

# A Dynamic Multinomial Model of Self-employment in the Netherlands

Elisabeth Beusch, Arthur van Soest

**NETSPAR ACADEMIC SERIES**



DP 08/2020-023

# A Dynamic Multinomial Model of Self-employment in the Netherlands\*

Elisabeth Beusch<sup>†</sup>      Arthur van Soest<sup>‡</sup>

Version date: 22nd August 2020

## Abstract

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.

---

\*This paper uses data of the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata (Tilburg University, The Netherlands). Elisabeth Beusch's research was funded by Instituut Gak through Netspar. We are grateful to Theo Nijman for his guidance and constructive comments. We also thank Marike Knoef, Rob Alessie, Caspar van Ewijk, Mauro Mastrogiacomo, and the participants of the Structural Econometric Group at Tilburg University for valuable comments and discussions.

<sup>†</sup>Tilburg University, e-mail: e.beusch@uvt.nl

<sup>‡</sup>Netspar, Tilburg University

## 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.<sup>1</sup> The main driver behind this growth have been the so called solo self-employed (SSE; in Dutch “zzp’ers” = zelfstandigen zonder personeel). As can be seen in Figure 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.

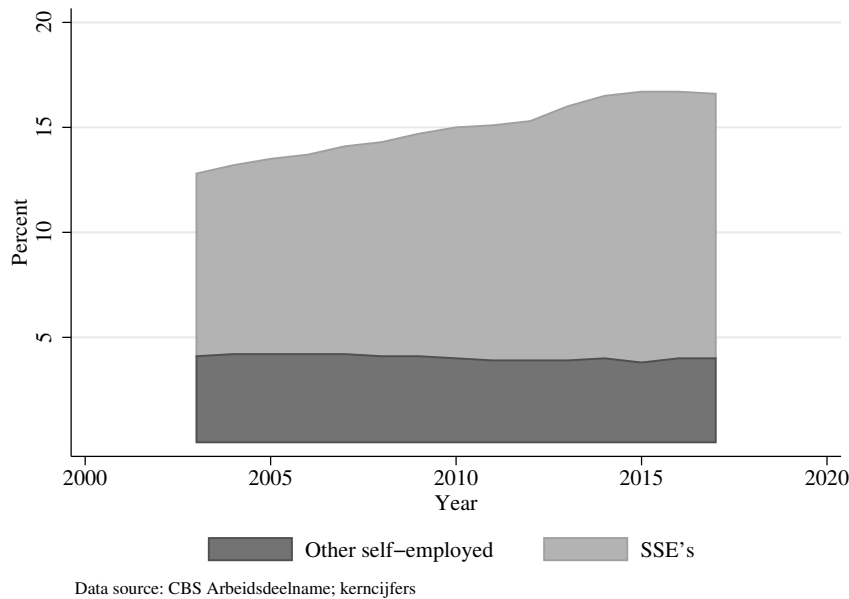


Figure 1: Cummulative share of self-employment in the working population

Because of the growing numbers, Dutch policy makers have become more interested in the effects the self-employed or SSE in particular may have on the labour market, social security, and government finances. Accordingly, several recent policy papers describe the trend in self-employment and the characteristics of the self-employed, or analyze their performance; see, e.g., Bosch et al. (2012), Bosch (2014) or Bolhaar et al. (2016). One key concern of the Dutch policy makers in relation to social security is adequacy of the pension savings of the self-employed; see, e.g., Mastrogiacomo and Alessie (2015) or Knoef et al. (2016). While the pay-as-you-go pension (the so called AOW, the first pillar of the Dutch pension-system) covers all individuals who have lived in the Netherlands between ages 15 and 65, contributions to a fully funded pension plan (the second pillar) are, unlike for the large majority of employees, neither mandatory nor accessible for most of the self-employed.<sup>2</sup> Instead, the self-employed

<sup>1</sup>All numbers are based on CBS Statline, *Arbeidsdeelname; kerncijfers*, downloaded on 4 May 2018.

<sup>2</sup>Most second pillar pensions are built up via employer based pension plans or industry specific pen-

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.<sup>3</sup> This then raises the question whether the self-employed save enough in the third pillar.

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

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

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

---

sion 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.

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

in the Dutch policy debate.<sup>4</sup>

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

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

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

The paper continues as follows. Section 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

---

<sup>4</sup>As illustrated in Figure 1 most of the recent dynamics in self-employment seem to be driven by the SSE.

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.<sup>5</sup> 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 income from both employment and self-employment are classified as self-employed, together with those individuals who report only income from self-employment.<sup>6</sup> 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

---

<sup>5</sup>Households that do not own a computer or Internet connection are provided with such so that they can nevertheless participate.

<sup>6</sup>About 40% of all self-employed have income from both employment and self-employment.

business). Next, we classify all individuals as unemployed who report receiving unemployment benefits and no other source of income, ignoring other social benefits. Because this will only classify individuals who are unemployed for a whole calendar year as unemployed, the unemployment definition is rather strict, covering a smaller number of individuals than the work and schooling based definition which refers to one point in time. Finally, individuals with no income from any of these sources are classified as not in the labour force. A comparison of this classification with the work and schooling based classification is given in Appendix A.

### *2.2. Sample selection*

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

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

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

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

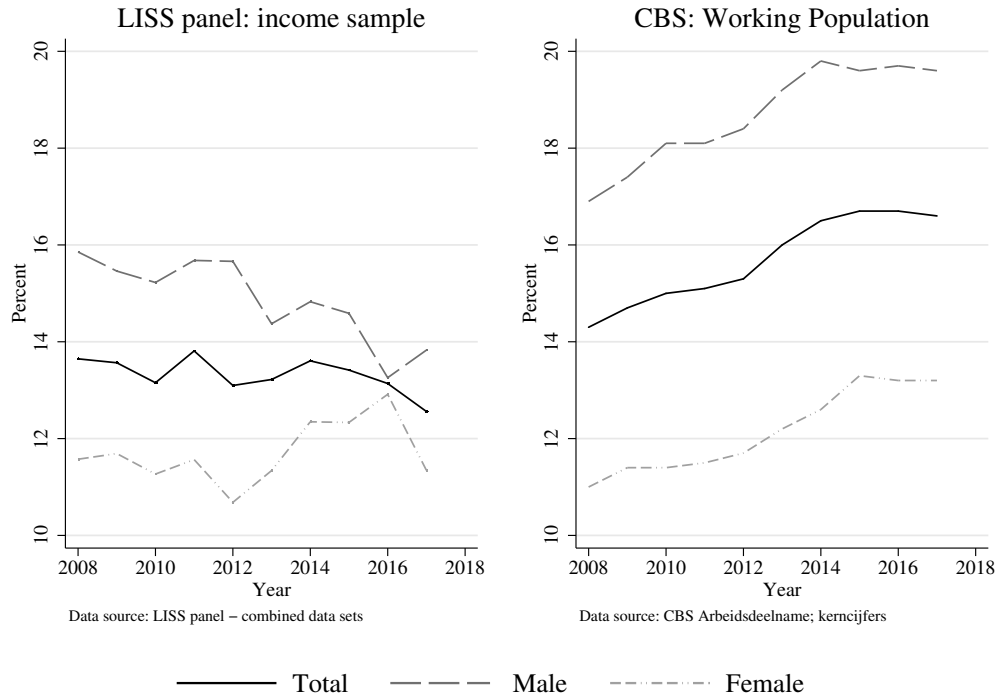


Figure 2: Comparison of self-employment shares in regression sample and population

Figure 2 compares the self-employment share in the income sample with CBS population data already seen in Figure 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 education. 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.<sup>7</sup> Weighing only leads to an increase of one percentage point in self-employment shares overall and does not change the trend. Furthermore, Wooldridge (2007, p. 1293) cautions about using weighting in panels. We therefore decided not to use these weights. To correct for selection on unobservables, we

<sup>7</sup>The weights are determined using population data downloaded from CBS Statline.

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.<sup>8</sup> For each of the five factors we then sum up the points on an individual basis and standardise these values with the mean and standard deviation of the complete LISS sample for each year, allowing us to interpret coefficients of the factors in terms of changes relative to the standard deviation. The Entrepreneurship-Prone Big-Five Personality Profile Distance (EP distance) measure is calculated using the non-standardised factor values following Obschonka et al. (2013).

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

Since health variables are more standard in the literature, we refer to Appendix D for the construction of the health index. In addition, we include the individual characteristics age, gender, and dummy variables for medium level education (VMBO, VWO, or MBO diploma) and higher education (university (WO) and applied science university (HBO) degrees), which have been shown to have some correlation with the choice to be self-employed.<sup>9</sup> 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).

---

<sup>8</sup><https://ipip.ori.org/newBigFive5broadKey.htm> retrieved on July 6, 2018.

<sup>9</sup>See Bosch et al. (2012), Bosch (2014), CBS (2014) and Bolhaar et al. (2016).

Table 1: Means and standard deviations (in brackets) of covariates

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

Source: LISS panel, own calculations.

## 2. Data

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

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

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

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

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.

### 2.3.1. Observed labour market state dynamics

In Table 2 the observed transition probabilities of the labour market states by gender are shown.<sup>10</sup> 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.

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.<sup>11</sup>

## 3. Model

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

### 3.1. Dynamic multinomial model of labour states

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

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

---

<sup>10</sup>The transition matrix is, of course, affected by the imputations for the gaps in the data. The changes are relatively small, however. See Table 14 in Appendix E.

<sup>11</sup>The CBS shares presented in Table 2 are calculated for the 25 to 60 year olds, taking the average for 2008–2017.

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

$$y_{ijt}^* = X_{it}\beta_j + y_{i,t-1}^\top\gamma_j + \alpha_{ij} + \epsilon_{ijt} \quad (2)$$

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  $\gamma_j$  and  $\beta_j$ ,  $j = 1, 2, 3$ , are to be estimated;  $\gamma_0$  and  $\beta_0$  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  $\epsilon_{ijt}$  are identically and independently distributed, independent of  $X_{it}$  and  $\alpha_{ij}$  and drawn from a Type 1 extreme value distribution. This implies that the labor market state probabilities, given  $X_{it}$ ,  $\alpha_{ij}$  and  $y_{i,t-1}$ , are the well-known multinomial logit probabilities (see Appendix C for details). 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  $\alpha_{ij}$ ,  $j = 0, \dots, 3$ , we model  $(\alpha_{i0}, \dots, \alpha_{i3})$  as follows, following Wooldridge (2005):

$$\alpha_{ij} = y_{i0}^\top\delta_j + \mu_{ij} \quad (3)$$

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

<sup>12</sup>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.

### 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

$$y_{ijt}^* = X_{it}\beta_j^s + \alpha_i^s + \epsilon_{ijt}^s \quad \text{with} \quad \alpha_i^s = \mu_i^s \quad (4)$$

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, chapter 9.3.3) we use Halton draws to simulate the multivariate normal distribution of  $\mu_i$ . See Appendix C.3 for details and reasoning.

#### 3.2. Attrition Bias

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

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

in a first stage). Formally the attrition model extends equations (1) and (2) as follows:

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

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

where  $A_{it}^*$  is latent and  $A_{it}$  indicates whether an individual is observed in the LISS panel at time  $t$  or not.<sup>13</sup> 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 and Trivedi (2015) and use the number of days that individuals took to answer the last wave of the *Economic Situation: Income* core study after they received the invitation. We further include a dummy that controls for whether the individual answered within the deadline of the first call to participate in the survey, or only after the reminder, as well as an interaction term of the two. The random term  $\nu_i$  is assumed to be a time invariant random effect while  $\psi_{1,it}$  is assumed to be iid standard normally distributed.

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

## 4. Estimation results

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

<sup>13</sup>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).

<sup>14</sup>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.

<sup>15</sup>The pooled regression results are available in Appendix E

#### 4.1. First stage — Heckman correction

In the first stage regression the sample also includes all individuals in the selected age range for whom we only have only one observation.<sup>16</sup> Detailed results for women and men are presented in Tables 17 and 18 in Appendix E.

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

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

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

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

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

<sup>16</sup>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.

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

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

#### 4.2. Second stage — static models

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

For the static model, we find similar effects of personal and household characteristics in all specifications. See e.g. the results for the model with Big-Five factor markers for women in Table 3 and for men in Table 4.<sup>19</sup> For women, only few of the personal and household characteristics are individually significant in the self-employment equation. In contrast to this, almost all of them are highly statistically significant (i.e., most at the 1%-, and some at the 5%-level) in the other two equations. In other words, personal characteristics do not help us much to explain the difference between being self-employed or working as an employee, but they are helpful in explaining the difference between employment and unemployment or not participating in the labour market. The only variables that have a significant coefficient in the self-employment equation are age and household size (both at the 1%-level). We find that the self-employed are on average older, and that the larger the household, the more likely women are to choose self-employment over wage-employment.

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

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

For the Big-Five factor markers, effects differ by gender. For both women and men, individuals who are more open to experiences are also more likely to be self-employed than employees. The effect is however twice as strong in magnitude for men. For women, we find that also emotional stability is significant at the 10%-level. The negative sign implies that higher emotional stability reduces the chances to become self-employed. For men on

---

<sup>19</sup>The pooled regression results for the model with Big-Five factor markers are in Appendix E, Table 20. The results for all other models are available upon request.

#### 4. Estimation results

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

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

4. Estimation results

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

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

the other hand, we find that high scores for extraversion increase the likelihood of being self-employed, and that high scores for agreeableness reduce it. These effects are in line with our expectations.

In all static models, the inverse Mills ratio is highly significant for the self-employed and for the equation explaining “not in the labour force.” It has a positive sign in the self-employment equation and a negative sign in the equation for not in the labour force. For women, it is also significant and negative in the unemployment equation. This suggests that, keeping observed characteristics constant, attrition is correlated with the unobserved factors driving someone’s labor market status, implying that it is important to correct for attrition bias.

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

#### 4.3. Second stage — dynamic models

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

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

Third, when we test for the joint significance of the inverse Mills ratio we fail to reject the null hypothesis that the coefficients are jointly equal to zero at any conventional significance level (p-values 0.188 for women and 0.395 for men). The results for the regressions without

---

<sup>20</sup>See the corresponding regression results in Appendix E, Table 25 (pooled), 26 (women), and 27 (men). The coefficient on the EP distance is always insignificant in the self-employment equation.

<sup>21</sup>The results for the model with the lagged health index can be found in Appendix E, Table 28 (pooled), 29 (women), and 30 (men).

the Heckman correction are shown in Table 5 for women and Table 6 for men.<sup>22</sup> They indeed do not differ much from the results with the Heckman correction. This suggests that correcting for attrition bias is not essential once we estimate the dynamic model.

How do the results change from the static models once we include dynamic effects? For women, age loses its significance in the self-employment equation while for men, age remains significant at the 5%-level. The coefficient on household size for women, and on the dummy for having at least one child for men are no longer significant in the self-employment equation either.

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

Looking at the lagged labour market state variables we find persistence for both genders: Having been in a given labour state one period earlier increases the probability to be in that state – i.e. the diagonal in the block of coefficients for lags shows the largest values. This is what we would expect given the pattern of the transition probabilities in Table 2. The coefficient on lagged self-employment in the self-employment equation stands out as the largest of all, implying stronger state dependence in self-employment than in other labor market states.

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

---

<sup>22</sup>The regression results with the Heckman correction are in Appendix E, Tables 21 (women), 22 (men), and 23 (pooled). The pooled regression results without the Heckman correction are in Table 24.

<sup>23</sup>Initial values are also strongly significant with substantial coefficients. This indicates that the individual effects are correlated with the initial observation, as expected.

4. Estimation results

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

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

4. Estimation results

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

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

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

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

Based on dynamic model with Big-Five factor markers, 15310 observation pairs (n=2694).

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

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

## 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.<sup>24</sup> 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  $\mu_i$  from the multivariate normal distribution (with mean zero and covariance matrix given by  $\hat{L}\hat{L}^\top$ ). In each time period for each individual and labour state, we then draw independent error terms  $\epsilon_{ijt}$  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.

<sup>24</sup>This assumption is not too farfetched. All covariates have relatively small within variation.

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

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

Based on dynamic model with Big-Five factor markers, 19138 observation pairs (n=3325).

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

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.<sup>25</sup> The left panels show the results for all men and women and the right panels show the results for the 45 to 60 year old subsamples. Comparing the total shares of all labour market states with the shares in Table 2, 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

<sup>25</sup>We choose the model without Heckman correction since we could not reject the null hypothesis that the correction terms all have coefficient zero.

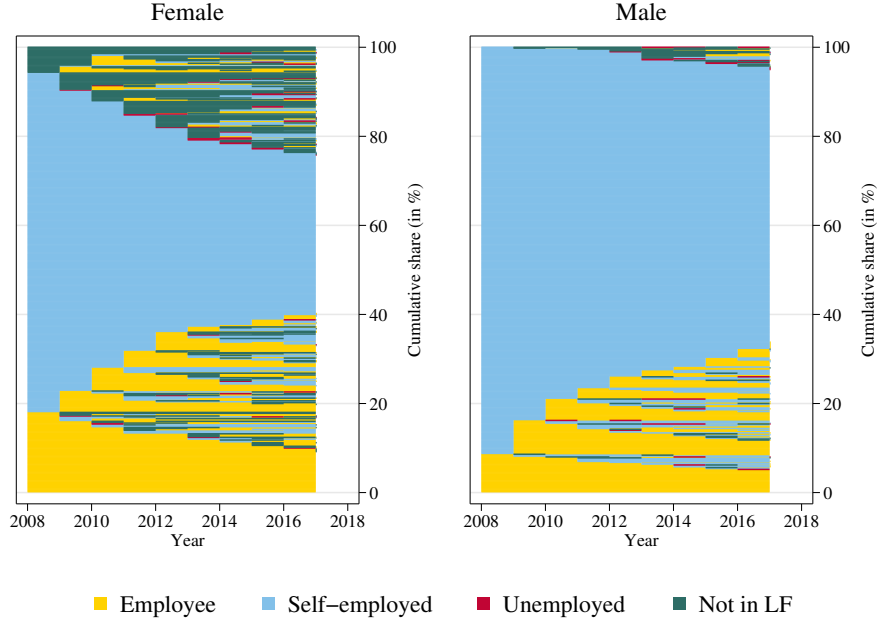


Figure 3: Simulated employment paths by gender: self-employed in 2007 (median characteristics)

personal and household characteristics equal to the median by gender for the self-employed in 2007. Thus the benchmark individuals have a partner and one child. The self-employed male benchmark has high education, the female has medium education. The median values for the Big-Five factor markers corresponding to these personal characteristics (for age ranging from 45 to 54 years) are presented in Table 16 in Appendix E.

Next we take 500 draws of the benchmark individual’s unobserved heterogeneity vector  $\mu_i$ , and then proceed in the same way as with the forward simulations for the transition probabilities, drawing the  $\epsilon_{ijt}$  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 Figure 3.

The sequence index plots show all simulated employment paths. The paths are stacked vertically on top of each other and each path is a (thin) horizontal line on which each year is coloured according to the labour market state the path takes in that year. We order the paths by labour market states, starting with the state in the first period, followed by the second period, etc. The scale of the y-axis reflects cumulative shares (in percent). On the righthand side of Figure 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

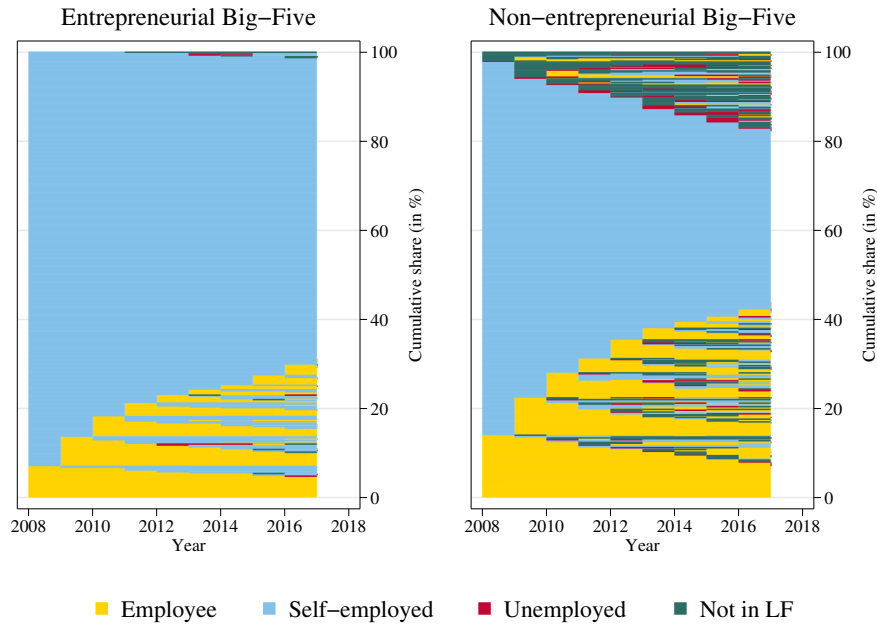


Figure 4: Simulated employment paths: male, self-employed in 2007

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.

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

Finally, since we are interested how personality affects individual probabilities to remain self-employed, we follow Obschonka et al. (2013) and define an “entrepreneurial” profile with high levels of extraversion, conscientiousness, emotional stability, and openness, and a low level of agreeableness. Taking the standardization of the factor markers into account we assign high and low values 2.5 and -2.5. In addition, we define a “non-entrepreneurial” profile that takes the opposite values. Figures 4 and 5 shows the sequence index plots for the

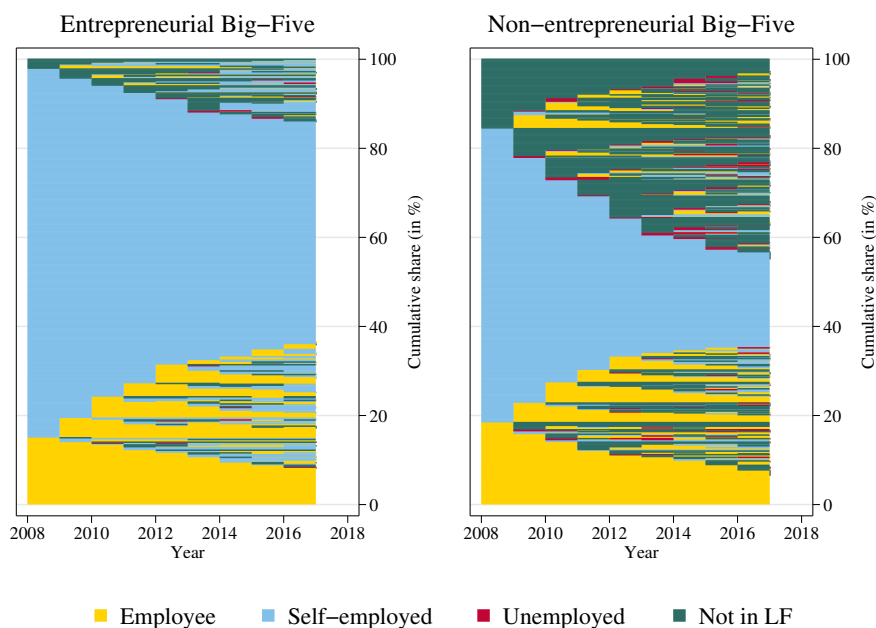


Figure 5: Simulated employment paths: female, self-employed in 2007

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 Figure 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.<sup>26</sup> Note that the probability of self-employment is always lower for a female individual in comparison to a male benchmark individual, also in cases where both have exactly the same characteristics. Furthermore, a shift from high to medium education increases the difference in probabilities between the “entrepreneurial” and “non-entrepreneurial” case. In addition to this, the difference between the lower and upper bound (the final two columns) within each case also increases. Finally, we also find that the

<sup>26</sup>See Figure 10 in Appendix F for a graphic representation of the evolution of the shares over time and type of benchmark individual.

5. Simulations

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

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

Based on dynamic model with Big-Five factor markers and no Heckman correction (n=500).

The benchmark individual is self-employed in 2007. The rightmost column gives the probability that the individual is self-employed in at least 5 of the 10 years.

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 Figure 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 microsimulation would over-predict the share of self-employed with (too) low pension savings.

### 5.3. *The impact of the macro economy*

As the number of self-employed in the Netherlands has increased in recent years, one might wonder if changes in the macro economy have been drivers of this increase. Our model can answer this question insofar that we estimate time effects for each year that capture the macro-economic changes. The estimates of the time effects for the dynamic model with the Big-Five factor markers and no Heckman correction are shown in Table 10. We find large positive effects for the unemployment probability, in line with the notion that 2008 was the star of the economic crisis. For self-employment, we find that most of the years do not differ significantly from the base year (2008). The only significant time effects on self-employment are a negative effect for men in 2009, and positive effects for women in the final two years (2016 and 2017). Keeping the trend as shown in Figure 1 in mind, this finding is actually as expected as there is no visible impact on self-employment by e.g. the Great Recession.

We can study the impact of macro variables further by repeating the simulation but leaving out the time effects. This has the same effect as assuming that the macroeconomic situation in 2008 holds for all years, which might be particularly interesting considering that the Great Recession started in the second quarter of that year. Table 11 shows the simulation results for this exercise.<sup>27</sup> Comparing them with the results in Table 9 we observe that the changes for women and men generally go in opposite directions: Without the year effects, the male benchmark individuals would all have a higher probability of remaining self-employed while the female individuals would have a lower probability. This is in line with the negative time effects for men and the mainly positive time effects for women in Table 10. Still, the differences between the probabilities of persistent self-employment (the final columns of Tables 11 and 9) accounting and not accounting for the macro-economic developments are rather small compared to the differences that can be explained by characteristics such as education level or an entrepreneurial individual attitude.

---

<sup>27</sup>See Figure 11 in Appendix F for a graphical representation.

## 6. Conclusion

Table 10: Time effects; dynamic model with Big-Five factor markers and no Heckman correction

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

Time effects from the regressions shown in Table 5 and 6. 2008 is the base year.

(Non-robust) standard errors in parentheses; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

## 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 estimated 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 edu-

Table 11: Probability (in %) that different benchmark individuals who are self-employed in 2007, are self-employed in later years: macro-economic situation as in 2008

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

Based on dynamic model with Big-Five factor markers and no Heckman correction (n=500).

The benchmark individual is self-employed in 2007. The rightmost column gives the probability that the individual is self-employed in at least 5 of the 10 years.

## 6. Conclusion

---

cation (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.

---

## References

- Albarrán, P., R. Carrasco, and J. M. Carro (2019). Estimation of dynamic nonlinear random effects models with unbalanced panels. *Oxford Bulletin of Economics and Statistics* 81(6), 1424–1441.
- Been, J. and M. Knoef (2017). Job-search requirements for unemployed at the end of working life effects on unemployment dynamics and self-employment probabilities. *Journal of Human Resources* 52(2), 491–530.
- Beugelsdijk, S. and N. Noorderhaven (2005). Personality characteristics of self-employed; an empirical study. *Small Business Economics* 24(2), 159–167.
- Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B* 35(7), 677–693.
- Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B* 37(9), 837–855.
- Blanchflower, D. G. (2000). Self-employment in OECD countries. *Labour Economics* 7(5), 471–505.
- Bolhaar, J., A. Brouwers, and B. Scheer (2016). De flexibele schil van de Nederlandse arbeidsmarkt: een analyse op basis van microdata. CPB Achtergronddocument, CPB.
- Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources* 43(4), 972–1059.
- Bosch, N. (2014). Succes als startende zelfstandige. CPB Achtergronddocument, CPB.
- Bosch, N., G. Roelofs, D. Van Vuuren, and M. Wilkens (2012). De huidige en toekomstige groei van het aandeel zzp'ers in de werkzame beroepsbevolking. CPB Achtergronddocument, CPB.
- Buddelmeyer, H. and M. Wooden (2011). Transitions out of casual employment: the Australian experience. *Industrial Relations: A Journal of Economy and Society* 50(1), 109–130.
- CBS (2014, December). Achtergrondkenmerken en ontwikkelingen van zzp'ers in Nederland. cbs.nl.
- Cheng, T. C. and P. K. Trivedi (2015). Attrition bias in panel data: a sheep in wolf's clothing? a case study based on the Mabel survey. *Health Economics* 24(9), 1101–1117.
- Cobb-Clark, D. A. and S. Schurer (2012). The stability of Big-Five personality traits. *Economics Letters* 115(1), 11–15.

- de Bresser, J. and M. Knoef (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., A. Van Soest, and E. Villagomez (2004). Mobility in the urban labor market: a panel data analysis for Mexico. *Economic Development and Cultural Change* 53(1), 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*, pp. 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*, pp. 114–178. MIT Press, Cambridge, MA.
- Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics* 16(2), 163–178.
- Kabátek, J. (2020). Happy birthday, you’re fired! the effects of an age-dependent minimum wage on youth employment flows in the Netherlands. *ILR Review*, forthcoming.
- Knoef, M., J. Been, R. Alessie, K. Caminada, K. Goudswaard, and A. Kalwij (2016). Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics & Finance* 15(1), 55–89.
- Mastrogiacomo, M. (2016). De pensioenpuzzel van zelfstandigen. Netspar Brief 7, Netspar.
- Mastrogiacomo, M. and R. J. Alessie (2015). Where are the retirement savings of self-employed? An analysis of ‘unconventional’ retirement accounts. Working Paper 454, De Nederlandsche Bank.
- Obschonka, M., E. Schmitt-Rodermund, R. K. Silbereisen, S. D. Gosling, and J. Potter (2013). The regional distribution and correlates of an entrepreneurship-prone personality profile in the United States, Germany, and the United Kingdom: A socioecological perspective. *Journal of Personality and Social Psychology* 105(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 & Economic Statistics* 30(3), 411–431.
- Rietveld, C. A., H. van Kippersluis, and A. R. Thurik (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. and T. Nijman (1992). Testing for selectivity bias in panel data models. *International Economic Review* 33(3), 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., M. N. Harris, and X. Zhao (2012). Ill-health and transitions to part-time work and self-employment among older workers. Technical report, Lancaster University WP 12/04. Available at SSRN 2269449.
- Zwinkels, W., M. Knoef, J. Been, C. Caminada, and K. Goudswaard (2017). Zicht op zzp-pensioen. Netspar Design Paper 91, Netspar.

## A. Self-employed individuals in the LISS panel

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

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

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

---

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

<sup>29</sup>In most households with more than one adult member, several adults participate in the LISS panel.

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

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

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

<sup>33</sup>See Appendix B.1 for a discussion of the inclusion of DGAs.

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

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

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

Despite the timing being different in the *Work and Schooling* classification compared to the income based classification the two should lead to a similar outcome. We expect small differences to arise because individuals generally do not switch occupations in January only. But since we have data over several years we will still pick-up the changes in labour market states in both data sets. Table 12 shows a comparison of the assigned labour market state according to the income based classification (“Income”), compared with the work and schooling based classification (“Work and schooling”) as well as the classification based upon the household box with background variables (“Background variables”).<sup>36</sup> The rows show the share of matches with labour market states in the “work and schooling” and “background variables” classifications for each labour market state in the income based classification. If the classifications would match perfectly, we would observe shares of 100% on the diagonal and zero elsewhere. As expected, we find that the match for the unemployed is not that good. Focusing on the self-employed and employees, we see that one fifth of the individuals who report income from self-employment activities are classified as employees in the work and schooling survey. Similarly, 26% of those with self-employment income is classified as an employee according to the background variables based classification. Only 63% of those with self-employment income are classified as self-employed according to the latter classification, compared to 71% for the work and schooling based classification.

The corresponding self-employment shares, calculated as the number of self-employed individuals per year as a fraction of the respective number of individuals in the working population (i.e. the sum of employees and self-employed), in the two samples after these restrictions are shown in Figure 6.

---

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

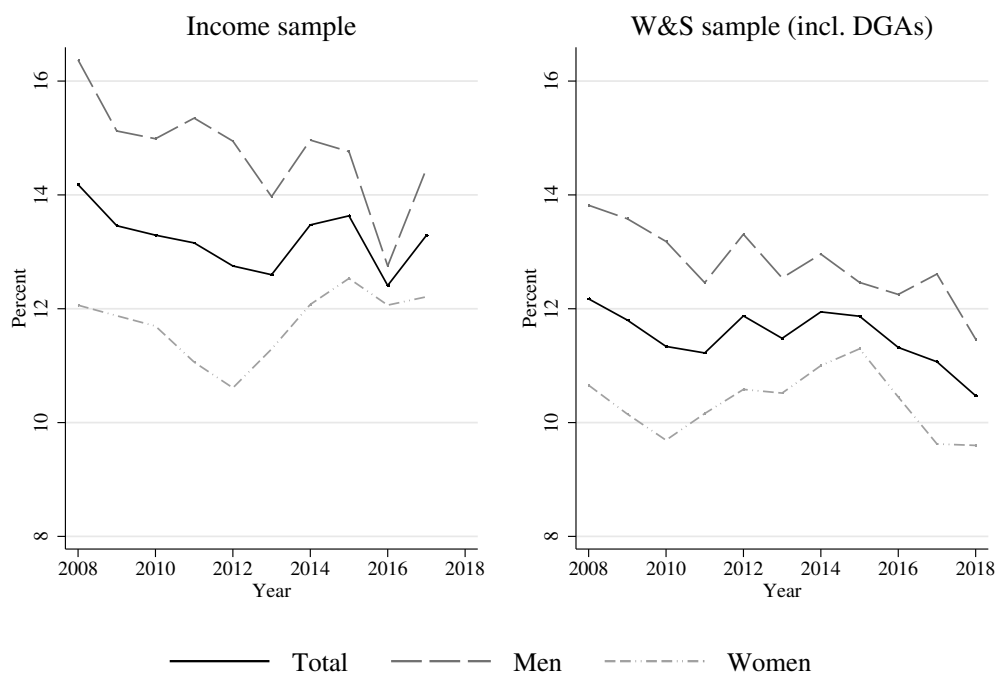
<sup>35</sup>Note that there are almost 11,000 observations combined for the years 2007 and 2018 that we cannot use because of the different timing of the two studies.

<sup>36</sup>The comparison is made for all observations in the corrected income sample excluding the corrected breaks (see section ??).

Table 12: Matching of labour market states across definitions (in % of income definition)

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

Source: LISS panel, own calculations.



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

Figure 6: Comparison of self-employment shares in working population of 25-60 year olds across samples.

There are two observations that we can make based on this figure. First, the share of self-employed in the working population is lower than in the actual population shown in Figure 1. While the share in the income based classification looks more or less equal in 2008, we have to take into consideration that CBS classifies individuals with multiple occupations based on the one in which they work most hours. Hence, as we also include part-time self-employed, the share in the LISS panel, if it were representative, should actually be higher than the population share.<sup>37</sup> Second, we observe an upward trend in self-employment shares according to none of the classifications in the LISS samples. These two observations hint that we may have to deal with both initial selection (lower initial shares) and attrition bias (lack of upward trend) in the LISS panel. We will discuss these issues in more detail below.

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

The model based sample restriction, i.e. having at least two consecutive observations, not only leads to a loss of data but also to a change in the evolution of employment shares over time in the new sample. In both the income and work based sample we see a clear downward trend after discarding the individuals with no sequence as well as any observations following a break in a sequence. The left panel of Figure 7, in comparison with the left panel in Figure 6 illustrates this for the income survey based classification. Thus the restriction of the sample to sequences creates (or worsens) attrition bias.

A closer look at the observations that are discarded because of the sequence restriction reveals that only a third of these observations belong to individuals who only participate in one wave of the income survey. That is, two thirds of the observations could be retained if we can correct for the break in the corresponding sequence. Furthermore, these are the observations (rather than the single wave answers) that are driving the downward trend shown in the left panel of Figure 7.

We take a conservative approach to filling in the breaks in sequences by imputing plausible values. That is, we only fill in one-year breaks, which account for the majority of all the breaks in the sequences. We however do not limit the correction to one break per individual but may fill in several breaks as long as these are only one period long. Nor do we extend a series beyond what we observe — that is, if an individual answers the work and schooling survey for more periods than the income survey, we do not use these other periods from the work and schooling survey to top up the income survey series. The procedure is as follows: we only consider the labour status information from the work based definition to fill in gaps in the income sample and vice versa, ignoring the information that is available in the background variables. We choose to do this because the number of correct matches between our definitions and a classification based on the background variables (as shown in Table 12) is quite low and we want to avoid generating false labour market state transitions

---

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



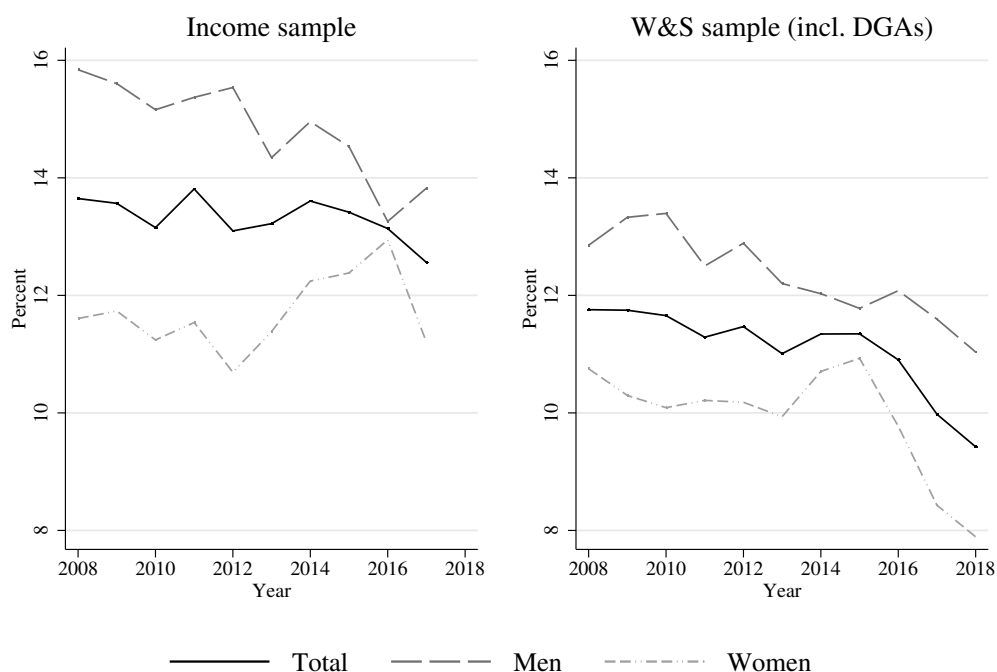
Data source: LISS Panel – Work & Schooling, Income, Background Variables

Figure 7: Comparison of self-employment shares based on income before/after correction

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

Based on this strategy we end up filling up at least one gap for more than one thousand individuals in the income survey sample. This in turn helps us to retain more than three thousand observations that would otherwise have to be discarded. More importantly, the downward trend in the self-employment shares is less pronounced after the correction. Figure 7 illustrates this for the income based classification. The impact of the correction on the work based classification is not that strong. This is likely because the number of additional observations that can be retained is rather low compared to the number of observations that is affected by the sequence restriction.

Figure 8 shows the corrected regression samples for both definitions. We can see that both samples display a slightly stronger downward trend in self-employment shares compared to the raw data in Figure 6. This negative trend also seems to worsen during later years and the effect is stronger in the work based classification. Because of this, and also due to the



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

Figure 8: Comparison of self-employment shares in corrected income and work panel

lower self-employment shares, we choose the income based classification for our analysis. The results discussed below, unless otherwise mentioned, are therefore for the income based classification (and the sample corresponding to it).

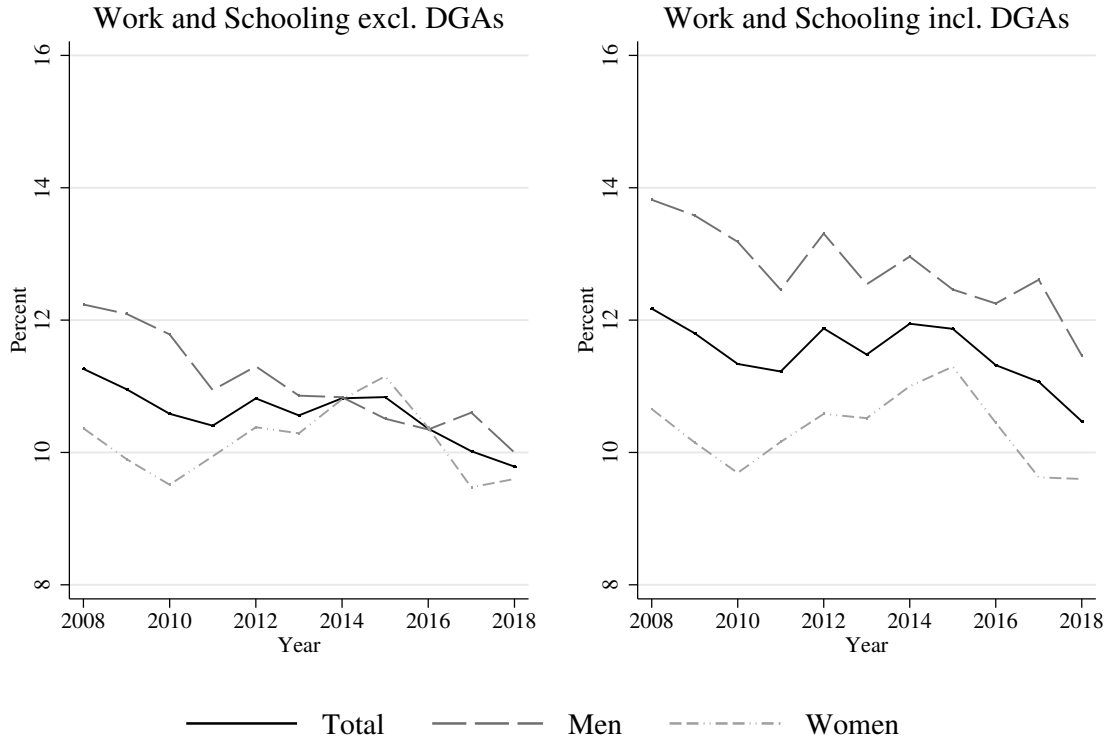
## B. DGAs and the self-employed

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

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

<sup>38</sup>Taken from the English version codebook of the Work and Schooling core study wave 10.

including DGAs based on question cwxxx121 in the self-employment definition should not lead to a mistake in our definition of the self-employed.



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

Figure 9: Comparison of self-employment shares with and without DGAs in raw sample

Figure 9 illustrates this further. By comparing the left and the right panel we see that the inclusion of DGAs shifts the shares more or less equally across all periods, which is consistent with evidence presented in CBS (2014, p.13, figure 3.2.1). It does so mostly through increasing the share of self-employed among the working men, which is consistent with the finding reported by CBS in the same report that 80% of DGAs were men in 2012.<sup>39</sup>

### B.2. DGAs in the income survey

In the income survey, question cixxx008 individuals are asked: *Did you receive income as employee in [t-1]?* including the explanation that *A [DGA] generally receives income as an employee as well and so please answer YES here.* The question block on self-employment on the other hand makes no special mention of DGAs. The section is however vague when asking individuals *[w]hich work situation as described [...] in [t - 1] applied to you (or for a*

<sup>39</sup>The shares presented in Figure 9 are calculated for all individuals from 25 to 60 years of age. The sample is further restricted to those for whom the covariates of the regression analysis are not missing. It is not conditional on observations being available for at least two periods consecutively.

part of  $[t - 1]$ ? and offering statements such as *work as an entrepreneur or as a freelancer (alongside a job)*, or *a company owner*, or *make profit (or losses) through enterprise in some other way [...]* in questions cixxx37 to cixxx044. DGAs are therefore only included in our definition of the self-employed if they self-identify as one of these options, otherwise they will be categorised as employees.

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

## C. Model details

### C.1. Correlation among the random effects

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

$$\mu_i = \xi_i^\top L^\top \tag{7}$$

where  $\xi_i$  is a  $J \times 1$  vector of independent standard normal distributed variables, and  $L$  is the lower triangular Cholesky factorization of  $\mu_i$ 's covariance matrix  $W$ , such that  $LL^\top = W$ .

Substituting (7) and (3) in (2), and writing the utilities in vectorised form we have

$$y_{it}^* = X_{it}\beta + y_{it-1}\gamma + y_{i0}\delta + \xi_i^\top L^\top + \epsilon_{it} \quad i = 1, \dots, N ; t = 2, \dots, T \tag{8}$$

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

---

<sup>40</sup>See <https://www.cpb.nl/publicatie/macro-economische-verkenning-mev-2019>

### C.2. Likelihood function

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

$$Prob(y_{it} = j | X_{it}, y_{it-1}, y_{i0}, \xi_i) = \frac{\exp(y_{it-1}\gamma_j + y_{i0}\delta_j + X_{it}\beta_j + \xi_i^\top L_j^\top)}{\sum_{k=0}^J \exp(y_{it-1}\gamma_k + y_{i0}\delta_k + X_{it}\beta_k + \xi_i^\top L_k^\top)} \quad (9)$$

where  $L_j$  is the  $j^{\text{th}}$  row of  $L$ .

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

$$Prob(y_i | X_i, y_{i0}, \xi_i) = \prod_t \prod_j Prob(y_{it} = j | X_{it}, y_{it-1}, y_{i0}, \xi_i)^{\mathbb{D}_{ijt}} \quad (10)$$

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

$$Prob(y_i | X_i, y_{i0}) = \int_{\xi_i} Prob(y_i | X_{it}, y_{it-1}, y_{i0}, \xi_i) f(\xi_i) d\xi_i \quad (11)$$

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

$$\log \mathcal{L} = \sum_{i=1}^N \log Prob(y_i | X_i, y_{i0}) \quad (12)$$

### C.3. Maximum Simulated Likelihood

The results of this section hold for both the static and the dynamic model. Note that in its current form the multinomial logit model described by equations (9) – (12) is not identified as there are too many parameters. For identification purposes we therefore take  $j = 0$ , that is wage-employment, as the base category.  $\beta_0, \gamma_0, \delta_0$  and the first column in  $L$  are normalised to zero. The parameters for the other alternatives  $j = 1, 2, 3$  must be interpreted relative to the base category.<sup>41</sup> For example, the covariance matrix is given by

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

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

Furthermore, the probabilities given by equation (11), i.e.  $\int_{\xi_i}$  in particular, have to be simulated. As Train (2009, chapter 10) writes,  $\log \hat{P}$  is not an unbiased estimator for  $\log P$

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

because of the non-linear log operation, even if  $\hat{P}$  is an unbiased estimator of  $P$ . Thus the bias in the simulator for  $\log Prob(y_i|X_i, y_{i0})$  translates into a bias in the maximum simulated likelihood estimator. This bias however diminishes as more draws are used in the simulation. Using a large number of draws for  $\xi_i$  on the other hand increases the computational burden of the estimation. One way to reduce this burden is, instead of using independent random draws, to use an alternative method that provides better coverage of the support of the distribution of the individual and therefore leads to greater accuracy for a given number of draws. This can be achieved using Halton draws. Both Bhat (2001) and Train (2009, chapter 9.3.3) have shown that e.g. 100 Halton draws can provide more precise results than 1000 random draws.

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

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

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

The simulated loglikelihood function is thus given by

$$\log \mathcal{S}\mathcal{L} = \sum_{i=1}^N \log \left[ \frac{1}{R} \sum_{r=1}^R \prod_{t=2}^T \left( \prod_{j=1}^J Prob(y_i|X_{it}, y_{it-1}, y_{i0}, \xi_i)^{\mathbb{D}_{ijt}} \right)^{\mathbb{S}_{it}} \right] \quad (13)$$

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

All model specifications for the correlated random effects models reported in section 4 are estimated using own code written in Matlab 2017a. We solve them as an unconstrained minimisation problem using KNITRO as a solver and supplying the gradient as defined in the next subsection. As the best guess for the starting values we first estimate a pooled multinomial regression on the same covariates, including the lags, for each specification. The only difference is the non-inclusion of the unobserved individual heterogeneity. For this step we make use of the function "mnrfit" in Matlab's statistics toolbox.

#### C.4. First order derivatives

Let for notational ease  $P(y_{ijtr}) \equiv \text{Prob}(y_{ijtr} | X_{it}, y_{it-1}, y_{i0}, \xi_i)$  and similarly for all other probabilities. The simulated loglikelihood function from equation (13) is then given by

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

Furthermore let vector  $\theta$  be a vector with  $\kappa$  elements containing the stacked columns of all the coefficient matrices  $\beta, \gamma, \delta$ , as well as a vector of all the elements of the Cholesky factorisation  $[l_{11}, l_{21}, l_{22}, l_{31}, l_{32}, l_{33}]'$ . And lastly, let  $V_{it}$  be a vector of the covariates, where  $X_{it}, y_{it-1}, y_{i0}$  are stacked accordingly to the elements in  $\theta$ . Note that e.g. for a covariate in  $X_{it}$  its index  $\kappa$  in  $V_{it}$  is jointly defined by  $k$  and  $j$ . The last rows in  $V_{it}$  are given by the vector  $[\xi_1, \xi_1, \xi_2, \xi_1, \xi_2, \xi_3]'$ .

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

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

## D. Health index

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

*D. Health index*

---

regression results.

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

Table 13: Regression results for health index

|                                   | Regression sample only | incl. single observations |
|-----------------------------------|------------------------|---------------------------|
|                                   | Health in general      | Health in general         |
| Health compared to $t - 1$ : poor | -1.555***<br>(0.0692)  | -1.536***<br>(0.0674)     |
| — moderate                        | -0.585***<br>(0.0238)  | -0.581***<br>(0.0231)     |
| — very good                       | -0.0120<br>(0.0225)    | -0.0166<br>(0.0217)       |
| — excellent                       | 0.455***<br>(0.0443)   | 0.421***<br>(0.0421)      |
| Hinder in daily life (index)      | -0.290***<br>(0.0114)  | -0.282***<br>(0.0110)     |
| Long-standing disease             | 0.264***<br>(0.0210)   | 0.279***<br>(0.0203)      |
| Regularly pain in joints          | -0.101***<br>(0.0200)  | -0.107***<br>(0.0193)     |
| — Hearth problems                 | -0.185***<br>(0.0471)  | -0.194***<br>(0.0456)     |
| — Breathing problems              | -0.166***<br>(0.0362)  | -0.163***<br>(0.0350)     |
| — Coughing, flu, etc.             | -0.208***<br>(0.0217)  | -0.194***<br>(0.0210)     |
| — Stomach/intestinal problems     | -0.207***<br>(0.0249)  | -0.210***<br>(0.0241)     |
| — Headaches                       | -0.107***<br>(0.0209)  | -0.113***<br>(0.0202)     |
| — Fatigue                         | -0.253***<br>(0.0198)  | -0.245***<br>(0.0191)     |
| — Sleeping problems               | -0.0814***             | -0.0864***                |

*D. Health index*

Table 13: Regression results for health index (continued)

|                            | Regression sample only<br>Health in general | incl. single observations<br>Health in general |
|----------------------------|---|--|
|                            | (0.0219)                                    | (0.0213)                                       |
| Other recurrent complaints | -0.249***                                   | -0.244***                                      |
|                            | (0.0259)                                    | (0.0251)                                       |
| No recurrent complaints    | 0.119***                                    | 0.113***                                       |
|                            | (0.0231)                                    | (0.0223)                                       |
| Taking medicine for:       |   |  |
| — high blood pressure      | -0.123***                                   | -0.131***                                      |
|                            | (0.0291)                                    | (0.0283)                                       |
| — diabetes                 | -0.311***                                   | -0.327***                                      |
|                            | (0.0507)                                    | (0.0495)                                       |
| — joint pain/infection     | -0.166***                                   | -0.173***                                      |
|                            | (0.0355)                                    | (0.0345)                                       |
| — hormonal osteoporosis    | -0.371***                                   | -0.325**                                       |
|                            | (0.137)                                     | (0.135)  |
| Not taking any medicine    | 0.231***                                    | 0.214***                                       |
|                            | (0.0252)                                    | (0.0243)                                       |
| Hospital stay              | -0.0646**                                   | -0.0749***                                     |
|                            | (0.0256)                                    | (0.0248)                                       |
| Female                     | 0.0583***                                   | 0.0653***                                      |
|                            | (0.0155)                                    | (0.0151)                                       |
| Age                        | -0.00781***                                 | -0.00756***                                    |
|                            | (0.000829)                                  | (0.000798)                                     |
| Underweight                | -0.140**                                    | -0.0958*                                       |
|                            | (0.0567)                                    | (0.0542)                                       |
| Overweight                 | -0.162***                                   | -0.170***                                      |
|                            | (0.0163)                                    | (0.0158)                                       |
| Obese                      | -0.343***                                   | -0.342***                                      |
|                            | (0.0233)                                    | (0.0226)                                       |
| Constant cut1              | -4.397***                                   | -4.305***                                      |
|                            | (0.123)                                     | (0.120)  |
| Constant cut2              | -1.964***                                   | -1.888***                                      |
|                            | (0.115)                                     | (0.112)  |
| Constant cut3              | 0.725***                                    | 0.778***                                       |
|                            | (0.114)                                     | (0.112)  |
| Constant cut4              | 1.960***                                    | 2.013***                                       |

E. Additional tables

Table 13: Regression results for health index (continued)

|                                  | Regression sample only | incl. single observations |
|----------------------------------|------------------------|---------------------------|
|                                  | Health in general      | Health in general         |
|                                  | (0.115)                | (0.112)                   |
| Depression, anxiety controls     | Yes                    | Yes                       |
| Difficulty with actions controls | Yes                    | Yes                       |
| Other medication controls        | Yes                    | Yes                       |
| Education controls               | Yes                    | Yes                       |
| Year fixed effects               | Yes                    | Yes                       |
| Observations                     | 28926                  | 30669                     |

Standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

E. Additional tables

Table 14: Observed transition probabilities (in %) excluding/including corrections

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

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

Source: LISS Panel, own calculations.

Table 15: Simulated transition probabilities (in %), pooled

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

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

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

*E. Additional tables*

Table 16: Personal characteristics for individual simulation

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

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

*E. Additional tables*

Table 17: First stage Heckman regression results, women

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

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

E. Additional tables

Table 18: First stage Heckman regression results, men

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

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

*E. Additional tables*

Table 19: First stage Heckman regression results, pooled

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

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

E. Additional tables

Table 20: Static model with Big-Five factor markers, pooled

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 21: Dynamic model with Big-Five factor markers, women

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 22: Dynamic model with Big-Five factor markers, men

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 23: Dynamic model with Big-Five factor markers, pooled

|                              | Self-employed              | Unemployed                 | Not in Labour Force        |
|------------------------------|----------------------------|----------------------------|----------------------------|
| Constant                     | -6.81***<br>(0.7856)       | -7.17***<br>(0.7682)       | -4.89***<br>(0.5285)       |
| Age                          | 0.02*<br>(0.0084)          | 0.05***<br>(0.0087)        | 0.05***<br>(0.0063)        |
| Female                       | -0.04<br>(0.262)           | 0.48**<br>(0.2256)         | 0.33*<br>(0.1885)          |
| Has partner                  | 0.05<br>(0.2356)           | -0.35<br>(0.2375)          | -0.37**<br>(0.1859)        |
| Has child                    | -0.25<br>(0.2619)          | -0.28<br>(0.2505)          | -0.23<br>(0.1949)          |
| Female x partner             | 0.22<br>(0.3025)           | -0.3<br>(0.268)            | 0.73***<br>(0.2181)        |
| Female x has child           | -0.16<br>(0.2704)          | 0.19<br>(0.2548)           | -0.1<br>(0.1969)           |
| Middle education             | -0.13<br>(0.3972)          | -0.79***<br>(0.2647)       | -1.12***<br>(0.2138)       |
| High education               | 0.21<br>(0.4078)           | -1.53***<br>(0.2901)       | -1.78***<br>(0.2322)       |
| Household size               | 0.17*<br>(0.0876)          | 0<br>(0.0846)              | 0.03<br>(0.0663)           |
| F1: extraversion             | 0.15**<br>(0.0728)         | -0.04<br>(0.0714)          | -0.03<br>(0.0515)          |
| F2: agreeableness            | -0.11<br>(0.0708)          | 0.04<br>(0.0695)           | -0.02<br>(0.0502)          |
| F3: conscientiousness        | -0.14*<br>(0.077)          | -0.15*<br>(0.0772)         | -0.21***<br>(0.0577)       |
| F4: emotional stability      | -0.17***<br>(0.0655)       | -0.36***<br>(0.0652)       | -0.28***<br>(0.0481)       |
| F5: openness for experience  | 0.26***<br>(0.0711)        | 0.19***<br>(0.0716)        | 0.11**<br>(0.0538)         |
| Last state: self-employed    | 4.06***<br>(0.1651)        | 0.93**<br>(0.3819)         | 1.14***<br>(0.2277)        |
| Last state: unemployed       | 0.3<br>(0.3483)            | 2.75***<br>(0.2023)        | 1.55***<br>(0.1664)        |
| Last state: not in LF        | 0.63***<br>(0.1985)        | 1.19***<br>(0.1604)        | 1.89***<br>(0.0885)        |
| Initial state: self-employed | 4.95***<br>(0.3848)        | 1.76***<br>(0.3905)        | 2.16***<br>(0.2708)        |
| Initial state: unemployed    | 2.92***<br>(0.5375)        | 3.78***<br>(0.3869)        | 3.44***<br>(0.3484)        |
| Initial state: not in LF     | 2.51***<br>(0.2842)        | 3.53***<br>(0.23)          | 4.27***<br>(0.1894)        |
| Inverse Mills Ratio          | -0.36<br>(0.96)            | -1.03<br>(1.1517)          | -2.06***<br>(0.7955)       |
| L                            | 2.2***<br>(0.1536)         |                            |                            |
|                              | 1.11***<br>(0.1756)        | 1.48***<br>(0.147)         |                            |
|                              | 1.1***<br>(0.1372)         | 1.52***<br>(0.1216)        | 0.65***<br>(0.1206)        |
| W                            | 4.8371<br>2.4439<br>2.4218 | 2.4439<br>3.4189<br>3.4722 | 2.4218<br>3.4722<br>3.9505 |
| Observations:                | 26402                      |                            |                            |
| Nr. of Individuals:          | 5914                       | 57                         |                            |
| Loglikelihood:               | -10201.11                  |                            |                            |

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

*E. Additional tables*

Table 24: Dynamic model with Big-Five factor markers and no Heckman correction, pooled

|                              | Self-employed        | Unemployed           | Not in Labour Force  |
|------------------------------|----------------------|----------------------|----------------------|
| Constant                     | -6.98***<br>(0.5708) | -7.67***<br>(0.5313) | -5.89***<br>(0.3637) |
| Age                          | 0.02**<br>(0.0067)   | 0.05***<br>(0.0066)  | 0.06***<br>(0.0049)  |
| Female                       | -0.05<br>(0.2604)    | 0.46**<br>(0.223)    | 0.28<br>(0.1864)     |
| Has partner                  | 0.05<br>(0.2342)     | -0.34<br>(0.2374)    | -0.35*<br>(0.1855)   |
| Has child                    | -0.26<br>(0.2604)    | -0.29<br>(0.2499)    | -0.26<br>(0.1942)    |
| Female x partner             | 0.21<br>(0.3007)     | -0.32<br>(0.2678)    | 0.7***<br>(0.2174)   |
| Female x has child           | -0.16<br>(0.2689)    | 0.2<br>(0.2547)      | -0.07<br>(0.1959)    |
| Middle education             | -0.12<br>(0.3894)    | -0.75***<br>(0.2615) | -1.04***<br>(0.2114) |
| High education               | 0.23<br>(0.3974)     | -1.47***<br>(0.2856) | -1.69***<br>(0.2292) |
| Household size               | 0.17*<br>(0.0873)    | -0.01<br>(0.0832)    | 0.01<br>(0.0657)     |
| F1: extraversion             | 0.15**<br>(0.0724)   | -0.05<br>(0.0713)    | -0.03<br>(0.0514)    |
| F2: agreeableness            | -0.11<br>(0.0707)    | 0.04<br>(0.0692)     | -0.02<br>(0.0501)    |
| F3: conscientiousness        | -0.13*<br>(0.0692)   | -0.11*<br>(0.0659)   | -0.14***<br>(0.0511) |
| F4: emotional stability      | -0.17***<br>(0.0651) | -0.37***<br>(0.0647) | -0.29***<br>(0.0477) |
| F5: openness for experience  | 0.26***<br>(0.0704)  | 0.18***<br>(0.0705)  | 0.09*<br>(0.0533)    |
| Last state: self-employed    | 4.04***<br>(0.1567)  | 0.86**<br>(0.3751)   | 1.04***<br>(0.2224)  |
| Last state: unemployed       | 0.29<br>(0.3482)     | 2.77***<br>(0.2017)  | 1.58***<br>(0.166)   |
| Last state: not in LF        | 0.64***<br>(0.1961)  | 1.22***<br>(0.1567)  | 1.95***<br>(0.085)   |
| Initial state: self-employed | 4.95***<br>(0.3845)  | 1.77***<br>(0.3913)  | 2.16***<br>(0.2694)  |
| Initial state: unemployed    | 2.91***<br>(0.5352)  | 3.77***<br>(0.3865)  | 3.43***<br>(0.3468)  |
| Initial state: not in LF     | 2.5***<br>(0.2837)   | 3.52***<br>(0.2298)  | 4.26***<br>(0.1888)  |
| L                            | 2.2***<br>(0.1535)   |                      |                      |
|                              | 1.12***<br>(0.1759)  | 1.47***<br>(0.1482)  |                      |
|                              | 1.1***<br>(0.1371)   | 1.52***<br>(0.1221)  | 0.66***<br>(0.1204)  |
| W                            | 4.8245               | 2.4551               | 2.4063               |
|                              | 2.4551               | 3.4135               | 3.4555               |
|                              | 2.4063               | 3.4555               | 3.9316               |
| Observations:                | 26402                |                      |                      |
| Nr. of Individuals:          | 5914                 |                      |                      |
| Loglikelihood:               | -10204.73            | 58                   |                      |

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 25: Dynamic model with EP distance, pooled

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 26: Dynamic model with EP distance, women

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

*E. Additional tables*

Table 27: Dynamic model with EP distance, men

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 28: Dynamic model with health index, pooled

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 29: Dynamic model with health index, women

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

E. Additional tables

Table 30: Dynamic model with health index, men

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

Regression including year fixed effects.

(Non-robust) standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.10

F. Additional figures

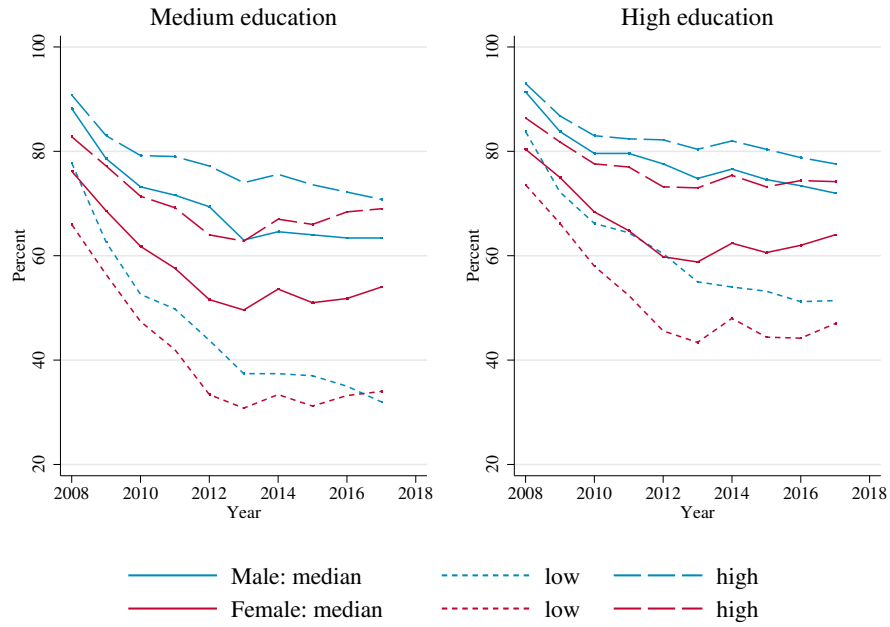


Figure 10: Share of employment paths spent in self-employment for different Big-Five factor markers (benchmark individual starts as self-employed in 2007)

F. Additional figures

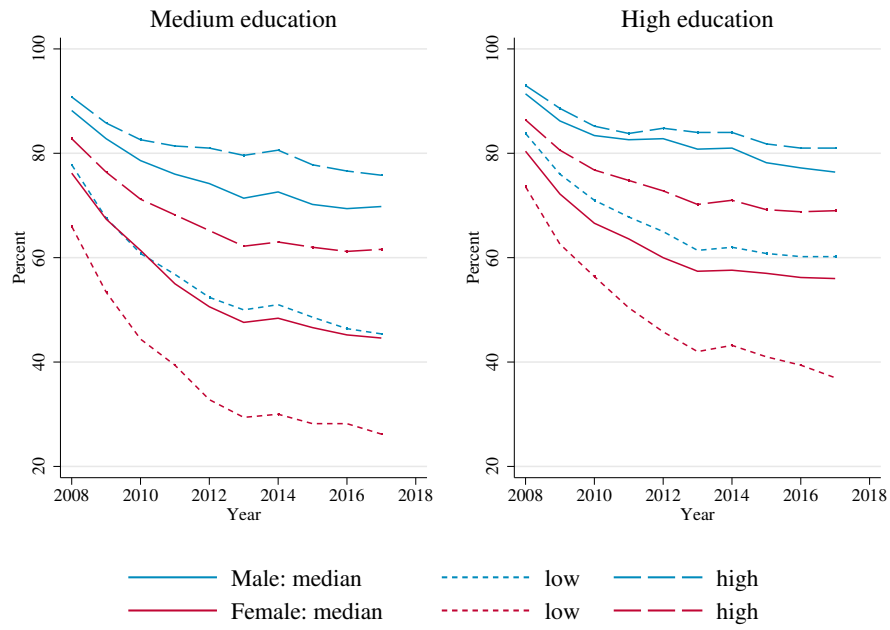


Figure 11: Share of employment paths spent in self-employment for different Big-Five factor markers (benchmark individual starts as self-employed in 2007, no year effects)