillingh, R. (2016). Entrepreneurs without wealth? An overview of their portfolio using different data sources for the Netherlands. In: *Netspar design paper 37*. Netspar.
- Munnell, A., Sanzenbacher, G., & Walters, A. (2019). How do older workers use nontraditional jobs? Forthcoming from Journal of Pension Economics and Finance. Cambridge: Cambridge University Press.
- Parker, S. C. (2004). The economics of self-employment and entrepreneurship. Cambridge: Cambridge University Press.
- Scherer, S. (2001). Early career patterns: A comparison of Great Britain and West Germany. European Sociological Review, 17(2), 119–144.
- SER. (2010, October). 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. In: *LIVES working papers 24*. University of Geneva Institute for Demographic and Life Course Studies.
- Studer, M., & Ritschard, G. (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(2), 481–511.
- SZW. (2016, July). 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. Retrieved July 30, 2019 from http://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?locat ions=NL.
- Tophoven, S., & Tisch, A. (2016). Employment trajectories of German baby boomers and their effect on statutory pension entitlements. *Advances in Life Course Research*, *30*, 90–110.
- 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(301), 236–244.
- Zacher, H., Biemann, T., Gielnik, M. M., & Frese, M. (2012). Patterns of entrepreneurial career development: An optimal matching analysis approach. *International Journal of Developmental Sci*ence, 6(3–4), 177–187.
- Zwinkels, W., Knoef, M., Been, J., Caminada, C., & Goudswaard, K. (2017). Zicht op zzp-pensioen. In: Netspar Design Paper 91. Netspar.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.