E. Beusch, A. van Soest[1]

A dynamic multinomial model of self-employment in the Netherlands

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

Keywords: labour market transitions; big-five; mixed logit; state dependence.

JEL classification: J62; C23; C25.

1. Introduction

In recent years the number of self-employed in the Netherlands has grown substantially, leading to an increase of almost 30% in their share in the working population: from 12.8% in 2003 to 16.6% in 2017[2]. The main driver behind this growth have been the so called solo self-employed (SSE; in Dutch "zzp'ers" = zelfstandigen zonder personeel). As can be seen in Fig. 1, the share of SSE in the working population increased from 8.1% in 2003 to 12.3% in 2015 and has remained rather stable since then. The share of other self-employed has, on the other hand, seen a slight decline since the financial crisis.

Because of the growing numbers, Dutch policy makers have become more interested in the effects the self-employed or SSE in particular may have on the labour market, social security, and government finances. Accordingly, several recent policy papers describe the trend in self-employment and the characteristics of the self-employed, or analyze their performance; see, e.g., (Bosch et al., 2012; Bosch, 2014) or (Bolhaar et al., 2016). One key concern of the Dutch policy makers in relation to social security is adequacy of the pension savings of the self-employed; see, e.g., (Knoef et al., 2016) or (Mastrogiacomo, Alessie, 2015). While the pay-as-you-go pension (the so called AOW, the first pillar of the Dutch pension-system) covers all individuals who have lived in the Netherlands between ages 15 and 65, contributions to a fully

---

[1] **Beusch, Elisabeth** — Tilburg University, Netherlands; e.beusch@uvt.nl.
    Van Soest, Arthur — Netspar, Tilburg University, Netherlands; avas@uvt.nl.

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

Fig. 1. Shares of solo self-employed (SSE) and other self-employed in the working population
(*Data source*: CBS Arbeitsdeelname; kerncijfers)

funded pension plan (the second pillar) are, unlike for the large majority of employees, neither mandatory nor accessible for most of the self-employed[3]. Instead, the self-employed are expected to save themselves through (voluntary) savings (the third pillar). Such pension savings are tax-favoured for everyone with an "incomplete" second-pillar pension, in order to stimulate that individuals indeed save enough for their pension[4]. This then raises the question whether the self-employed save enough in the third pillar. It turns out that the policy makers' concerns have some basis. Mastrogiacomo (2016) shows that while the self-employed have the same savings ambitions as the employed, they are not more likely to save in the third pillar. Only one third of the self-employed contributes to the third pillar, which indicates that the majority will fall short on their savings. In line with this finding, earlier studies by De Bresser, Knoef (2015) and Knoef et al. (2016) found that the self-employed are less likely to meet their retirement expenditure or saving goals. Zwinkels et al. (2017) focus on the solo self-employed and estimate that more than 40% of SSE households fall short on their savings if a target replacement rate of 70% of earnings is used.

One simplifying assumption made in the pension wealth projections by Zwinkels et al. (2017), but also by De Bresser, Knoef (2015) and Knoef et al. (2016), is that the observed individuals remain in the labour state in which they were at the point in time when the data were collected. To our knowledge this assumption ("static micro-simulation") is standard in the pension

---

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

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

E. Beusch, A. van Soest

literature and its consequences have not been discussed so far. Still, given that the savings in the second (and third) pillar — a large share of most individuals' pension wealth — are linked directly to the individuals' labour state, it may be worthwhile to study the validity and consequences of this assumption. This paper therefore studies the dynamics in the Dutch labour market, considering self-employment as one of the labour market states. For instance, it asks how likely it is that somebody who is observed in self-employment will remain self-employed, depending on the individual's characteristics.

We will use data from the LISS (Longitudinal Internet Studies for the Social sciences) panel, a representative sample of adult individuals in the Netherlands administered by CentERdata (affiliated with Tilburg University). It is based upon a random sample of Dutch households drawn by Statistics Netherlands. Individuals of age 16 and older in the participating households are invited to answer survey questions on a monthly basis. The surveys cover domains such as work, education and income, but also a wide range of other topics, like health and personality, thus offering a rich set of information on which we can build our analysis. It also allows to distinguish between employees, SSE and other self-employed. Because sample size limitations, the main analysis is done without a distinction between different self-employment types, even though such a distinction might be desirable given the specific interest in the SSE in the Dutch policy debate[5].

In addition to a set of personal and household characteristics also included in most of the studies cited above, we control for personality traits and a (lagged) health index. Recent work on the economic importance of personality traits (see e.g., (Borghans et al., 2008)) has shown that personality traits matter for different labour market outcomes and, particularly, for the decision to become self-employed (Beugelsdijk, Noorderhaven, 2005). This has also been found in sociological research on career counseling. For instance, Obschonka et al. (2013) construct an Entrepreneurship-Prone Big-Five Profile (EP) Distance measure and find that in the US, the EP distance's geographical distribution corresponds to observed entrepreneurial activity. We therefore also include the EP distance in our analysis. Good health has been identified as a factor that increases the probability to become self-employed for older workers in the US (Rietveld et al., 2015). We use the rich nature of the LISS data to construct a health index and study the role of health for self-employment transitions in the Netherlands.

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

---

[5] As illustrated in Fig. 1 most of the recent dynamics in self-employment seem to be driven by the SSE.

all individuals of working age and use survey data, which has the advantage of providing rich background information such as personality or health indicators, as already emphasized above.

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

The paper continues as follows. Section 2 discusses the LISS panel and our sample selection process. Our model is presented in section 3 and the corresponding estimation results in section 4. Section 5 presents the simulation results based on the estimations. Section 6 concludes the paper.

2. Data

In this paper we make use of the LISS (Longitudinal Internet Studies for the Social sciences) panel. The LISS panel consists of monthly Internet surveys to a representative sample of households drawn from the Dutch population register[6]. Among the monthly surveys there are ten annual or biennial longitudinal core studies. Additionally, individuals are asked to fill in a basic survey, the household box, about the most important general characteristics of their household and its members such as age, gender, education, marital status, as well as their primary occupation and gross income. Individuals, or the contact person of the household — if there is more than one member of the household participating in the LISS panel — are asked to fill in the household box at the beginning when joining the panel and then prompted every month before each survey to fill in changes if such have occurred.

2.1. Self-employment

Among all longitudinal surveys, there are three instances within the LISS panel through which we can identify self-employed individuals. First, information about an individual's labour market status is stored in the household box. Instead, self-employed individuals can be identified using either the *Work and Schooling* or the *Economic Situation: Income* core study. We will base our analysis on the income study for two main reasons. First, the income study allows us to identify solo-self-employed (SSE) while the work and schooling study does not allow for a distinction between SSE and other self-employed. Second, the income study based sample suffers less from selection or attrition bias than the work and schooling study. The income survey has a different timing from other LISS studies. Individuals are supposed to use documentation on income taxes in the previous year to fill out the survey questions. Therefore, the survey asks individuals in period t about sources of income in the calendar year $t-1$. For example, the 2008 survey asks about all income received in 2007. We classify individuals as employees who report receiving only income from employment over the whole year. Individuals with

---

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

income from both employment and self-employment are classified as self-employed, together with those individuals who report only income from self-employment[7]. An individual is classified as self-employed if indicating at least one type of entrepreneurial work activity. The activities that the income survey covers are (part-time) work as an entrepreneur or freelancer, SSE, owning a company (including a private limited liability company or a limited partnership), or participating in a partnership (either a so called *maatschap* or *vennootschap onder firma, VOF*) and, lastly, making a profit (or loss) through an enterprise in some way (except as spouse or partner cooperating in the business). Next, we classify all individuals as unemployed who report receiving unemployment benefits and no other source of income, ignoring other social benefits. Because this will only classify individuals who are unemployed for a whole calendar year as unemployed, the unemployment definition is rather strict, covering a smaller number of individuals than the work and schooling based definition which refers to one point in time. Finally, individuals with no income from any of these sources are classified as not in the labour force. A comparison of this classification with the work and schooling based classification is given in (Beusch, van Soest, 2020).

2.2. Sample selection

The surveys in the LISS panel generally have a response rate between 75 and 80%. Thus, every year we have answers of around 5000–6500 individuals, of which a few are incomplete. Individuals who leave the panel (i.e. stop answering the surveys all together) are replaced in later waves through refreshment samples. Most of these are stratified to improve the representativeness of the panel, aiming at oversampling difficult to reach groups with below-average response rates. For the surveys that we use, there are in total more than 14000 individuals across the 11 waves and some 81000 observations. We have information on the labour status for some 60000 observations if we use the income based classification. We restrict the sample to individuals from age 25 up to and including age 60, i.e. individuals' prime working years. We choose the lower bound at 25 because, first, the minimum wage increases with the worker's age until 23 in some sectors. As a result of this it seems that young workers in these sectors may have a higher risk of becoming unemployed close to their birthdays (Kabátek, 2020). Second, students who are finishing their education are harder to classify. They may hold a (side) job, while studying, and can also be considered first time job seekers. By age 25 most individuals should no longer be students. The age limit at 60 years stems from the idea that individuals older than 60 may have access to (early) retirement. The age restriction reduces the sample size to approximately 32000 individuals.

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

---

[7] About 40% of all self-employed have income from both employment and self-employment.
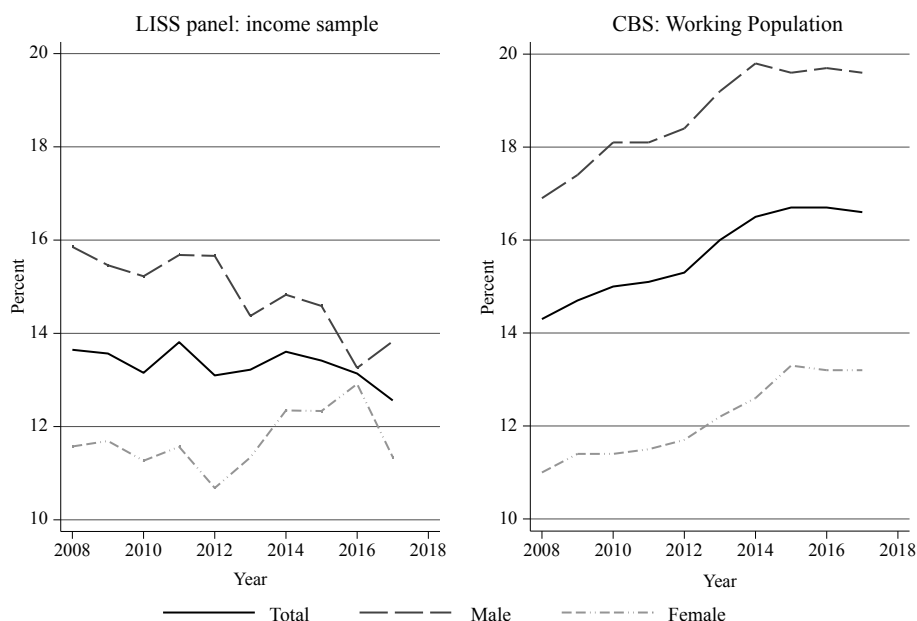
Fig. 2. Comparison of self-employment shares in regression sample and population
(*Data source*: CBS Arbeidsdeelname; kerncijfers)

in the years 2008–2012 and again from 2014–2018, we do not include the 2014–2018 block. These restrictions make us lose approximately 15% of the observations.

As shown in (Beusch, van Soest, 2020, Appendix A), the final restriction potentially creates (or worsens) attrition bias in the sample. Because a large share of the dropped observations belongs to individuals that participate in more than one wave, we correct for breaks in sequences with information from the work survey.

Figure 2 compares the self-employment share in the income sample with CBS population data already seen in Fig. 1. It shows that the LISS panel replicates neither the magnitude nor the time trend of the share of self-employment. This difference can be due to selection bias, i.e. self-employed individuals are less likely to respond when invited for the LISS panel, or attrition bias, i.e. self-employed individuals are more likely to stop responding in a later wave. Most likely, the problem is a mixture of both. Examining the evolution of self-employment shares over time by recruitment wave shows that they all display similar patterns. They start with relatively high shares at recruitment, followed by a drop, and ending with a flat or slightly downward trend. The fact that shares are initially closer to population figures suggests that the trend is more affected by attrition bias than by initial selection.

Both selection and attrition bias are of less concern if they are driven by observables like age, gender, or eduction. In such a case one can correct for the bias by weighing the observations accordingly. We therefore tried weighing observations using weights based upon the observable characteristics that enter our model[8]. Weighing only leads to an increase of one percentage point in self-employment shares overall and does not change the trend. Furthermore, Wooldridge (2007, p. 1293) cautions about using weighting in panels. We therefore decided

---

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

not to use these weights. To correct for selection on unobservables, we would have to impose a structure on the selection process. As this would require strong assumptions, and because the self-employment shares are initially not that different from population shares, we refrain from correcting for selection bias and focus on attrition bias only. Correcting for attrition bias requires weaker assumptions, especially if we can use an exclusion restriction. We test and correct for attrition bias in section 3.2.

2.3. Explanatory variables

Because personality traits are less commonly used, we discuss them in some detail. The *Personality* core study of the LISS panel focusses on respondents' "personality and characteristics". Its questions are based on established questionnaires from the field of psychology that each have a different focus. One of these questionnaires is the short 50 question set for the Big-Five factor markers by Goldberg (1992). Individuals answer these questions on a 1 to 5 scale. We code their answers according to the corresponding International Personality Item Pool key[9]. For each of the five factors we then sum up the points on an individual basis and standardise these values with the mean and standard deviation of the complete LISS sample for each year, allowing us to interpret coefficients of the factors in terms of changes relative to the standard deviation. The Entrepreneurship-Prone Big-Five Personality Profile Distance (EP distance) measure is calculated using the non-standardised factor values following (Obschonka et al., 2013).

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

Since health variables are more standard in the literature, we do not describe the health index' construction in detail (see (Beusch, van Soest, 2020, Appendix D)). In addition, we include the individual characteristics age, gender, and dummy variables for medium level education (VMBO, VWO, or MBO diploma) and higher education (university (WO) and applied science university (HBO) degrees), which have been shown to have some correlation with the choice to be self-employed[10]. In addition, we use household specific variables: dummy variables controlling for whether an individual lives with a partner and/or has children, as well as the size of the household. These variables also have been found to have explanatory power in regressions explaining the decision to be self-employed; see, e.g., the overview of research on entrepreneurship by Blanchflower (2000).

---

[9] https://ipip.ori.org/newBigFive5broadKey.htm.

[10] See (Bosch et al., 2012; Bosch, 2014; CBS, 2014; Bolhaar et al., 2016).

Table 1. Means and standard deviations (in parentheses) of covariates by labour market state

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

Source: LISS panel, own calculations.

Table 1 reports the means by labour market status for all covariates. Overall we see that women are slightly over-represented in the sample as they make up 55% of all observations. Approximately half of the individuals in the sample have at least one child and the majority lives with a partner. Unsurprisingly, we find that those not in the labour force are mostly women, and that the majority of them lives with a partner. The two education dummy variables account for 96% of the sample, implying that only 4% of the sample has the lowest education level. Furthermore, we can see that the distribution of the two dummy variables varies between the working and non-working population — the higher educated are much more likely to do paid work. The distribution of the health index is left skewed with the mode at 0.76, and we can see directly that individuals in the working population have a higher health status than those not working. There are also differences in Big-Five factor markers between the working and non-working individuals. Comparing across the labour status groups, we find that the means of all variables except for the health index and the fourth personality factor marker are statistically different between employees and the self-employed. This difference is in line with the literature: men are more likely to be self-employed than women, the self-employed are on average older (compared to the employed), and the self-employed more frequently have a higher level of education. They are also more likely to have children. The argument for the EP distance by Obschonka et al. (2013) predicts that entrepreneur prone individuals are more extraverted, less agreeable, and more open to new experiences. We find all of this reflected in the differences of the means of the three factor markers. However, the theory also argues that entrepreneurs should be more conscientious, and we find the opposite. Still, the EP distance measure has a lower average value for the self-employed than for employees, as theory would predict.

E. Beusch, A. van Soest

2.3.1. Observed labour market state dynamics

In Table 2 the observed transition probabilities of the labour market states by gender are shown[11]. We see that employees are more likely to remain in the same state than the self-employed, whose probabilities to exit to working as an employee are 8 to 10% per year. This already suggests that the assumption that self-employment is persistent has its limitations. The majority of the self-employed who do not continue as such switch to employment, and since the number of employees is much larger than the numbers in other labor market states, the largest contribution in numbers to entrants into self-employment are individuals making the transition from employment.

Table 2. Observed transition probabilities (in %) by gender

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

Notes. Based on 12014 and 14496 observation pairs for men and women, respectively.
Source: LISS Panel and CBS Arbeidsdeelname; kerncijfers, own calculations.

Comparing the transition matrices of men and women, we see that women are less likely to remain in self-employment than men are, and much more likely to remain out of labour force. Because of the substantial differences between men and women, we estimate all models separately by gender. Finally, we observe, in percent as well as in absolute numbers, very few changes from unemployment to self-employment and vice versa. This may be due to our categorisation approach in the income based definition of self-employment, or to sample selection. Finally, comparing with CBS population figures shows that self-employed are under represented, as already discussed[12].

3. Model

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

---

[11] The transition matrix is, of course, affected by the imputations for the gaps in the data. The changes are relatively small, however (see (Beusch, van Soest, 2020) for details).

[12] The CBS shares presented in Table 1 are calculated for the 25 to 60 year olds, taking the average for 2008–2017.

3.1. Dynamic multinomial model of labour states

We model the observed labour market state of an individual as the outcome of a utility maximisation process[13]. Each individual re-evaluates the potential states every period, and chooses the labour state j that maximises utility for that period. In terms of the econometric specification we thus consider a discrete choice model where an individual i derives utility y_{ijt}^* from state j at time t. In other words:

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

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

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

$$
y_{ijt}^* = X_{it}\beta_j + y_{i,t-1}\gamma_j + \alpha_{ij} + \varepsilon_{ijt}.
\tag{2}
$$

Here $y_{i,t-1}$ is a vector of dummy variables describing the individual's labour state in the previous period, X_{it} is a vector of k observed strictly exogenous variables, and the coefficient vectors γ_j and β_j, $j = 1,2,3$, are to be estimated; γ_0 and β_0 are normalized to zero. The variables in X_{it} are individual as well as household characteristics that may influence the utility y_{ijt}^*. These variables have been discussed in section 2.3. X_{it} also includes time dummies, to control for macro-economic effects.

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

$$
\alpha_{ij} = y_{i0}\delta_j + \mu_{ij}.
\tag{3}
$$

Here $\mu_i = (\mu_{i0},\dots,\mu_{i3})$ is independent of y_{i0} and all X_{it} and ε_{ijt}. Like Gong et al. (2004) and Been and Knoef (2017), we assume that μ_i is drawn from a J-dimensional multivariate normal distribution with mean zero and covariance matrix W. W is parameterized using its Choleski decompostion as LL^{T}, where L is a lower triangular $J \times J$ matrix with parameters to be estimated. The δ_j are 4-dimensional parameter vectors to be estimated. This essentially boils down to including the vector of dummies y_{i0} as additional regressors when estimating the model[14]. Note

---

[13] More details on the model specification and its implementation can be found in (Beusch, van Soest, 2020, Appendix C).

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

that due to the presence of the unobserved heterogeneity terms, the independence of irrelevant alternatives (IIA) assumption is not imposed. This IIA assumption is often seen as a drawback of the standard multinomial logit model. The estimates of the covariances will give an indication whether individuals who prefer one labour state are also more likely to prefer any particular other labour state. For example, if the covariance for states 1 and 2 (self-employed or unemployed) is positive, we should expect an individual, ceteris paribus, to have a higher probability of choosing self-employment when he or she has a high individual parameter for unemployment.

Static multinomial model of labour states. As already mentioned, the static model is a special case of the dynamic model. That is, in the static model we exclude the past period's labour state and as a consequence also drop the initial conditions. Equations (2) and (3) can then be rewritten as

$$y_{ijt}^{*} = X_{it}\beta_j^s + \alpha_i^s + \varepsilon_{ijt}^s \quad \text{with} \quad \alpha_i^s = \mu_i^s, \tag{4}$$

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

Estimation. The probabilities implied by the static and dynamic model discussed above have to be simulated. Following (Bhat, 2001) and (Train, 2009, Section 9.3.3), we use Halton draws to simulate their multivariate normal distribution.

3.2. Attrition bias

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

In our benchmark specification of the model, we find that this variable enters the multinomial model significantly for men, providing evidence that the model for men suffers from attrition bias, but not for women (*p*-values are 0.001 for men and 0.41 for women). We therefore estimate an extension of the model adding a Heckman correction term (estimated in a first stage). Formally the attrition model extends equations (1) and (2) as follows:

$$A_{it} = \begin{cases} 1, & \text{if } A_{it}^* > 0 \quad \text{for } i = 1,\dots,N; \ t = 2,\dots,T, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

$$A_{it}^* = X_{it-1}\beta^H + y_{it-1}\gamma^H + z_{it-1}\theta + \nu_i + \psi_{1,it}, \quad i = 1,\dots,N; \ t = 2,\dots,T, \tag{6}$$

where A_{it}^* is latent and A_{it} indicates whether an individual is observed in the LISS panel at time t or not[15]. Together, equations (5) and (6) model the attrition process in the first stage of estimation. X_{it-1} contains the same vector of regressors as X_{it} in equation (2) but one period lagged. y_{it-1} is a vector of dummies indicating the labour state of the individual one period past. Lastly, z_{it-1} is a vector of variables entering the attrition equation but not the other equations in the model (the "exclusion restrictions"). For the exclusion restriction we follow (Cheng, Trivedi, 2015) and use the number of days that individuals took to answer the last wave of the *Economic Situation: Income* core study after they received the invitation. We further include a dummy that controls for whether the individual answered within the deadline of the first call to participate in the survey, or only after the reminder, as well as an interaction term of the two. The random term ν_i is assumed to be a time invariant random effect while $\psi_{1,it}$ is assumed to be iid standard normally distributed.

Stage one is estimated separately as a panel probit model with random effects. The second stage model is then given by equations (1) and (2), but with the inverse Mills ratio $\varphi(\cdot)/\Phi(\cdot)$, estimated using the panel probit model, as an additional regressor in equation (2)[16].

4. Estimation results

We estimate four different models[17]: the baseline model with the basic personal and household characteristics, and three other models in which we add the (lagged) health index, the EP distance measure, and the Big-Five factor markers, respectively. Because the first stage of the Heckman correction is the same for both the static and dynamic model, we will discuss these first stage results first. In our description of the estimation results and the effects we can derive, all statements are meant as partial effects, keeping all other observed and unobserved characteristics constant, and averaging over the complete sample. We will focus on the results for the separate regressions for men and women, since some of the the estimated coefficients differ substantially between men and women.

---

[15] We exclude individuals from this step if they leave our sample because they turned 61 years of age. The first stage only aims at correcting for an individual's own choice to participate in the survey or not; it does not correct for the sample selection choices we made (see section 2).

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

[17] Estimation results for models pooling men and women, detailed estimates of first stage equations modeling attrition, and alternative model specifications are available upon request.

E. Beusch, A. van Soest

4.1. First stage — Heckman correction

In the first stage regression the sample also includes all individuals, within the selected age range, for whom we have only one observation in the LISS panel[18]. For women, we find rather weak effects for the labour market state indicators for self-employment and not participating in the labour force on the chances to remain in the sample. The same effects are statistically more significant and of larger magnitude for men. We find the same signs for both genders though. That is, compared to employees, self-employed individuals are less likely to be observed in the following period, and individuals not in the labour force are more likely to be observed again. In particular for men, and in combination with the findings of the variable addition test, this supports our argument for the need to correct for attrition bias: attrition is at least part of the explanation why the self-employed are underrepresented in later waves of the panel.

Furthermore, we see that the first stage estimates for the basic personal and household characteristics are robust across the different model specifications. Only age has a statistically significant positive coefficient, implying that older individuals are more likely than younger individuals to continue participating in the LISS panel. For women, we do not find a statistically significant effect for any other of the basic covariates. This is in contrast to men, where the dummy variable for high education is statistically significant at the 1% level, and the medium education dummy variable is statistically significant at the 5% level. Together the coefficients on these dummies show that men are more likely to continue participating in the LISS panel the higher their education.

The variables that are excluded from the main model ("the exclusion restrictions") have the expected signs in the first stage. The more days individuals take to answer after the invitation to participate in the survey has been issued, the less likely they are to return in the following year. For women, this effect is particularly strong for those who answer after the first call for participation in a survey[19] (as shown by the significant interaction terms). In terms of the variables driving the different model specifications, we do not find statistically significant effects of the respondents' contemporary health status on the probability to be observed in the next period. This holds for both men and women. The EP distance measure is not statistically significant for women either, but it is significantly negative at the 5% level for men. Recall that the EP distance's interpretation is that the lower its value is, the more likely it should be that an individual is self-employed. The results show that conditional on employment or self-employment status, less entrepreneurial individuals are less likely to stay in the sample, perhaps because they are more pressed with time.

For both men and women, more conscientious individuals are statistically significantly more likely to continue participating in the LISS panel, as expected. For men, we do not find any other individually significant effects, but we do find that the Big-Five factor markers are jointly statistically significant (at a 2% level)[20]. For women, we find that emotional stability and openness

---

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

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

[20] The results for the corresponding restricted model are available on request.

for experience have marginally significant negative effects on the probability to stay in the sample, which is not what we would have expected[21].

4.2. Second stage — static models

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

When we look at the static estimation results, we find similar effects of personal and household characteristics in all specifications. See, e.g. the results for the model with Big-Five factor markers for women in Table 3 and for men in Table 4. For women, only few of the personal and household characteristics are individually significant in the self-employment equation. In contrast to this, almost all of them are highly statistically significant (i.e., most at the 1%-, and some at the 5% level) in the other two equations. In other words, personal characteristics do not help us much to explain the difference between being self-employed or working as an employee, but they are helpful in explaining the difference between employment and unemployment or not participating in the labour market. The only variables that have a significant coefficient in the self-employment equation are age and household size (both at the 1% level). We find that the self-employed are on average older, and that the larger the household, the more likely women are to choose self-employment over wage-employment.

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

For both genders, we find that the lagged health index does not enter the self-employment equation significantly. Health does have significant effects in the other equations though, and the coefficients are also jointly significant. The negative sign of the coefficients suggests that individuals who had bad health one period earlier are more likely to be observed in unemployment or out of the labour force relative to being employed[22].

Similarly, we do not find a statistically significant effect of the EP distance for women, although the sign is, as we would expect, positive in the other two equations. For men on the other hand, we find a significant effect (p-value < 0.01) also in the self-employment equation. The sign is negative which is in line with our expectations: as the EP distance decreases, the more likely individuals become to be self-employed. With respect to the Big-Five factor markers we find

---

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

[22] In the pooled regression the coefficient is also statistically significant (–0.25, with p-value < 0.01) in the self-employment equation. It is of smaller absolute magnitude compared to the other two equations (–0.88 and –0.89). In other words, bad health a period earlier will increase the relative risk for an individual in all three categories relative to being employed. But an individual is then more likely to be self-employed compared to the other two states.

Table 3. Static model with Big-Five factor markers, women

	Self-employed	Unemployed	Not in labour force
Constant	−17.84***	−5.96***	−4.65***
	(1.2323)	(1.0431)	(0.8464)
Age	0.13***	0.07***	0.1***
	(0.0169)	(0.0135)	(0.012)
Has partner	0.45	−0.53**	0.75***
	(0.305)	(0.2302)	(0.2087)
Has child	−0.64	−0.46*	−0.73***
	(0.3988)	(0.277)	(0.2374)
Middle education	−0.24	−1.28***	−1.51***
	(0.6524)	(0.4414)	(0.3927)
High education	0.39	−2.83***	−2.92***
	(0.6965)	(0.4805)	(0.4226)
Household size	0.47***	0.26**	0.34***
	(0.1511)	(0.1203)	(0.0958)
F1: extraversion	0.13	−0.09	−0.07
	(0.1555)	(0.1128)	(0.0955)
F2: agreeableness	−0.05	0.18	0.19*
	(0.1572)	(0.1118)	(0.0972)
F3: conscientiousness	0.09	−0.58***	−0.65***
	(0.1581)	(0.121)	(0.0998)
F4: emotional stability	−0.24*	−0.47***	−0.32***
	(0.1417)	(0.1059)	(0.0864)
F5: openness for experience	0.36**	0.36***	0.17*
	(0.145)	(0.1097)	(0.0987)
Inverse Mills ratio	10.49***	−6.63***	−8.06***
	(1.2557)	(1.3516)	(1.0045)
L (lower triangular matrix with parameters determining $W = LL^{\mathrm{T}}$)	6.99***	0	0
	(0.3628)		
	2.48***	3.46***	0
	(0.2235)	(0.2033)	
	3.23***	3.65***	1.55***
	(0.2143)	(0.1653)	(0.1167)
W (covariance matrix of μ_{ij} in equation (3))	48.8702	17.3339	22.5515
	17.3339	18.1315	20.6267
	22.5515	20.6267	26.1073

Notes. 14435 observations on 3267 individuals; log-likelihood = −7704.03. Regression including year fixed effects. (Non-robust) standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.10$.

E. Beusch, A. van Soest

Table 4. Static model with Big-Five factor markers, men

	Self-employed	Unemployed	Not in labour force
Constant	−22.07***	−10.28***	−2.5***
	(1.4174)	(1.5229)	(0.9608)
Age	0.2***	0.13***	0.07***
	(0.0181)	(0.019)	(0.0127)
Has partner	0.08	−0.77**	−0.76***
	(0.3333)	(0.3629)	(0.2579)
Has child	−1.65***	−0.38	−0.73**
	(0.348)	(0.4375)	(0.309)
Middle education	0.67	−1.46***	−2.45***
	(0.7667)	(0.5278)	(0.3789)
High education	1.6**	−2.6***	−3.83***
	(0.7987)	(0.5816)	(0.4332)
Household size	0.21	−0.03	0.16
	(0.1608)	(0.1919)	(0.1335)
F1: extraversion	0.39**	−0.07	−0.06
	(0.1608)	(0.1678)	(0.1028)
F2: agreeableness	−0.32**	0.13	0.05
	(0.1311)	(0.1309)	(0.0938)
F3: conscientiousness	0.02	−0.3**	−0.46***
	(0.1423)	(0.1407)	(0.1037)
F4: emotional stability	−0.17	−0.6***	−0.67***
	(0.1417)	(0.1422)	(0.0988)
F5: openness for experience	0.73***	0.34**	0.23**
	(0.1359)	(0.1464)	(0.1009)
Inverse Mills ratio	21.18***	−0.23	−7.86***
	(1.4209)	(2.2355)	(1.2027)
L (lower triangular matrix with parameters determining $W = LL^{\mathrm{T}}$)	6.37***	0	0
	(0.3519)		
	2.23***	3.15***	0
	(0.273)	(0.1969)	
	2.33***	2.44***	1.53***
	(0.2468)	(0.1824)	(0.1421)
W (covariance matrix of μ_{ij} in equation (3))	40.5225	14.2217	14.8488
	14.2217	14.9076	12.8843
	14.8488	12.8843	13.7103

Notes. 11967 observations on 2647 individuals; log-likelihood = −5244.5. Regression including year fixed effects. (Non-robust) standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.10$.

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

In all static models, the coefficient of the inverse Mills ratio is highly statistically significant for the self-employed and for the equation explaining "not in the labour force". It has a positive sign in the self-employment equation and a negative sign in the equation for not in the labour force. For women, the coefficient is also statistically significant and negative in the unemployment equation. This suggests that, keeping observed characteristics constant, attrition is correlated with the unobserved factors driving someone's labor market status, implying that it is important to correct for attrition bias. Looking at the estimates for the elements driving the unobserved heterogeneity components in the mixed logit model (the matrix L driving the covariance matrix), we see that all coefficients are highly significant. The variance of the unobserved heterogeneity in self-employment is much larger than for the other two states. In particular for men the covariances have about the same magnitude as the variance for unemployment or being out of the labour force. For women, the covariances differ more from each other and we see in particular that, for given observed characteristics, self-employed individuals are also more likely to be out of the labour force.

4.3. Second stage — dynamic models

The first observation that we can make unequivocally when comparing the dynamic with the static results, is that the dynamic model is preferable for all four models and samples. The likelihood ratio test rejects the null hypothesis that the dynamic factors play no role, i.e. the lagged labour states are jointly significant (*p*-value of 0.0000 for all) and in most cases also individually significant.

Second, we find that the Big-Five factor markers are jointly statistically significant. In models which replace the big five with the EP distance or the health index, we also find that the EP distance or the health index enter significantly (using LR tests, even at the 0.1% level)[23]. Hence, adding either personality traits or information on an individual's health improves the model compared to a model with only the core personal and household characteristics. Comparing the two models with personality traits, Akaike's information criterion suggests that we should choose the model with the Big-Five factor markers over the one with the EP distance for both men and women[24]. In the following, we will therefore focus our discussion on the model including the Big-Five. We also prefer the Big-Five model since the lagged health index does not enter

---

[23] The regression results of the basic model, both for the full sample, as well as with the adjusted sample sizes for the likelihood ratio tests against the basic model, are available upon request.

[24] Neither for men, nor for women the coefficient on the EP distance is significant in the self-employment equation in none of the three samples.

E. Beusch, A. van Soest

significantly in the self-employment equation of the dynamic model either. This model offers a more detailed explanation for which individuals become self-employed[25].

Third, when we test for the joint significance of the inverse Mills ratio we fail to reject the null hypothesis that the coefficients are jointly equal to zero at any conventional significance level[26]. The results for the regressions without the Heckman correction are shown in Table 5 for women and Table 6 for men. They indeed do not differ much from the results with the Heckman correction. This suggests that correcting for attrition bias is not essential once we estimate the dynamic model.

How do the results change from the static models' once we include dynamic effects? For women, age loses its statistical significance in the self-employment equation while it remains significant at the 5% level for men. In other words, once we condition on women's past labour state, we no longer find a significant effect of age on the choice between self-employed or employee. Both the coefficient on household size for women, and on the dummy for having at least one child for men are no longer significant in the self-employment equation either.

With respect to the factor markers, we find that only the fifth factor, openness for experience, remains statistically significant (at the 5% level) for women. The coefficient still has a positive sign. We find an increase in the probability for a woman to be self-employed over being an employee by a factor of 1.28 if her score on openness for experience increases by one standard deviation. For men, we also find that the coefficients on the factor markers decrease in magnitude, with the largest change occurring in the fifth factor, translating to a 40% smaller increase in the relative probability for self-employment. Its impact is now also of approximately the same size as for women, and the *p*-value increases to around 0.023. Furthermore, for men we find a change in the other factor markers too. Agreeableness is no longer significant but emotional stability and conscientiousness are significant at the 5% level. Their signs are negative and thus opposite to what we would expect based on the arguments underlying the EP distance. Regarding conscientiousness this is, however, not entirely surprising considering that we already saw in section 2.3 that the self-employed in the LISS panel are on average scoring lower than employees. Last but not least, the coefficient on extraversion remains statistically significant (also at the 5% level) and has, as expected, a positive sign. This implies that for men, an increase in the score for extraversion by one standard deviation increases the relative probability to be self-employed by a factor of 1.32 compared to being employed.

Looking at the lagged labour market state variables we find the following for both genders: Having been in the same labour state one period earlier increases the relative probability for an individual to be in the same labour state compared to being an employee — i.e. the diagonal in the block of coefficients for lags shows the largest values. This is what we would expect given the transition probabilities in Table 1. The coefficient on lagged self-employment in the self-employment equation stands out as the largest of all, implying stronger state dependence in self-employment than in other labor market states. We find that all coefficients for the lags are positive. That is, given that an individual is not an employee (past and initial), the individual is more likely to end up in any of the other three states rather than in employment. It should also be noted that in terms of the relative size of coefficients, the individual and household characteristics, as well as the personality traits have much less influence on the choice probabilities than

---

[25] Note also, that the two models do not differ much in the estimates for all the other coefficients. Hence, the statements we can make on either hold in general also for the other.

[26] The *p*-value is 0.188 for women and 0.395 for men.

Table 5. Dynamic model with Big-Five factor markers and no Heckman correction, women

	Self-employed	Unemployed	Not in labour force
Constant	−6.98***	−6.56***	−5.49***
	(0.8184)	(0.6488)	(0.4526)
Age	0.01	0.03***	0.05***
	(0.0096)	(0.0082)	(0.0063)
Has partner	0.29	−0.67***	0.35**
	(0.2415)	(0.1886)	(0.1692)
Has child	−0.34	−0.23	−0.43**
	(0.299)	(0.2331)	(0.192)
Middle education	−0.11	−0.8**	−0.9***
	(0.5784)	(0.3503)	(0.2933)
High education	0.18	−1.53***	−1.41***
	(0.6016)	(0.3807)	(0.3122)
Household size	0.13	0.04	0.05
	(0.1201)	(0.0977)	(0.0812)
F1: extraversion	0.03	−0.06	−0.04
	(0.0972)	(0.0895)	(0.0665)
F2: agreeableness	−0.07	0.03	0.01
	(0.1107)	(0.0907)	(0.0698)
F3: conscientiousness	−0.05	−0.14*	−0.14**
	(0.099)	(0.0839)	(0.0653)
F4: emotional stability	−0.1	−0.35***	−0.21***
	(0.0894)	(0.0827)	(0.0609)
F5: openness for experience	0.25**	0.23***	0.11
	(0.1004)	(0.0895)	(0.0713)
Last state: self-employed	4.00***	1.07**	1.13***
	(0.2327)	(0.4764)	(0.2906)
Last state: unemployed	0.13	2.96***	1.76***
	(0.5087)	(0.2809)	(0.2164)
Last state: not in LF	0.61**	1.56***	2.14***
	(0.2611)	(0.1973)	(0.1042)
Initial state: self-employed	4.92***	1.41**	2.3***
	(0.5324)	(0.5596)	(0.36)
Initial state: unemployed	3.47***	3.35***	3.18***
	(0.693)	(0.5308)	(0.4772)
Initial state: not in LF	2.84***	3.17***	4.27***
	(0.3794)	(0.2931)	(0.2381)
L (lower triangular matrix with parameters determining $W = LL^T$)	2.26***	0	0
	(0.2119)		
	1.12***	1.26***	0
	(0.2288)	(0.2282)	
	1.25***	1.43***	0.63***
	(0.1803)	(0.1641)	(0.1701)
W (covariance matrix of μ_{ij} in equation (3))	5.0893	2.5314	2.8172
	2.5314	2.8507	3.2082
	2.8172	3.2082	4.002

Notes. 14435 observations on 3267 individuals; log-likelihood = −6089.36. Regression including year fixed effects. (Non-robust) standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.10$.

E. Beusch, A. van Soest

Table 6. Dynamic model with Big-Five factor markers and no Heckman correction, men

	Self-employed	Unemployed	Not in labour force
Constant	–6.82***	–8.77***	–6.25***
	(0.7965)	(0.9056)	(0.5853)
Age	0.02**	0.08***	0.07***
	(0.0099)	(0.0117)	(0.0087)
Has partner	–0.08	–0.34	–0.32
	(0.2537)	(0.3139)	(0.2259)
Has child	–0.28	–0.07	–0.11
	(0.3129)	(0.4071)	(0.2749)
Middle education	–0.09	–0.62	–1.12***
	(0.5254)	(0.4082)	(0.3191)
High education	0.27	–1.38***	–1.99***
	(0.5303)	(0.4576)	(0.3549)
Household size	0.18	–0.1	–0.07
	(0.1278)	(0.1811)	(0.1269)
F1: extraversion	0.28**	–0.03	–0.03
	(0.1097)	(0.1277)	(0.0855)
F2: agreeableness	–0.14	0.02	–0.05
	(0.0927)	(0.1225)	(0.0823)
F3: conscientiousness	–0.19**	–0.12	–0.15*
	(0.0958)	(0.1119)	(0.0884)
F4: emotional stability	–0.25**	–0.33***	–0.36***
	(0.1026)	(0.116)	(0.0825)
F5: openness for experience	0.24**	0.15	0.08
	(0.1035)	(0.1187)	(0.085)
Last state: self-employed	4.17***	0.36	0.74*
	(0.2254)	(0.7462)	(0.3905)
Last state: unemployed	0.24	2.65***	1.26***
	(0.5737)	(0.3338)	(0.3108)
Last state: not in LF	0.57*	0.66**	1.59***
	(0.3281)	(0.2959)	(0.1751)
Initial state: self-employed	4.75***	2.7***	2.28***
	(0.5798)	(0.7191)	(0.4486)
Initial state: unemployed	2.60***	4.13***	3.76***
	(0.9068)	(0.5922)	(0.5508)
Initial state: not in LF	2.07***	4.04***	4.31***
	(0.5227)	(0.4126)	(0.3508)
L (lower triangular matrix with parameters determining $W = LL^{\mathrm{T}}$)	2.1***	0	0
	(0.2383)		
	1.48***	1.27***	0
	(0.3285)	(0.3264)	
	1.17***	1.62***	0.23
	(0.2578)	(0.2006)	(0.4156)
W (covariance matrix of μ_{ij} in equation (3))	4.4139	3.1113	2.454
	3.1113	3.7993	3.7779
	2.454	3.7779	4.0285

Notes. 11967 observations on 2647 individuals; log-likelihood = –4067.57. Regression including year fixed effects. (Non-robust) standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.10$.

E. Beusch, A. van Soest

the lagged values of an individual's labour state[27]. Finally, we also see that the estimated variance in the unobserved heterogeneity becomes substantially smaller once we include the dynamic variables. The variance in the unobserved heterogeneity for the self-employed is still larger than for the other two states but by a much smaller factor. The same holds for the estimated covariances. Looking at the covariances, we find that self-employed women are more likely to be out of the labour force than the unemployed, whereas the opposite holds for men.

5. Simulations

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

5.1. Transition probabilities

To correct for attrition, we simulate under the counterfactual assumption that none of the individuals leave the sample. For those who do in reality, we need to impute the values of the covariates. We assume that, apart from age, personal and household characteristics remain the same[28]. The missing Big-Five factor marker values are completed with the same mean values used to fill in the initial gaps as described in section 2.3. We then, for each individual i, draw one vector of unobserved heterogeneity components μ_i from the multivariate normal distribution (with mean zero and covariance matrix given by $\hat{L}\hat{L}$). In each time period for each individual and labour state, we then draw independent error terms ε_{ijt} from a Type 1 extreme value distribution. Taking the first labour market state that we observe for an individual as given, we then simulate individuals' labour market state outcomes for the following time periods.

Table 7. Simulated transition probabilities (in %), men

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

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

---

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

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

Table 8. Simulated transition probabilities (in %), women

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

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

Tables 7 and 8 show the simulated transition probabilities for men and women, based on the dynamic model with Big-Five factor markers and no Heckman correction[29]. The left panels show the results for all men and women and the right panels show the results for the 45 to 60 year old subsamples. Comparing the total shares of all labour market states with the shares in Table 1, we find that the simulations for the whole sample lead to results that are similar to what we observed initially in the data.

We also see that our simulations, in particular for women, overestimate the probability that individuals remain self-employed. As a consequence, the transitions out of self-employment into other labour market states are underestimated but the general pattern seen in the data is nevertheless reproduced. Overall, it looks like we fare slightly better for men than for women in terms of replicating the observed transition probabilities.

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

5.2. Individual simulations

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

We therefore simulate the employment paths from 2008 until 2017, for benchmark individuals who are self-employed in 2007. We fix age to 45 years at the start and set other personal and household characteristics equal to the median by gender for the self-employed in 2007. Thus

---

[29] We choose the model without Heckman correction since we could not reject the null hypothesis that the correction terms all have coefficient zero.
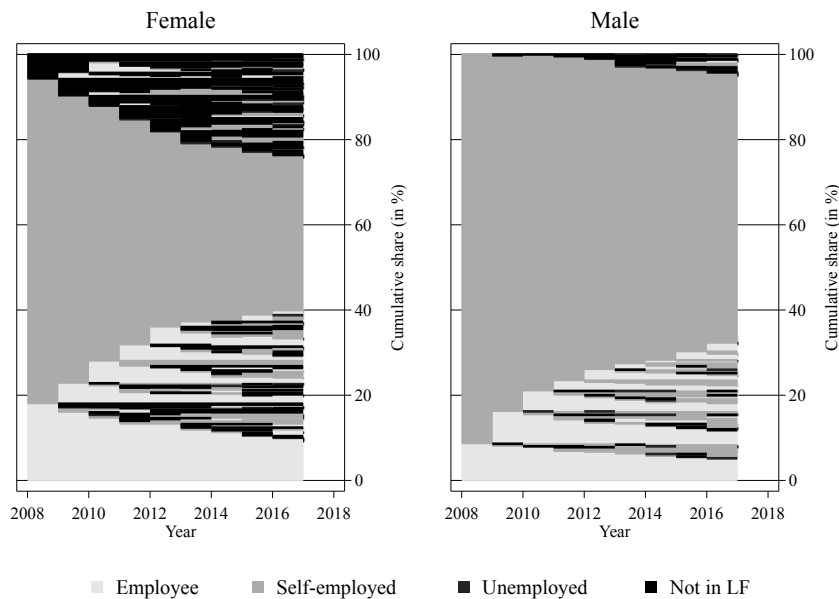
Fig. 3. Simulated employment paths by gender: self-employed in 2007 (median characteristics)

the benchmark individuals have a partner and one child. The self-employed male benchmark has high education, the female has medium education.

Next we take 500 draws of the benchmark individual's unobserved heterogeneity vector μ_i, and then proceed in the same way as with the forward simulations for the transition probabilities, drawing the ε_{ijt} from a Type 1 extreme value distribution once in each of the time periods for each labour market state. The resulting employment paths for the self-employed benchmark man and woman are shown as sequence index plots in Fig. 3. The sequence index plots show all simulated employment paths. The paths are stacked vertically on top of each other and each path is a (thin) horizontal line on which each year is coloured according to the labour market state the path takes in that year. We order the paths by labour market states, starting with the state in the first period, followed by the second period, etc. The scale of the y-axis reflects cumulative shares (in percent). On the righthand side of Fig. 3, we can see that the benchmark male has a chance of approximately 60% to remain self-employed throughout the ten years. For the benchmark female, this probability is only 35%. The main reason for the difference is the larger probability to leave the labour force for women.

Are these the probabilities we need to evaluate mobility? We view them as lower bounds on the probabilities of "persistent self-employment", because individuals can also leave self-employment for a short time and then return. If we are concerned about pension savings (our initial motivation), which tend to be much lower for the self-employed than for employees, the paths where self-employment is interrupted by a short employee spell are close to "persistent self-employment", since occupational pension wealth accumulated during the short employee spell will be small. A static micro-simulation which keeps everyone who is self-employed always in self-employment will then be a good approximation. We therefore also define an upper bound on the probability of "persistent self-employment", counting all paths where at least half of the time, five years, is spent in self-employment. This gives probabilities 80% for the male and 58.6% for the female benchmark individual starting as self-employed in 2007.
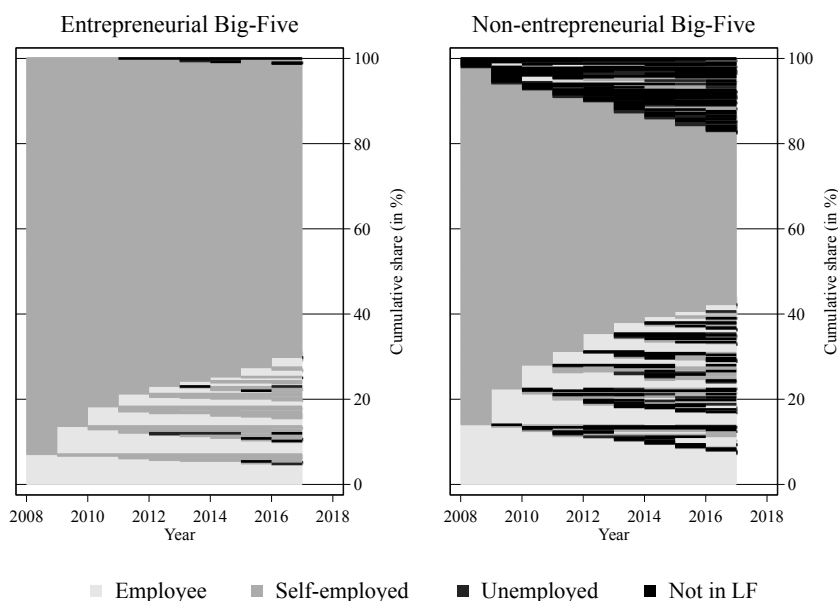
Fig. 4. Simulated employment paths: male, self-employed in 2007

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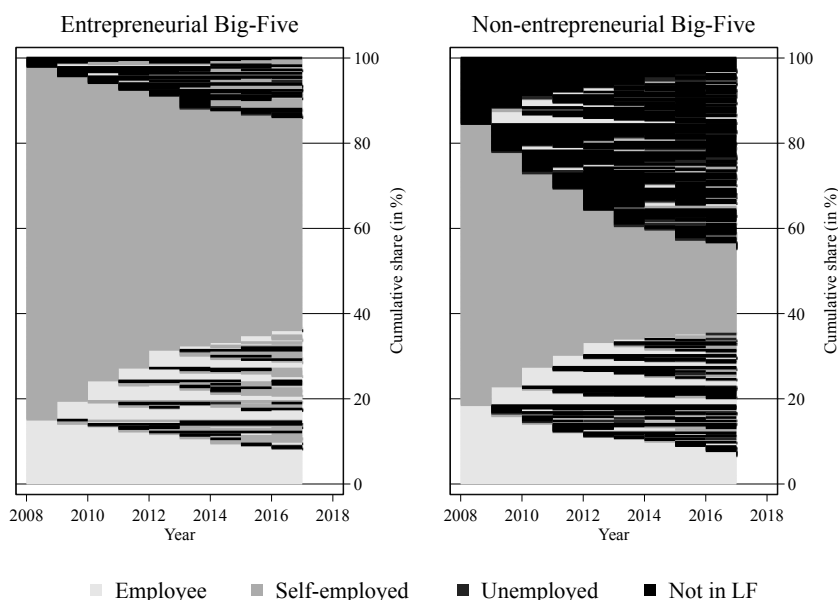

Fig. 5. Simulated employment paths: female, self-employed in 2007

E. Beusch, A. van Soest

Finally, since we are interested how personality affects individual probabilities to remain self-employed, we follow (Obschonka et al., 2013) and define an "entrepreneurial" profile with high levels of extraversion, conscientiousness, emotional stability, and openness, and a low level of agreeableness. Taking the standardization of the factor markers into account we assign high and low values 2.5 and –2.5. In addition, we define a "non-entrepreneurial" profile that takes the opposite values. Figures 4 and 5 shows the sequence index plots for the corresponding male and female benchmark individuals who are self-employed in 2007 (other characteristics are the same as before). As expected, the benchmark individuals with "entrepreneurial" Big-Five factor markers have a larger probability to remain self-employed than the "non-entrepreneurial" ones. The differences are large: the probabilities to stay self-employed throughout the ten year period is 67.6% (59.2%) for the entrepreneurial male (female) and 38.4% (22.2%) for the non-entrepreneurial male (female). The changes in the probabilities compared to Fig. 3 are larger for the "non-entrepreneurial" than for the "entrepreneurial" individuals, in line with the entrepreneurial characteristics of the self-employed benchmark individuals used in these earlier figures.

Table 9 summarises the simulation results for each combination of education level and personality profile. In the two rightmost columns, the table shows the lower and upper bounds based on our earlier definitions. The other columns show the share of employment paths in self-employment at a specific point in time. Note that the probability of self-employment is always lower for a female individual in comparison to a male benchmark individual, also in cases where both have exactly the same characteristics. Furthermore, a shift from high to medium education increases the difference in probabilities between the "entrepreneurial" and "non-entrepreneurial" case.

Table 9. Probabilities (in %) that different benchmark individuals who are self-employed in 2007, are self-employed in later years

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

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

In addition to this, the difference between the lower and upper bound (the final two columns) within each case also increases. Finally, we also find that the share of paths observed in self-employment in a year stabilizes after the fifth to sixth year in the simulation.

These results point at the limitations of a static microsimulation approach for pension wealth and income projections. If, for example, we take the most stable of the benchmark individuals, a male with high education and "entrepreneurial" Big-Five factor marker levels (left panel of Fig. 4), we still get chance of 12.6% to spend at least seven of the next ten years as an employee. Given the context of the Dutch pension system, such an individual would accumulate pension savings in the second pillar during at least seven of the ten years — and would therefore end up in a better financial position than what a model assuming that self-employment is an absorbing labour market state would predict. That is, a static micro-simulation would over-predict the share of self-employed with (too) low pension savings.

6. Conclusion

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

There is clear evidence of persistence (state dependence) in self-employment, but self-employment is not an absorbing state. The probability to remain self-employed in the next year is on average around 90% for men and slightly lower for women. But the probabilities to remain self-employed for a longer time period are much smaller, as illustrated by simulated employment paths over multiple years for benchmark individuals. Our benchmark self-employed male has an 80% chance to spend the majority of the next ten years in self-employment. This probability falls to 62% if the individual does not have an entrepreneurial personality, or even to 47% if, in addition, he has medium education rather than high education (the benchmark). For the self-employed benchmark woman, the chances to spend the majority of the next ten years in self-employment are only 59%, and this falls to below 40% if the woman does not have an entrepreneurial personality.

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

E. Beusch, A. van Soest

References

Albarrán P., Carrasco R., Carro J. M. (2019). Estimation of dynamic nonlinear random effects models with unbalanced panels. *Oxford Bulletin of Economics and Statistics*, 81 (6), 1424–1441.

Been J., Knoef M. (2017). Job-search requirements for unemployed at the end of working life effects on unemployment dynamics and self-employment probabilities. *Journal of Human Resources*, 52 (2), 491–530.

Beugelsdijk S., Noorderhaven N. (2005). Personality characteristics of self-employed; an empirical study. *Small Business Economics*, 24 (2), 159–167.

Beusch E., van Soest A. (2020). A dynamic multinomial model of self-employment in the Netherlands. *Netspar Discussion Paper* (forthcoming*)*. Netspar.

Bhat C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, 35, (7), 677–693.

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. *CPB Achtergronddocument*. CPB. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Achtergronddocument-17nov2016-De-flexibele-schil-van-de-nederlandse-arbeidsmarkt-een-analyse-op-basis-van-microdata.pdf.

Borghans L., Duckworth A. L., Heckman J. J., Ter Weel B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43 (4), 972–1059.

Bosch N. (2014). Succes als startende zelfstandige. *CPB Achtergronddocument*. CPB. https://www.cpb.nl/sites/default/files/publicaties/download/cpb-achtergronddocument-succes-als-startende-zelfstandigen.pdf.

Bosch N., Roelofs G.,van Vuuren D., Wilkens M. (2012). De huidige en toekomstige groei van het aandeel zzp'ers in de werkzame beroepsbevolking. *CPB Achtergronddocument*. CPB. https://www.cpb.nl/sites/default/files/publicaties/download/achtergronddocument-scenarios-zzpers.pdf.

Buddelmeyer H., Wooden M. (2011). Transitions out of casual employment: The Australian experience. *Industrial Relations: A Journal of Economy and Society*, 50 (1), 109–130.

CBS (2014). Achtergrondkenmerken en ontwikkelingen van zzp'ers in Nederland. Statistics Netherlands, The Hague. https://www.cbs.nl/-/media/imported/documents/2014/51/009093%20zzpers%20in%20nederland.pdf.

Cheng T. C., Trivedi P. K. (2015). Attrition bias in panel data: A sheep in wolf's clothing? A case study based on the Mabel survey. *Health Economics*, 24 (9), 1101–1117.

Cobb-Clark D. A., Schurer S. (2012). The stability of Big-Five personality traits. *Economics Letters*, 115 (1), 11–15.

De Bresser J., Knoef M. (2015). Can the Dutch meet their own retirement expenditure goals? *Labour Economics*, 34, 100–117.

Goldberg L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4 (1), 26–42.

Gong X., van Soest A., Villagomez E. (2004). Mobility in the urban labor market: A panel data analysis for Mexico. *Economic Development and Cultural Change*, 53 (1), 1–36.

Heckman J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5 (4), 475–492.

Heckman J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47 (1), 153–161.

Heckman J. J. (1981a). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: C. F. Manski and D. L. McFadden (eds.). *Structural Analysis of Discrete Data and Econometric Applications*, 179–195. MIT Press, Cambridge, MA.

Heckman J. J. (1981b). Statistical models for discrete panel data. In: C. F. Manski and D. L. McFadden (eds.). *Structural Analysis of Discrete Data and Econometric Applications*, 114–178. MIT Press, Cambridge, MA.

Kabátek J. (2020). Happy birthday, you're fired! The effects of an age-dependent minimum wage on youth employment flows in the Netherlands. *Industrial and Labor Relations Review*, forthcoming.

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.

Mastrogiacomo M. (2016). De pensioenpuzzel van zelfstandigen. *Netspar Brief 7*. Netspar. https://issuu.com/netspar/docs/d20161108_netsparbrief7_mastrogiaco.

Mastrogiacomo M., Alessie R. J. (2015). Where are the retirement savings of self-employed? An analysis of 'unconventional' retirement accounts. *Working Paper* 454. De Nederlandsche Bank.

Obschonka M., Schmitt-Rodermund E., Silbereisen R. K., Gosling S. D., Potter J. (2013). The regional distribution and correlates of an entrepreneurship-prone personality profile in the United States, Germany, and the United Kingdom: A socioecological perspective. *Journal of Personality and Social Psychology*, 105 (1), 104.

Oguzoglu U. (2016). Disability and multi-state labour force choices with state dependence. *Economic Record*, 92 (296), 28–46.

Prowse V. (2012). Modeling employment dynamics with state dependence and unobserved heterogeneity. *Journal of Business and Economic Statistics*, 30 (3), 411–431.

Rietveld C. A., van Kippersluis H., Thurik A. R. (2015). Self-employment and health: Barriers or benefits? *Health Economics*, 24 (10), 1302–1313.

Train K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Verbeek M., Nijman T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 681–703.

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

Wooldridge J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141 (2), 1281–1301.

Zucchelli E., Harris M. N., Zhao X. (2012). Ill-health and transitions to part-time work and self-employment among older workers. *Working Paper* 12/04. University of York. https://ssrn.com/abstract=2269449.

Zwinkels W., Knoef M., Been J., Caminada C., Goudswaard K. (2017). Zicht op zzp-pensioen. *Design Paper* 91. Netspar. https://www.netspar.nl/assets/uploads/P27112017_DP91_Knoef.pdf.