

Retaining through Training: Even for Older Workers*

Matteo Picchio^{a,e,†} and Jan C. van Ours^{a,b,c,d,e}

^a *Tilburg University, CentER, ReflecT, The Netherlands*

^b *University of Melbourne, Australia*

^c *CESifo, Germany*

^d *CEPR, United Kingdom*

^e *IZA, Germany*

April 29, 2011

Abstract

This paper investigates whether on-the-job training has an effect on the employability of workers. Using data from the Netherlands we disentangle the true effect of training incidence from the spurious one determined by unobserved individual heterogeneity. We also take into account that there might be feedback from shocks in the employment status to future propensity of receiving firm-provided training. We find that firm-provided training significantly increases future employment prospects. This finding is robust to a number of robustness checks. It also holds for older workers, suggesting that firm-provided training may be an important instrument to retain older workers at work.

Keywords: training, employment, human capital, older workers.

JEL classification codes: C33, C35, J21, J24, M53

*We acknowledge financial support for this research by Stichting Instituut GAK, through Reflect, the Research Institute for Flexicurity, Labor Market Dynamics and Social Cohesion at Tilburg University.

[†]Corresponding author. Department of Economics, Tilburg University, PO BOX 90153, 5000 LE Tilburg, The Netherlands. Tel: +31 13 4662534. Fax: +31 134663042.

E-mail addresses: m.picchio@uvt.nl (M. Picchio), vanours@uvt.nl (J.C. van Ours).

1 Introduction

Active Labour Market Policies (ALMP) aim to increase employment rates by stimulating job finding rates and reducing job separation rates. In recessions ALMP are used to dampen the effects of the downturn in employment. In the recent crisis temporary shorter working hours arrangements, often in combination with increased training of workers, were used as instruments. Indeed several countries reported measures to provide training to existing workers at risk of job loss (OECD, 2010).

On-the-job training is an investment by individuals and firms which is aimed at upgrading individuals' human capital and which might pay back in terms of better labour market performances and higher productivity in the future. Training helps individuals to meet the needs of a changing economic and technological environment, enlarge the spectrum of competences, increase productivity, send good signals to employers, and thereby avoid unemployment events and return faster at work in case of job loss.

The question whether adult training affects labour market performances and productivity of individuals has been the core of substantial debates and research investigations. A strand of the literature analyses whether training programmes are tools to integrate the unemployed into the work force. Training programmes are found to have a modest or no effect on unemployment exit rates (Gerfin and Lechner, 2002; Andr n and Andr n, 2006; Cr pon et al., 2007; Lechner et al., 2008; Lalive et al., 2008; Sianesi, 2008) and on wages (Heckman et al., 1999). Positive effects are instead found on the future employment stability (Cr pon et al., 2007; Lechner et al., 2008) and Jones et al. (2009) show that training is significantly and positively associated with job satisfaction.

A similar branch of this literature has tried to understand how on-the-job training affects wages and other performance measures. Gritz (1993) concludes that participation in training improves the employment prospects, especially for women, the youth, and minorities. Bonnal et al. (1997) report that, in the private sector, on-the-job training increases the employment rates, especially for young workers. Bartel (1995) shows that, at firm level, wages and productivity are positively affected by on-the-job training. Finally, several other studies find that on-the-job training has a significantly positive impact on productivity (see, among others, Bartel, 1994; Barrett and O'Connell, 2001; Conti, 2005; Dearden et al., 2006).

In the Netherlands, previous evaluation studies have tried to infer the impact of different kind of training programmes on future employment prospects. Ridder (1986) distinguishes between training, recruitment, and employment programmes. He finds that especially women and ethnic minorities benefit from training programmes. Sanders and de Grip (2004) investigate whether training participation might affect low-skilled workers'

firm-internal and firm-external mobility. They find that training increases firm-internal mobility, but it does not affect firm-external employability. Lastly, [Pavlopoulos et al. \(2009\)](#) analyse the effect of training on low pay mobility. They find that training increases the chances for upward wage mobility and therefore improves earnings prospects.

All in all, the effect of training on employment prospects is often found to be somewhat disappointing. Either training does not increase job finding rates significantly or even modest negative effects are found (see also [Kluve \(2010\)](#) for a recent overview study). Whereas the effect of training on job finding rates has been studied quite frequently, the effect on job separation rates is rarely investigated.¹

Using data on the Netherlands from the European Community Household Panel (ECHP), we investigate whether firm-provided training enhances the probability of retaining workers into the workforce in the Netherlands. From the econometric viewpoint, it is challenging to disentangle the pure effect of training from the spurious one determined by individual unobserved heterogeneity. Unobserved heterogeneity might indeed jointly determine the likelihood of training participation and the labour market performances: motivation, labour market attachment, innate ability. Using semiparametric techniques to control for the endogeneity of training participations, we explicitly model the inter-related dynamics leading to training and determining the future employment prospects. We also take into account that there might be feedback from current employment shocks to individuals' future probability of receiving firm-provided training. As a result, we are able to estimate policy-relevant effects of on-the-job training participation on employment prospects later in life.

We find that in the Netherlands firm-provided training significantly improves future employability, i.e training leads to retaining.² We also focus on the effect for older workers. As in many other European countries, the labour market position of older workers is cause for concern in the Netherlands, given that the demographic trends are causing an ageing of the workforce and that older workers' job separations are often a one-way street out of the labour force and into long-term unemployment. We find that older workers who receive training are more likely to remain employed. We suggest that additional on-the-job training of workers, especially older workers, can be influenced by government

¹Using cross-country time series data on unemployment rates [Boone and van Ours \(2009\)](#) find that training has a significant negative effect on unemployment. They do not attribute this to the positive effect of training on the job finding rate but to the negative effect on the job separation rate. Training increases the human capital of participants and therefore the quality of their post-unemployment job increases, leading to lower job separation rates.

²In a companion paper, [Picchio and van Ours \(2011\)](#), we show that in the Netherlands firm-sponsored training is affected by labour market imperfections but not by product market competition. If labour mobility goes up employers are less willing to invest in training. If product market competition increases employer sponsored training is unaffected.

policy, for example by providing the employers with age-specific subsidies to stimulate firm-provided training. Furthermore, an age-specific firing tax may persuade employers to train older workers, increasing thereby older workers' employability.

This paper is set up as follows. The data are described in Section 2. Section 3 formalizes the econometric model and clarifies the identification strategy. The estimation results are presented and discussed in Section 4. Section 5 concludes.

2 Data Description

The data used in this paper are from the 1994–2001 waves of the longitudinal dimension of the ECHP, a rotating panel survey based on harmonized methodology and definitions across several European countries. The ECHP contains nationally representative samples of households and covers a large set of topics such as work, income, financial situation, housing, family, health, training and education, and social relations. We select data for the Netherlands, where the survey was annually conducted by Statistics Netherlands, under the coordination of Eurostat. The longitudinal ECHP data for the Netherlands comprise a number of individual records that range from 12,000 to 13,000 per year over the time window 1994–2001, for a total of 100,716 records.

From the original Dutch ECHP panel data, we lose the 1994 wave as information on training was not collected in 1994 in the Netherlands. We focus on prime age and older workers, i.e. workers who are older than 26 and younger than 64 years of age and who are either employee or not employed. Self-employed workers are deemed to be structurally different from employees and therefore are excluded from the sample. We drop observations with missing values in the variables used in the econometric analysis and we drop individuals that are not in the sample for at least three consecutive time periods between 1994 and 2001. The latter restriction is due to the fact that we estimate a dynamic model of order one with unobserved effects. Hence, one time period is lost because of the model dynamics. A further period is lost as we will use initial values to correct for initial conditions induced by the presence of unobserved effects.

After the application of these sample selection criteria, we have an unbalanced panel of 7,257 individuals, for a total of 33,348 individual-year observations, from 1996 until 2001.³ Table 1 clarifies the structure of our data.

³We have an unbalanced panel due to attrition, missing information, and sample renewal for issues of representativeness over time. We assume that attrition and missing information are random. It would have been interesting to use more recent data. Unfortunately, it is not possible to use data the EU database on Statistics on Income and Living Conditions (SILC) as this database does not contain information about training.

Table 1: The Structure of the Unbalanced Panel

Years of observation (the initial year $t = 0$ is not included)	Individual records		Total records	
	Absolute frequencies	Relative frequencies	Absolute frequencies	Relative frequencies
2000–2001	496	.070	992	.030
1999–2000	67	.010	134	.004
1999–2001	407	.057	1,221	.037
1998–1999	36	.006	72	.002
1998–2000	44	.006	132	.004
1998–2001	223	.032	892	.027
1997–1998	53	.009	106	.003
1997–1999	50	.007	150	.005
1997–2000	32	.005	128	.004
1997–2001	248	.035	1,240	.037
1996–1997	514	.073	1,028	.031
1996–1997/2000–2001	41	.006	164	.005
1996–1998	522	.072	1,566	.047
1996–1999	574	.087	2,296	.069
1996–2000	473	.074	2,365	.071
1996–2001	3,477	.451	20,862	.626
Total	$N = 7,257$	1.000	$NT = 33,348$	1.000

We are interested in whether and to what extent the employability of a worker – the probability of remaining employed – is affected by firm-provided training. The non-employment indicator is constructed on the basis of the ILO definition of employment status. It is denoted by y_{it} and it is equal to 1 if individual i is not in the workforce at time t and 0 otherwise. The firm-provided training indicator w_{it} is instead equal to 1 if employee i attended vocational education courses paid or organized by the firm since the beginning of the previous year and 0 otherwise.⁴

Table 2 reports the probabilities of being out of the workforce conditional and unconditional of previous employment situation. The unconditional non-employment probability is 30.8% and it shows a strong persistence, possibly due to individual observed and unobserved heterogeneity: someone not employed at $t - 1$ is almost 20 times as likely not to be employed at t as someone employed at $t - 1$. The non-employment probability is lower for those who attended some firm-provided training in the past than those who did not: 1.9% against 5.5%. Note that the probability of attending firm-provided training courses seems to be strongly affected by the past employment condition. This might be due to individual observed and unobserved heterogeneity but it might also reflect feedback effects going from current shocks in the employment status to future probability of attending firm-provided training.

Table 3 displays the observed transitions between employment positions and, as ex-

⁴We build the non-employment indicator on the basis of variables PE003 and PE004 of the ECHP survey. The firm-provided training indicator is built on the basis of variables PT001 and PT017.

Table 2: Raw Conditional and Unconditional Nonemployment Probabilities

<i>Employment status at t</i>	<i>Employment status at t – 1</i>			Total
	Not employed	Employed with firm-provided training	Employed without firm-provided training	
Not employed	.883	.019	.055	.308
Employed with firm-provided training	.004	.285	.043	.041
Employed without firm-provided training	.113	.696	.902	.651
Total	1.000	1.000	1.000	1.000
Observations	10,243	1,389	21,716	33,348

pected, most of the individuals show a strong persistence in employment. The identification of the effect of training on employees' employability comes from observations out of the diagonal of this transition matrix.

Table 3: Absolute (Relative) Frequencies of Transitions between Labour Market Positions

<i>Employment status at t</i>	<i>Employment status at t – 1</i>			Total
	Not employed	Employed with firm-provided training	Employed without firm-provided training	
Not employed	9,048 (.271)	26 (.001)	1,194 (.036)	10,268 (.308)
Employed with firm-provided training	36 (.001)	396 (.014)	931 (.028)	1,363 (.041)
Employed without firm-provided training	1,159 (.035)	967 (.029)	19,591 (.588)	21,717 (.651)
Total	10,243 (.307)	1,389 (.042)	21,716 (.651)	33,348 (1.000)

Table 4 presents summary statistics of the outcome variables and of the variables used in the specification of the employment equation. We control for gender, education, age, years of potential work experience, health status, number of household components, presence of children in the household (younger than 12 years old), position in the family, and time indicators.⁵ The average age is about 43 years with 18 years of potential working experience. More than 53% of the people in the sample are women, 54% have a secondary degree, and more than 23% do not have a good health situation. On average each household has 3 members, while 35% of the sample has a child younger than 12 years of age in the household. Almost 86% of the people are living in a couple (married or unmarried).

Table 5 shows summary statistics of the covariates entering the training equation for employees. In this case further variables capturing job and employment characteristics are used to explain employees' probability of receiving firm-provided training: contract arrangement, part-time indicator, occupational dummies, job tenure, and sector and firm

⁵In the model specification we also included the interactions between gender and presence of children.

Table 4: Summary Statistics of the Pooled Sample

	Mean	Std. Dev.	Minimum	Maximum
Female	.531	.499	.000	1.000
Education ISCED 5-7	.205	.404	.000	1.000
Education ISCED 3	.536	.499	.000	1.000
Education ISCED 0-2	.259	.438	.000	1.000
Age	43.498	10.047	26.000	64.000
Potential experience (years)	18.102	13.662	.000	52.000
Bad health ^(b)	.232	.422	.000	1.000
Number of household members	3.019	1.287	1.000	8.000
Presence of kids younger than 12	.354	.478	.000	1.000
Individual is cohabiting	.856	.351	.000	1.000
ln(household net income) ^(b)	3.822	1.947	.000	6.661
1996	.168	.374	.000	1.000
1997	.179	.384	.000	1.000
1998	.172	.377	.000	1.000
1999	.169	.375	.000	1.000
2000	.165	.371	.000	1.000
2001	.147	.354	.000	1.000
Observations <i>NT</i>			33,348	
Number of individuals <i>N</i>			7,257	

^(a) We build the health indicator on the basis of variable PH001, which reports self-perceived health. It is equal to one in case of fair, rather bad, or bad health conditions. It is equal to zero in case of either good or very good health conditions.

^(b) The household net income is computed from the variables HI100 and PI100. It does not include the income of the corresponding individual and it is in constant prices (2000 prices). It is deflated by using the Consumer Price Index (CPI), gathered by Statistics Netherlands.

size indicators. About 82% of the employees have a permanent job and more than 30% work on a part-time basis. Almost half of the employees are high-skilled white collar workers, more than 71% work in the service sector, and more than 50% work in firms with more than 100 employees. More than 26% of the workers have a job in the public sector.

3 Econometric Modelling

In this Section we describe a multivariate discrete response model for panel data to investigate whether the employment probability is affected by participation in firm-provided training courses. There are reasons to suspect that the training indicator is a potentially endogenous human capital variable. First, there might be self-selection issues related to unobserved heterogeneity: time-invariant individual characteristics, unobservable by the econometrician, that jointly determine the probability of being at work and participating in training. Innate ability, intelligence, motivations, and labour market attachments are examples of such endowments that, if ignored, may lead to biased parameter estimates (Heckman, 1981; Hyslop, 1999). Second, there might be feedback effects from employ-

Table 5: Summary Statistics of Employees

	Mean	Std. Dev.	Minimum	Maximum
Female	.459	.498	.000	1.000
Education ISCED 5-7	.238	.426	.000	1.000
Education ISCED 3	.538	.499	.000	1.000
Education ISCED 0-2	.223	.416	.000	1.000
Age	41.364	8.902	26.000	64.000
Potential experience (years)	2.441	11.506	.000	52.000
Bad health	.166	.372	.000	1.000
Number of household members	3.055	1.267	1.000	8.000
Presence of kids younger than 12	.375	.484	.000	1.000
Individual is cohabiting	.860	.347	.000	1.000
ln(household net income)	3.785	1.922	.000	6.561
1996	.163	.370	.000	1.000
1997	.174	.379	.000	1.000
1998	.171	.376	.000	1.000
1999	.170	.376	.000	1.000
2000	.170	.376	.000	1.000
2001	.151	.358	.000	1.000
Permanent contract	.818	.386	.000	1.000
Part-time job	.301	.459	.000	1.000
Blue collar worker ^(a)	.257	.437	.000	1.000
Low-skilled white collar worker ^(a)	.246	.431	.000	1.000
High-skilled white collar worker ^(a)	.497	.500	.000	1.000
Agriculture	.012	.110	.000	1.000
Industry	.195	.396	.000	1.000
Services	.711	.453	.000	1.000
Unknown sector	.082	.274	.000	1.000
Public employment	.261	.439	.000	1.000
Unknown job tenure	.140	.347	.000	1.000
Job tenure 0-4 years	.301	.459	.000	1.000
Job tenure 5-9 years	.184	.387	.000	1.000
Job tenure 10-14 years	.127	.332	.000	1.000
Job tenure 15 years or more	.248	.432	.000	1.000
Firm size is not applicable	.145	.353	.000	1.000
Firm size 0-4 employees	.033	.178	.000	1.000
Firm size 5-19 employees	.117	.321	.000	1.000
Firm size 20-49 employees	.108	.310	.000	1.000
Firm size 50-99 employees	.089	.284	.000	1.000
Firm size 100-499 employees	.222	.416	.000	1.000
Firm size 500 employees or more	.286	.452	.000	1.000
Observations <i>NT</i>			23,080	
Number of individuals <i>N</i>			5,609	

^(a) We built the occupational dummies on the basis of variable PE006C. We define as high-skilled white collars those workers who reported to be legislators, senior officers, managers, professionals, technicians, or associate professionals. We define as low-skilled white collars those workers we were clerks, service workers, or shop/market sales workers. We define as blue collars those workers employed as skilled agricultural or fishery workers, craft and related trades workers, plant and machine operators and assemblers, or elementary occupations.

ment status to future training participation, i.e. shocks in the employment status affecting future probabilities of training participation. There are indeed reasons to expect that future participation in a training programme can be correlated to the recent labour market history (Bassi, 1984; Ham and LaLonde, 1996). For instance, individuals with a negative transitory shock in the employment probability can be seen as less reliable and less attached to the labour market and, therefore, employers might be less willing to provide them with training courses. Alternatively, individuals that involuntarily exit employment might change their behaviour and invest in their own human capital.

We use a discrete response unobserved effects model for panel data that can deal with these endogeneity issues. We jointly model the employment status and, in case of employment ($y_{it} = 0$), the firm-provided training participation. The model is designed with a dynamic recursive structure. The current employment status depends on the past employment condition and upon firm-provided training received in the past. Similarly, for those who are at work, the probability of receiving firm-provided training depends on the previous employment condition and past training participation. More in detail, the inter-related dynamics between employment situation and training participation are specified using a panel data bivariate unobserved effects probit model, i.e. for $i = 1, \dots, N$ and $t = 1, \dots, T$

$$y_{it} = 1[y_{it-1}\delta_1 + w_{it-1}\gamma_1 + \mathbf{x}'_{it}\beta_1 + c_{1i} + u_{1it} > 0] \quad (1)$$

$$w_{it} = 1[y_{it-1}\delta_2 + w_{it-1}\gamma_2 + \mathbf{z}'_{it}\beta_2 + c_{2i} + u_{2it} > 0] \quad \text{if } y_{it} = 0, \quad (2)$$

where:

- $1[\cdot]$ is the indicator function;
- \mathbf{x}_{it} is the vector of strictly exogenous covariates explaining the employment status and β_1 is the conformable vector of parameters;
- \mathbf{z}_{it} is the vector of strictly exogenous covariates explaining training participation and β_2 is the conformable vector of parameters;
- (c_{1i}, c_{2i}) is the time-invariant individual heterogeneity characterized by joint distribution with, *a priori*, unrestricted correlation structure;
- u_{1it} and u_{2it} are iid errors with standard normal distribution.

This model is a modified version of the one in Alessie et al. (2004) and similar to that used by Stewart (2007) to analyse the interrelated dynamics of unemployment and low-wage employment and by Picchio (2008) to study the stepping-stone effect of temporary jobs.

Equation (1) shows that in each time period the probability of individual i of being out of the workforce at time t is determined by a vector of observed characteristics, \mathbf{x}_{it} ,

by unobserved heterogeneity, c_{1i} , and by the previous employment situation (employment without training, employment with training, or nonemployment). The previous employment situation is described by the values taken by y_{it-1} , equal to one in case of nonemployment, and by the values taken by w_{it-1} , equal to one in case of employment with firm-provided training. The coefficients δ_1 and γ_1 are of particular interest. The former is the effect of previous nonemployment on the current employability with respect to the case of employment without firm-provided training. The latter is the effect of previous employment with firm-provided training on the current employability with respect to the case of employment without firm-provided training.

For those who are at work, equation (2) describes the process determining the probability of receiving firm-provided training. This is affected by a set of observed characteristics, \mathbf{z}_{it} , by unobservables, c_{2i} , and by past employment situation. The coefficient γ_2 is the effect of past employment with training, rather than without training, on the current probability of receiving training.

Although the u_{1it} and the u_{2it} are assumed iid, the composite error terms will be correlated over time due to the presence of the individual fixed effects c_{1i} and c_{2i} . They will also be correlated across equations due to unrestricted correlation structure between c_{1i} and c_{2i} . In other words, unconditional on c_{1i} and c_{2i} the nonemployment equation is correlated to the training equation, but once we condition on these unobserved factors (and on a set of observed characteristics) the two processes are independent. Note that if the two equations are independent, w_{it-1} is weakly endogenous in the employment equation and equation (1) could be estimated in a univariate framework with predetermined regressors.

3.1 Unobserved Heterogeneity and Initial Conditions

The dynamic unobserved effects probit model in equations (1) and (2) can distinguish between spurious effects determined by unobserved heterogeneity and the true effect of lagged variables (state dependence). However, the presence of unobserved heterogeneity generates two problems that must be faced when estimating such a non-linear model: first, how to get rid of the fixed effects c_{1i} and c_{2i} as it is well known that they cannot be treated as parameters to be estimated due to the incidental parameters problem (e.g. [Heckman, 1981](#)); second, the initial conditions problems that arise in a dynamic model when the initial observations of the outcome variables are correlated to the unobserved heterogeneity.

We solve for these problems by mixing parametric and semiparametric assumptions. First, we allow for dependence between observed and unobserved characteristics by using a [Mundlak \(1978\)](#) version of [Chamberlain's \(1984\)](#) approach. Second, the initial

conditions problem is addressed by using [Wooldridge's \(2005\)](#) approach.⁶ Formally, the parametric specification of the unobserved heterogeneity terms is assumed to be,

$$c_{1i} = \bar{\mathbf{x}}_i' \alpha_1 + y_{i0} \theta_1 + w_{i0} \psi_1 + v_{1i}, \quad (3)$$

$$c_{2i} = \bar{\mathbf{z}}_i' \alpha_2 + y_{i0} \theta_2 + w_{i0} \psi_2 + v_{2i}, \quad (4)$$

where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{z}}_i$ are the individual time averages of respectively \mathbf{x}_{it} and \mathbf{z}_{it} , and y_{i0} and w_{i0} are the realizations of the outcome variables at the date of entry into our sample. The term $\mathbf{v}_i \equiv (v_{1i}, v_{2i})$ is residual unobserved heterogeneity and it is assumed to be independent of observed characteristics. The unobserved time-invariant factors are allowed to be cross-correlated, so as to capture cross-equation correlation. In our preferred specification, we avoid too strict parametric assumptions on the distribution of the random unobserved heterogeneity. We follow [Heckman and Singer \(1984\)](#) and assume that the vector \mathbf{v}_i is a random draw from a discrete distribution function. More in detail, we assume that v_{1i} and v_{2i} have two points of support each with the following four probabilities:

$$\begin{aligned} p^1 &\equiv \Pr(v_1 = v_{1i}^1, v_2 = v_{2i}^1) & p^2 &\equiv \Pr(v_1 = v_{1i}^2, v_2 = v_{2i}^1) \\ p^3 &\equiv \Pr(v_1 = v_{1i}^1, v_2 = v_{2i}^2) & p^4 &\equiv \Pr(v_1 = v_{1i}^2, v_2 = v_{2i}^2) \end{aligned}$$

The probabilities associated to the mass points are specified as logistic transforms:

$$p^m = \frac{\exp(\lambda_m)}{\sum_{r=1}^4 \exp(\lambda_r)} \quad \text{with} \quad \lambda_4 = 0.$$

Note that v_1 and v_2 are independent if and only if $p^1 p^4 = p^2 p^3$ (see [van den Berg and Lindeboom, 1998](#); [van den Berg and Ridder, 1994](#)). This makes it easy to test whether the nonemployment and training equations are independent. Furthermore, it can be shown that the correlation ρ_v between v_1 and v_2 is given by

$$\rho_v = \frac{p^1 p^4 - p^2 p^3}{\sqrt{(p^1 + p^3)(p^2 + p^4)(p^1 + p^2)(p^3 + p^4)}}.$$

⁶An alternative correction of the initial conditions problem is in [Heckman \(1981\)](#) and it is based on a separate formulation of the processes leading to the first realizations of the outcome variables, in order to get an approximation of the conditional distribution of the initial conditions. In this study, we prefer [Wooldridge's \(2005\)](#) approach because the true processes are already ongoing when the first observations are recorded and they are likely to be generated in the same way as later observations. Moreover, [Wooldridge's \(2005\)](#) approach is computationally less demanding.

3.2 The Likelihood Function

Our assumptions with respect to the individual heterogeneity distribution and on the initial conditions allow us to rewrite the model in equations (1) and (2) as

$$y_{it} = 1[y_{it-1}\delta_1 + w_{it-1}\gamma_1 + \mathbf{x}'_{it}\beta_1 + \bar{\mathbf{x}}'_i\alpha_1 + y_{i0}\theta_1 + w_{i0}\psi_1 + v_{1i} + u_{1it} > 0] \quad (5)$$

$$w_{it} = 1[y_{it-1}\delta_2 + w_{it-1}\gamma_2 + \mathbf{z}'_{it}\beta_2 + \bar{\mathbf{z}}'_i\alpha_2 + y_{i0}\theta_2 + w_{i0}\psi_2 + v_{2i} + u_{2it} > 0] \text{ if } y_{it} = 0. \quad (6)$$

Since the unobserved heterogeneity term \mathbf{v}_i is not observed and is a random term from a bivariate distribution, it can be integrated out when the model is estimated by maximum likelihood (ML). The probability masses and the location of the points of support of the discrete unobserved heterogeneity distribution are estimated by ML jointly with all the other parameters. On the basis of model in equations (5) and (6) and the assumptions on the distribution of \mathbf{v}_i , the contribution to the likelihood function of individual i is given by

$$\begin{aligned} \mathcal{L}_i = & \sum_{m=1}^4 p^m \prod_{t=1}^T \Phi \left[(2y_{it} - 1)(y_{it-1}\delta_1 + w_{it-1}\gamma_1 + \mathbf{x}'_{it}\beta_1 + \bar{\mathbf{x}}'_i\alpha_1 + y_{i0}\theta_1 + w_{i0}\psi_1 + v_{1i}^m) \right] \\ & \times \Phi \left[(2w_{it} - 1)(y_{it-1}\delta_2 + w_{it-1}\gamma_2 + \mathbf{z}'_{it}\beta_2 + \bar{\mathbf{z}}'_i\alpha_2 + y_{i0}\theta_2 + w_{i0}\psi_2 + v_{2i}^m) \right]^{(1-y_{it})}, \end{aligned}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. The log-likelihood function is the sum over the sample of the log of the individual likelihood contribution, i.e. $\ell = \sum_{i=1}^N \ln(\mathcal{L}_i)$.⁷

4 Estimation Results

4.1 Dynamic Unobserved Effects Probit

Tables 6 and 7 display the estimation results of four specifications of the dynamic unobserved effects probit model. The first specification is the benchmark model as described in Section 3 and our preferred specification.

In the second specification, job characteristics are not included among the explanatory variables of the training equation. As job characteristics may not be strictly exogenous, we check thereby whether the results are sensitive to their (in)exclusion. As shown by Table 6, the exclusion of job characteristics from the training equation has little effect on the estimation results. Our preferred specification is however the first one, since job

⁷We used the Matlab minimizer *fminunc* with analytic first derivatives to obtain the ML estimates.

characteristics can help to control for time-varying factors (e.g. the quality of the job match) which might be correlated to the probability of receiving firm-provided training and, at the same time, to the probability of keeping the job.

In the third specification, we change the assumptions about the distribution of the unobserved heterogeneity term v_i and we impose a bivariate normal distribution with zero mean and correlation ρ_v , instead of a discrete distribution. We find that specification 1 is to be preferred according to likelihood criteria, e.g. the Akaike Information Criterion (AIC) reported at the bottom of Tables 6 and 7. While in specification 1 we can confidently reject the null hypothesis of independent equations, this is not the case in specification 3. As a matter of fact, the estimation results in specification 4, a single-equation model for employment status, are in line with those of specification 3. This suggests that imposing too strict parametric assumptions on the unobserved heterogeneity term v_i results in model misspecification and biased estimation results.

In the upper panel of Tables 6 and 7 we report usual coefficient estimates. In the second panel we report instead estimated predicted probabilities and average partial effects (APEs) that are of focal interest in this paper. They are aimed at quantifying the size of the effects under analysis. There are different ways in which the marginal effect of y_{it-1} and w_{it-1} on the nonemployment probability can be estimated in a dynamic unobserved effects probit model. At the sample mean of the exogenous regressor (\bar{x}), we define:

- π_1 as the probability of being currently nonemployed conditional on employment without firm-provided training in the previous period;
- π_2 as the probability of being currently nonemployed conditional on employment with firm-provided training in the previous period;
- π_3 as the probability of being currently nonemployed conditional on nonemployment in the previous period.

Consistent estimators of these probabilities are:

$$\hat{\pi}_1 = \frac{1}{N} \sum_{i=1}^N \sum_{m=1,2} \hat{p}^m \Phi(\bar{x}'\hat{\beta}_1 + \bar{x}'_i\hat{\alpha}_1 + y_{i0}\hat{\theta}_1 + w_{i0}\hat{\psi}_1 + \hat{v}_{1i}^m); \quad (7)$$

$$\hat{\pi}_2 = \frac{1}{N} \sum_{i=1}^N \sum_{m=1,2} \hat{p}^m \Phi(\hat{\gamma}_1 + \bar{x}'\hat{\beta}_1 + \bar{x}'_i\hat{\alpha}_1 + y_{i0}\hat{\theta}_1 + w_{i0}\hat{\psi}_1 + \hat{v}_{1i}^m); \quad (8)$$

$$\hat{\pi}_3 = \frac{1}{N} \sum_{i=1}^N \sum_{m=1,2} \hat{p}^m \Phi(\hat{\delta}_1 + \bar{x}'\hat{\beta}_1 + \bar{x}'_i\hat{\alpha}_1 + y_{i0}\hat{\theta}_1 + w_{i0}\hat{\psi}_1 + \hat{v}_{1i}^m). \quad (9)$$

We obtain the APEs by taking the difference between these quantities.⁸ Two APEs are particularly useful for discussion: $\hat{\pi}_2 - \hat{\pi}_1$ and $\hat{\pi}_2 - \hat{\pi}_3$. The former measures the effect on the nonemployment probability of previous employment with firm-provided training rather than without firm-provided training. It is a measure of whether and to what extent firm-provided training boosts employees's chances to be retained in the workforce in the future. The latter is the effect on the nonemployment probability of previous employment with firm-provided training rather than previous nonemployment.

The benchmark model (specification 1 in Table 6) gives an APE of past employment with training, rather than employment without training, of -0.052, statistically significant at the 1% confidence level. An employee with a given set of observed and unobserved characteristics is 5.2 percentage points less likely to be out of the workforce at t if she had been employed with training at $t - 1$ than if she had been employed without training at $t - 1$. It is a quite large effect: the nonemployment probability decreases from 32.6 percentage points to 27.4 percentage points, i.e. by 16%. As shown in Table 2, the corresponding figure from raw data is 65.5%. This points out that more than three fourths of the raw effect of firm-provided training on employability is spurious and determined by observed and unobserved characteristics.

Three other issues are worth mentioning. First, it is clear from the statistics in Table 2 that there is a sizeable persistence in nonemployment: people who were out of the workforce at $t - 1$ are about 46 (16) times more likely to be nonemployed at t than people who were employed with (without) firm-provided training at $t - 1$. These figures become much smaller but still sizeable when we get rid of spurious state dependence induced by individual heterogeneity (observed and unobserved): an individual is “only” about twice as likely to be nonemployed at t if she had been out of the workforce at $t - 1$ as if she had been employed with or without training at $t - 1$.⁹

Second, the estimated correlation between the unobserved determinants of nonemployment and training is negative, -0.665, and highly significant. Unobserved characteristics, like ability, motivation, and attachment to work, affect positively the probability of receiving firm-provided training and negatively the probability of being out of the workforce.

Third, in specification 3 we fail to capture the cross-equation correlation induced by unobserved heterogeneity, probably because of too strict parametric assumptions on its

⁸Standard errors of the predicted probabilities and of the APEs are estimated by bootstrapping the results (individual-cluster bootstrap with replacement).

⁹More in detail, with raw data the nonemployment probability decreases from 88.3% to 1.9% in case of employment with training and to 5.5% in case of employment without training. Once we net out spurious state dependence, the nonemployment probability decreases from 52% to 27.4% in case of employment with training and to 32.6% in case of employment without training.

Table 6: Dynamic Unobserved Effects Probit Model

Variable	Specification 1						Specification 2					
	Nonemployment equation			Firm-provided training equation			Nonemployment equation			Firm-provided training equation		
	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.
Nonemployment _{t-1}	1.715	***	.033	-.093		.101	1.717	***	.033	-.156		.099
Firm-provided training _{t-1}	-.546	***	.115	.738	***	.060	-.281	**	.111	.756	***	.059
Female	.205	***	.044	-.010		.051	.211	***	.044	-.056		.045
<i>Education - Reference: Education ISCED 0-2</i>												
Education ISCED 5-7	-.270	***	.046	-.113	*	.059	-.266	***	.046	.086		.053
Education ISCED 3	-.206	***	.035	.001		.046	-.205	***	.035	.057		.044
Age/10	-1.732	***	.154	.171		.259	-1.759	***	.154	.243		.254
Age squared/1000	2.408	***	.177	-.418		.336	2.441	***	.177	-.570	*	.332
Work experience/10	1.285	***	.185	-.279		.340	1.285	***	.185	-.635	**	.321
Work exper. squared/1000	.073	*	.037	-.013		.061	.077	**	.037	-.007		.059
Bad health	.218	***	.042	.062		.062	.218	***	.042	.050		.060
No. household members	.069	*	.040	.006		.051	.071	*	.040	-.030		.050
Kids<12 years	.227	*	.127	.012		.108	.234	*	.126	.001		.104
Kids<12 years*Female	.165		.146	-.311	**	.155	.156		.146	-.433	***	.150
Living in a couple	-.493	***	.136	.049		.139	-.498	***	.136	.098		.137
ln(household net income)	.072	***	.015	-.040	**	.020	.071	***	.015	-.041	**	.019
Permanent contract	-		-	.072		.062	-		-	-		-
Part-time job	-		-	-.310	***	.083	-		-	-		-
Public employment	-		-	.114		.071	-		-	-		-
<i>Occupation indicators – Reference: High-skilled white collar worker</i>												
Blue collar	-		-	-.215	**	.085	-		-	-		-
Low-skilled white collar	-		-	-.202	***	.063	-		-	-		-
<i>Sectoral indicators – Reference: Services</i>												
Agriculture	-		-	.041		.384	-		-	-		-
Industry	-		-	-.224	***	.086	-		-	-		-
Unknown sector	-		-	-.322	***	.081	-		-	-		-
<i>Job tenure indicators – Reference: 0-4 years</i>												
Unknown	-		-	-.312	**	.144	-		-	-		-
5-9 years	-		-	-.269	***	.065	-		-	-		-
10-14 years	-		-	-.227	**	.091	-		-	-		-
15 years or more	-		-	-.267	**	.121	-		-	-		-
<i>Initial conditions and unobserved heterogeneity support points</i>												
Nonemployment ₀	1.384	***	.075	.505	***	.115	1.407	***	.077	.004		.085
Firm-provided training ₀	-.059		.105	.357	***	.069	-.144		.102	.449	***	.071
\hat{v}_j^1	1.998	***	6.571	-2.199	***	.482	2.042	***	.304	-2.067	***	.477
\hat{v}_j^2	.539	*	1.812	-1.426	***	.497	.560	*	.297	-1.284	***	.472
$\hat{\rho}_v$	-.665	***	.208				.114		.144			
Predicted probability $\hat{\pi}_1$.326	***	.015	-		-	.325	***	.017	-		-
Predicted probability $\hat{\pi}_2$.274	***	.021	-		-	.297	***	.022	-		-
Predicted probability $\hat{\pi}_3$.520	***	.009	-		-	.518	***	.009	-		-
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.052	***	.012	-		-	-.028	***	.012	-		-
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.247	***	.025	-		-	-.221	***	.025	-		-
NT (N)			33,348 (7,275)						33,348 (7,275)			
Log-likelihood			-11,812.7						-11,946.7			
No. of parameters			103						67			
Pseudo-R ²			.542						.536			
Test of independent eq.			$\chi^2=5.32$						$\chi^2=.94$			
			p-value=.021						p-value=.332			

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Time dummies, form size indicators, and individual time average of the time-varying covariates are included in the model specification but not reported for the sake of brevity. The standars errors are robust to within-individual correlation and heteroskedasticity. The standard errors of the predicted probabilities, APEs, and ρ_v are obtained by bootstrapping the results 399 times (individual-cluster bootstrap with replacement).

Table 7: Dynamic Unobserved Effects Probit Model

Variable	Specification 3						Specification 4			
	Nonemployment equation			Firm-provided training equation			Nonemployment equation			
	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.	
Nonemployment _{t-1}	2.155	***	.035	-.254		.240	2.155	***	.035	
Firm-provided training _{t-1}	-.311	***	.086	.925	***	.045	-.304	***	.085	
Female	.143	***	.033	-.020		.046	.144	***	.033	
<i>Education - Reference: Education ISCED 0-2</i>										
Education ISCED 5-7	-.199	***	.037	-.087	*	.048	-.199	***	.038	
Education ISCED 3	-.143	***	.029	.016		.039	-.143	***	.029	
Age/10	-1.299	***	.135	.292		.224	-1.304	***	.135	
Age squared/1000	1.827	***	.161	-.592	**	.290	1.832	***	.160	
Work experience/10	1.300	***	.205	-.372		.276	1.296	***	.205	
Work exper. squared/1000	-.048		.031	-.016		.051	-.048		.031	
Bad health	.182	***	.041	.050		.058	.181	***	.042	
No. household members	.051		.035	.001		.045	.050		.035	
Kids<12 years	.171	**	.083	.000		.095	.173	**	.084	
Kids<12 years*Female	.123		.111	-.309	**	.137	.120		.111	
Living in a couple	-.398	***	.101	.080		.111	-.398	***	.101	
ln(household net income)	.064	***	.015	-.038	**	.017	.064	***	.015	
Permanent contract	-		-	.109		.071	-		-	
Part-time job	-		-	-.275	***	.077	-		-	
Public employment	-		-	.120	*	.070	-		-	
<i>Occupation indicators – Reference: High-skilled white collar worker</i>										
Blue collar	-		-	-.190	**	.074	-		-	
Low-skilled white collar	-		-	-.180	***	.065	-		-	
<i>Sectoral indicators – Reference: Services</i>										
Agriculture	-		-	.070		.320	-		-	
Industry	-		-	-.188	**	.087	-		-	
Unknown sector	-		-	-.289	***	.078	-		-	
<i>Job tenure indicators – Reference: 0-4 years</i>										
Unknown	-		-	-.223	*	.134	-		-	
5-9 years	-		-	-.244	***	.058	-		-	
10-14 years	-		-	-.187	**	.084	-		-	
15 years or more	-		-	-.205	*	.112	-		-	
Constant	.597	**	.266	-2.048	***	.407	.607	**	.266	
<i>Initial conditions and unobserved heterogeneity correlation</i>										
Nonemployment ₀	.560	***	.036	.140	*	.074	.559	***	.036	
Firm-provided training ₀	-.097		.084	.248	***	.053	-.096		.084	
$\hat{\rho}_v$	-.230		.235				-		-	
Predicted probability $\hat{\pi}_1$.260	***	.	-		-	.260	***	.021	
Predicted probability $\hat{\pi}_2$.221	***	.	-		-	.221	***	.026	
Predicted probability $\hat{\pi}_3$.603	***	.	-		-	.604	***	.016	
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.040	***	.	-		-	-.039	***	.011	
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.383	***	.	-		-	-.383	***	.040	
NT (N)			33,348 (7,275)				33,348 (7,275)			
Log-likelihood			-11,937.8				-7,441.8			
No. of parameters			99				31			
Pseudo- R^2			.537				.639			
Test of independent eq.			$\chi^2=.36$				$p\text{-value}=.843$			

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Time dummies, firm size indicators, and individual time average of the time-varying covariates are included in the model specification but not reported for the sake of brevity. The standard errors are robust to within-individual correlation and heteroskedasticity. The standard errors of the predicted probabilities and APEs are obtained by bootstrapping the results 399 times (individual-cluster bootstrap with replacement).

distribution. The state dependence of nonemployment is overestimated: the predicted probability \hat{p}_3 and the estimated coefficient of the lagged nonemployment indicator are indeed much larger than those in specification 1. The APE of employment with training instead of employment without training is equal to -0.040 and, therefore, it is slightly overestimated by the failure in capturing cross-equation correlation.

Looking at the impact of exogenous variables on the nonemployment probability, women are less likely to be at work. The profile of the age and nonemployment probability has an inverted U-shape while, *ceteris paribus*, the probability of being out of the workforce increases with potential work experience. Higher educated people and those living in a couple are more likely to be at work. Those with health problems or high household income are less likely to be employed.

With regard to the firm-provided training equation, there is evidence of state dependence: people getting firm-provided training at $t - 1$ are more likely to receive firm-provided training at t . It is interesting to note that people not employed at $t - 1$ are as likely to receive training at t as those who were at work without training at $t - 1$. Higher educated workers are less likely to receive training, but the coefficient is barely significant. There is therefore some evidence that education and firm-provided training are not complementary assets, in contrast to the findings in [Blundell et al. \(1999\)](#) for the US and the UK. The training probability is lower for women with young kids and increasing with the household net income, probably due to less attachment to the labour market and/or more family commitment. Part-time workers are less likely to get firm-provided training. High skilled white collar workers are more likely to get training, suggesting that tasks and human capital formation are complementary assets. Finally, firm-provided training is more present in the services sector, among public employees, and among newly hired workers.¹⁰

4.2 Retaining Older Workers

The workforce is ageing in many industrialized countries. The ageing of the workforce might be caused, in addition to demographic trends, also by the fading out of early retirement programs and by changes in the pension system like changes in the earliest possible or mandatory retirement age. In the 1980s and in the first half of the 1990s, the Netherlands had one of the lowest employment rates of elderly among the European countries. For example, in 1992 the Dutch employment rate of persons aged 55 to 64 was 28.7%

¹⁰Firm size indicators are included in the specification of the training equation (and not reported in Tables 6 and 7) but they are not significantly different from zero.

against an European average of 39.1%.¹¹ Given the economic stagnation in that period, the low employment rates of older workers were not seen as a major problem, whereas the high youth unemployment was thought of being problematic. Early retirement programs were promoted with the aim of giving a contribution to the employment of new entrants in the labour market. However, with the ageing of the population and the resulting pressures on the pension system, the Dutch early retirement system was no longer sustainable. A series of policy reforms were introduced with the aim of reducing the generosity of the early retirement schemes and creating incentives for postponing retirement. Hence, the response to population ageing was based on increasing labour supply and delaying retirement.¹² By 2009, the employment rate of older workers had increased to 52.6%, larger than the European average but still lower than the OECD average.

Given the ageing of the workforce, the labour market position of older workers is cause for increasing concern. If employed, their position is usually fine as they are not very likely to be dismissed. As a matter of fact, older workers are well-protected by seniority rules and employment protection legislation. Nevertheless, if older workers lose their job, they find it very hard to find a new one. [Gielen and van Ours \(2006\)](#) show that cyclical adjustments of the workforce in the Netherlands occur partly through fluctuations in separations for older workers. These separations are likely to be a one-way street out of the labour force into long-term unemployment.¹³ Employers are indeed reluctant to hire an older worker because of the pay-productivity gap and because of the possible obsolescence of general human capital.¹⁴ As training can refresh general human capital, avoid its obsolescence, and increase workers' productivity, it can be a channel through which retirement can be postponed and employability of the older workers increased.

In this Section, we focus on the effect of firm-provided training on employability by allowing this effect to be heterogeneous across three age categories: 26–35, 36–49, and 50–64. The corresponding indicator variables are interacted with the lags of the nonemployment indicator and of the firm-provided training indicator. The benchmark model is augmented by these interactions and by the age indicators and re-estimated.

Table 8 reports the estimation results of the coefficients and APEs of primary interest.

¹¹These figures are available in the Eurostat webpage http://stats.oecd.org/Index.aspx?DataSetCode=LFS_SEXAGE_I_R.

¹²Empirical studies found that the Dutch reforms had a positive effect on the labour force participation of older workers ([Euwals et al., 2010](#)).

¹³[Gielen and van Ours \(2006\)](#) suggest that training of older workers in public training programs would help them to acquire new skills and to adapt to new demands, such that these workers are more likely to retain their jobs.

¹⁴See the special issue of *De Economist* on “Ageing Workforces” ([Vandenberghe, 2011](#)) for recent studies on the age-productivity-pay nexus and [van Ours and Stoeldraijer \(2011\)](#) for evidence on the effect of age on productivity and wages in the Netherlands.

Table 8: Dynamic Unobserved Effects Probit Model with Age Interactions

Variable	Nonemployment equation		Firm-provided training equation			
	Coeff.	S.E.	Coeff.	S.E.		
<i>Reference: employed without firm-provided training and age within [26, 35]</i>						
Nonemployment $_{t-1}$	1.405	***	.058	-.083	.143	
Firm-provided training $_{t-1}$	-.547	***	.168	.766	***	.078
Nonemployment $_{t-1}$ *Age [36, 49]	.346	***	.063	-.023		.187
Nonemployment $_{t-1}$ *Age [50, 64]	.691	***	.073	-.104		.350
Firm-provided training $_{t-1}$ *Age [36, 49]	-.062		.218	-.008		.095
Firm-provided training $_{t-1}$ *Age [50, 64]	-.042		.287	-.372	**	.179
Age [36, 49]	-.102		.067	-.116		.077
Age [50, 64]	-.509	***	.105	-.046		.133
<i>Predictions and APEs if age [26, 35]</i>						
Predicted probability $\hat{\pi}_1$.339	***	.018	–	–	–
Predicted probability $\hat{\pi}_2$.283	***	.027	–	–	–
Predicted probability $\hat{\pi}_3$.502	***	.011	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.055	***	.017	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.218	***	.028	–	–	–
<i>Predictions and APEs if age [36, 49]</i>						
Predicted probability $\hat{\pi}_1$.328	***	.018	–	–	–
Predicted probability $\hat{\pi}_2$.268	***	.028	–	–	–
Predicted probability $\hat{\pi}_3$.532	***	.011	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.060	***	.018	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.265	***	.033	–	–	–
<i>Predictions and APEs if age [50, 64]</i>						
Predicted probability $\hat{\pi}_1$.288	***	.021	–	–	–
Predicted probability $\hat{\pi}_2$.228	***	.040	–	–	–
Predicted probability $\hat{\pi}_3$.525	***	.013	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.060	*	.032	–	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.297	***	.044	–	–	–
NT (N)			33,348 (7,257)			
Log-likelihood			-11,760.6			
No. of parameters			115			
Pseudo- R^2			.544			

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. All the variables included in the benchmark specification are also included here: the corresponding estimated coefficients are not reported for the sake of brevity. The standard errors are robust to within-individual correlation and heteroskedasticity. The standard errors of the predicted probabilities and APEs are obtained by bootstrapping the results 399 times (individual-cluster bootstrap with replacement).

The coefficient of the interactions between the lag training indicator and the age categories are (jointly) not significantly different from zero. This means that firm-provided training is able to reduce the future probability of being out of the workforce for younger workers as well as for older workers. Note also that the interactions between lag non-employment status and age categories are significantly different from zero and point out that older workers not employed at $t - 1$ are more like to be not employed at t than prime aged and young workers. This suggests that once older workers lose their jobs, they are less likely to find a new one. The coefficient of the indicator for older workers is instead significantly negative: older individuals are less likely to lose their jobs and therefore to be out of the workforce.

The estimation of the APEs at the sample means of the other variables confirm that firm-provided training reduces the probability of being not employed with the same magnitude over age classes. An employee with a given set of observed and unobserved characteristics and in the age range 50–64 is 6 percentage points less likely to be out of the workforce at t if she had been employed with training at $t - 1$ than if she had been employed without training at $t - 1$. This figure is equal to that for employees in the age range 36–49 and slightly bigger in size than that of young workers. There is evidence therefore that firm-provided training substantially increases the employability of older workers as well as the employability of young workers. Firm-provided training might be an important tool to lighten the burden of population ageing on the pension system in the Netherlands.

To some extent training is endogenous to retirement institutions. In 2006 in the public sector in the Netherlands pre-pension plans for every worker born after December 31, 1949 were abolished. To receive the same pension benefits the younger cohort has to postpone retirement for about 13 months. [Montizaan et al. \(2010\)](#) show that this change in future pension benefits had an effect on the expected retirement age and, through this, a positive effect on workers' training participation. We show that this is rational to do since training leads to retaining of jobs. To retain employability of older workers age-specific subsidies to stimulate job training might be used or alternatively age-specific layoff taxes may be introduced.¹⁵ The first type of policy would it make it more attractive for employers to train older workers thus increasing the likelihood that they retain their employment. The second type of policy would make it more expensive for employers to fire older workers thus making it more attractive to train these workers and thereby

¹⁵[Schnalzenberger and Winter-Ebmer \(2009\)](#) show that an age-specific firing tax affected the labour market position of older workers in Austria. Employers had to pay a tax of up to 170% of the gross monthly income when they gave notice to a worker age 50 or more. This tax caused a substantial reduction in layoffs for older workers.

increasing the likelihood that they retain their employability.

4.3 Further Robustness Checks

We perform two further sensitivity analyses to assess whether our estimates are robust to misspecification: i) due to omitting information about individuals who might have attended vocational training courses *not* provided by the firm; ii) of the dynamics.

With regard to the former, the problem might arise as an omitted time-varying variable indicating whether the employee has undertaken training courses not provided by the firm is very likely to be correlated to the participation to a firm-provided training and, at the same time, to the future employment status. To assess whether this might be a problem, we build an indicator variable q_{it} equal to 1 if employee i attended vocational education courses which were not paid or organized by the firm since the beginning of the previous year and 0 otherwise.¹⁶ The incidence of training not provided by the firm is equal to 3.5% among the employees of our sample. Firstly, we include the variable q_{it} in the model specification as an exogenous variable and then as a predetermined variable (weak exogenous). In both cases, we find estimation results of the quantities of interest that are very much in line with those of the benchmark model. Secondly, we jointly model the process determining training not provided by the firm and the other two equations of the benchmark model, yielding the following simultaneous three-equation model

$$\begin{aligned} y_{it} &= 1[y_{it-1}\delta_1 + w_{it-1}\gamma_1 + q_{it-1}\lambda_1 + \mathbf{x}'_{it}\beta_1 + \bar{\mathbf{x}}'_i\alpha_1 + y_{i0}\theta_1 + w_{i0}\psi_1 + q_{i0}\varphi_1 + v_{1i} + u_{1it} > 0] \\ w_{it} &= 1[y_{it-1}\delta_2 + w_{it-1}\gamma_2 + q_{it-1}\lambda_2 + \mathbf{z}'_{it}\beta_2 + \bar{\mathbf{z}}'_i\alpha_2 + y_{i0}\theta_2 + w_{i0}\psi_2 + q_{i0}\varphi_2 + v_{2i} + u_{2it} > 0] \text{ if } y_{it} = 0 \\ q_{it} &= 1[y_{it-1}\delta_3 + w_{it-1}\gamma_3 + q_{it-1}\lambda_3 + \mathbf{z}'_{it}\beta_3 + \bar{\mathbf{z}}'_i\alpha_3 + y_{i0}\theta_3 + w_{i0}\psi_3 + q_{i0}\varphi_3 + \kappa v_{2i} + u_{3it} > 0] \text{ if } y_{it} = 0, \end{aligned}$$

where κ is the loading factor determining, together with v_{2i} , the points of support of the equation of training not provided by firms. The loading factor is used to simplify the specification of the unobserved heterogeneity distribution and reduce the computational difficulty in estimating the model. The likelihood function of the benchmark model can be trivially extended to the three-equation case.

Table 9 reports the estimation results of the three-equation model. All the estimates of the nonemployment equation and of the firm-provided equation are in line with those of the benchmark model reported in Table 6. Looking at the equation for training not provided by the firm, it is noted that the past employment status and the past training matter: people who were nonemployed in the past are more likely to train in the future than those

¹⁶The indicator for vocational training not provided by the firm is built on the basis of variables PT001, PT008, and PT017 of the ECHP data.

Table 9: Dynamic Unobserved Effects Probit Model with 3 Simultaneous Equations

Variable	Nonemployment equation		Firm-provided training equation		Other vocational training equation	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Nonemployment _{t-1}	1.725 ***	.034	-.065	.102	.196	** .095
Firm-provided training _{t-1}	-.479 ***	.116	.779 ***	.060	.201	*** .076
Other vocational training _{t-1}	.244 ***	.093	.096	.089	1.386 ***	.061
Female	.208 ***	.044	-.009	.050	.044	.056
<i>Education - Reference: Education ISCED 0-2</i>						
Education ISCED 5-7	-.269 ***	.046	-.102 *	.058	-.027	.062
Education ISCED 3	-.206 ***	.035	.008	.045	-.023	.049
Age/10	-1.748 ***	.155	.395	.252	-.271	.187
Age squared/1000	2.431 ***	.179	-.697 **	.326	.194	.247
Work experience/10	1.290 ***	.186	-.350	.343	.156	.400
Work exper. squared/1000	.073 *	.038	-.003	.061	-.006	.075
Bad health	.218 ***	.042	.060	.062	-.130	.080
No. household members	.072 *	.040	.003	.052	-.106	.067
Kids<12 years	.229 *	.128	.012	.110	.160	.170
Kids<12 years*Female	.165	.147	-.315 **	.156	-.126	.223
Living in a couple	-.490 ***	.138	.055	.140	.051	.146
ln(household net income)	.071 ***	.015	-.039 **	.020	-.001	.023
Permanent contract	-	-	.083	.062	-.103	.079
Part-time job	-	-	-.298 ***	.083	.245	*** .085
Public employment	-	-	.116	.072	-.148	.092
<i>Occupation indicators - Reference: High-skilled white collar worker</i>						
Blue collar	-	-	-.205 **	.086	-.253	** .106
Low-skilled white collar	-	-	-.193 ***	.064	.040	.076
<i>Sectoral indicators - Reference: Services</i>						
Agriculture	-	-	.050	.381	.609	** .305
Industry	-	-	-.218 **	.086	-.030	.117
Unknown sector	-	-	-.321 ***	.082	-.152	.105
<i>Job tenure indicators - Reference: 0-4 years</i>						
Unknown	-	-	-.289 **	.144	-.009	.147
5-9 years	-	-	-.259 ***	.066	.077	.086
10-14 years	-	-	-.213 **	.092	.182	.144
15 years or more	-	-	-.245 **	.122	.230	.195
<i>Initial conditions and unobserved heterogeneity support points</i>						
Nonemployment ₀	1.389 ***	.077	.417 ***	.111	.119	.086
Firm-provided training ₀	-.083	.107	.348 ***	.067	.138	* .083
Other vocational training ₀	.004 **	.104	.166 **	.081	.363 ***	.069
\hat{v}_j^1	2.005 ***	.307	-2.552 ***	.475	-	-
\hat{v}_j^2	.540 *	.300	-1.803 ***	.509	-	-
$\hat{\rho}_v$	-.528 *	.293	-	-	.314	** .130
$\hat{\kappa}$	-	-	-	-	-	-
Predicted probability $\hat{\pi}_1$.324 ***	.015	-	-	-	-
Predicted probability $\hat{\pi}_2$.279 ***	.022	-	-	-	-
Predicted probability $\hat{\pi}_3$.520 ***	.009	-	-	-	-
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.045 ***	.015	-	-	-	-
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.241 ***	.025	-	-	-	-
NT (N)	33,348 (7,275)					
Log-likelihood	-14,535.4					
No. of parameters	176					
Pseudo-R ²	.501					
Test of independent eq.	$\chi^2_1=224.98$ p-value=.000					

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Time dummies, firm size indicators, and individual time average of the time-varying covariates are included in the model specification but not reported for the sake of brevity. The standard errors are robust to within-individual correlation and heteroskedasticity. The standard errors of the predicted probabilities and APEs are obtained by bootstrapping the results 239 times (individual-cluster bootstrap with replacement).

who were employed without any form of training. This type of “catch-up” response is in line with the empirical evidence in [Mroz and Savage \(2006\)](#), who found that recent unemployment spells has a significant positive effect on whether a young man trains today in the US. As pointed out by [Mroz and Savage \(2006\)](#), the “catch-up” response is predicted by the economic theory: people with a recent nonemployment spell enter the current period with a lower stock of human capital than otherwise identical individuals. When they re-enter the workforce, they dynamically re-optimize their investments in human capital and their new optimal time-path of human capital investments might lie above the one of otherwise identical individuals who did not suffer the employment shock and the loss of human capital.

Table 10: Dynamic Unobserved Effects Probit Model with Lag of Order Two

Variable	Nonemployment equation		Firm-provided training equation	
	Coeff.	S.E.	Coeff.	S.E.
Nonemployment $_{t-1}$	1.884 ***	.036	-.114	.126
Nonemployment $_{t-2}$.769 ***	.042	.154	.108
Firm-provided training $_{t-1}$	-.466 ***	.115	.760 ***	.060
<i>Predictions and APEs if the individual was employed at $t - 2$</i>				
Predicted probability $\hat{\pi}_1$.232 ***	.030	–	–
Predicted probability $\hat{\pi}_2$.172 ***	.035	–	–
Predicted probability $\hat{\pi}_3$.533 ***	.015	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.060 ***	.016	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.361 ***	.046	–	–
<i>Predictions and APEs if the individual was not employed at $t - 2$</i>				
Predicted probability $\hat{\pi}_1$.343 ***	.019	–	–
Predicted probability $\hat{\pi}_2$.274 ***	.031	–	–
Predicted probability $\hat{\pi}_3$.688 ***	.029	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_1$	-.068 ***	.019	–	–
APE: $\hat{\pi}_2 - \hat{\pi}_3$	-.414 ***	.056	–	–
NT (N)	33,348 (7,275)			
Log-likelihood	-8,999.0			
No. of parameters	103			
Pseudo- R^2	.550			

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. All the variables included in the benchmark specification are also included here: the corresponding estimated coefficients are not reported for the sake of brevity. The standars errors are robust to within-individual correlation and heteroskedasticity. The standard errors of the predicted probabilities and APEs are obtained by bootstrapping the results 399 times (individual-cluster bootstrap with replacement).

In a second sensitivity analysis, we check whether our findings are robust to the misspecification of the dynamics. We set up a dynamic model where employment status and firm-provided training enter up to the lag of order two. We find that the lag of order two of the firm-provided training indicator does not significantly affect the nonemployment probability (p -value equal to 0.134) and thereby removed from the model specification.

Table 10 reports the estimation results of the lagged variables. The sample size is now smaller: one more time period is lost as a consequence of the dynamic of higher order. A nonemployment event at $t - 2$ reinforces the positive effect of a nonemployment event at $t - 1$ on the probability of being out of the workforce at t . The estimated coefficient of lagged training is qualitatively in line with the one of the benchmark model. We estimated the APEs by conditioning on the nonemployment status at $t - 2$. In case of employment at $t - 2$, the APE of working with training at $t - 1$ rather than working without training at $t - 1$ is of -6 percentage points in the probability of being nonemployed (-26%). If not at work at $t - 2$, the APE is equal to -6.8 percentage points (-20%). This suggests that the estimated APE $\hat{\pi}_2 - \hat{\pi}_1$ of our benchmark model is not sensitive to the dynamic specification and, if any, it suffers from an upward bias. Note however that when we take the model with a dynamic of higher order, we lose observations and we restrict the analysis to individuals that are in the panel for at least four consecutive waves. This makes the assumption of no attrition less likely to hold. We therefore prefer to stick to the specification with a dynamic of order one.

5 Conclusions

This paper studies the relationship between on-the-job training and employability in the Netherlands. In our analysis we disentangle the true effect of training from the spurious effect that might be induced by self-selection of non-random individuals into training participation. We find that firm-provided training significantly improves employment prospects. For prime age workers who generally have a strong labour market position, in the sense that after job loss they find a new job quite easily, this relationship may be of limited interest. However, we also investigate whether for older workers training leads to higher employability. We find that older workers who receive on-the-job training are more likely to keep employment.

In many countries the labour market position of older workers is cause for concern. Older workers' job separations are often a one-way street out of the labour force and into long-term unemployment. This is a reason for concern since demographic trends are causing an ageing of the workforce. Therefore, improving the employment position of older workers is very important from a policy point of view. Our research findings suggest that on-the-job training may be an important instrument to achieve this goal. Our research does not provide direct evidence on how to stimulate on-the-job training. Nevertheless we suggest that, in order to retain employability of older workers, age-specific subsidies to stimulate job training might be used or, alternatively, age-specific layoff taxes may be

introduced. The first type of policy would make it more attractive for employers to train older workers, thus increasing the likelihood that they retain their employability. The second type of policy would make it more expensive for employers to fire older workers, thus making it more attractive to train these workers and thereby increasing the likelihood that their employment prospects improve.

References

- Alessie, R., S. Hochguertel, and A. van Soest**, “Ownership of Stocks and Mutual Funds: A Panel Data Analysis,” *Review of Economics and Statistics*, 2004, 86 (3), 783–796.
- Andr n, T. and D. Andr n**, “Assessing the Employment Effects of Vocational Training Using a One-Factor Model,” *Applied Economics*, 2006, 38 (21), 2469–2486.
- Barrett, A. and P.J. O’Connell**, “Does Training Generally Work? The Returns to In-Company Training,” *Industrial and Labor Relations Review*, 2001, 54 (3), 647–662.
- Bartel, A.P.**, “Productivity Gains From the Implementation of Employee Training Programs,” *Industrial Relations*, 1994, 33 (4), 411–425.
- , “Training, Wage Growth, and Job Performance: Evidence from a Company Database,” *Journal of Labor Economics*, 1995, 13 (3), 401–425.
- Bassi, L.J.**, “Estimating the Effect of Training Programs with Non-Random Selection,” *Review of Economics and Statistics*, 1984, 66 (1), 36–43.
- Blundell, R., L. Dearden, C. Meghir, and B. Sianesi**, “Human Capital Investments: The Returns from Education and Training to the Individual, the Firm, and the Economy,” *Fiscal Studies*, 1999, 20 (1), 1–23.
- Bonnal, L., D. Foug re, and A. S randon**, “Evaluating the Impact of French Employment Policies on Individual Labour Market Histories,” *Review of Economic Studies*, 1997, 64 (6), 683–713.
- Boone, J. and J.C. van Ours**, “Bringing Unemployed Back to Work: Effective Active Labor Market Policies,” *De Economist*, 2009, 157, 293–313.
- Chamberlain, G.**, “Panel Data,” in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North Holland, 1984, chapter 22, pp. 1248–1318.
- Conti, G.**, “Training, Productivity and Wages in Italy,” *Labour Economics*, 2005, 12 (4), 557–576.
- Cr pon, B., M. Ferracci, and D. Foug re**, “Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?,” 2007. IZA Discussion Paper No. 3215, Bonn.
- Dearden, L., H. Reed, and J. Van Reenen**, “The Impact of Training on Productivity and Wages: Evidence from British Panel Data,” *Oxford Bulletin of Economics and Statistics*, 2006, 68 (4), 397–421.
- Euwals, R., D. van Vuuren, and R. Wolthoff**, “Early Retirement Behaviour in the Netherlands: Evidence from a Policy Reform,” *De Economist*, 3 2010, 158, 209–236.

- Gerfin, M. and M. Lechner**, “A Microeconomic Evaluation of the Active Labour Market Policy in Switzerland,” *Economic Journal*, 2002, 112 (482), 854–893.
- Gielen, A.C. and J.C. van Ours**, “Age-specific Cyclical Effects in Job Reallocation and Labor Mobility,” *Labour Economics*, 2006, 13 (4), 493–504.
- Gritz, R.M.**, “The Impact of Training on the Frequency and Duration of Employment,” *Journal of Econometrics*, 1993, 57 (1), 21–51.
- Ham, J.C. and R.J. LaLonde**, “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training,” *Econometrica*, 1996, 64 (1), 175–205.
- Heckman, J.J.**, “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 with Econometric Applications*, Cambridge: MIT Press, 1981, chapter 4, pp. 179–195.
- **and B. Singer**, “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data,” *Econometrica*, 1984, 52 (2), 271–320.
- **, LaLonde R.J., and Smith J.A.**, “The Economics and Econometrics of Active Labor Market Programs,” in Ashenfelter O.C. and Card D., eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: Elsevier, 1999, chapter 31, pp. 1865–2097.
- Hyslop, D.R.**, “State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women,” *Econometrica*, 1999, 67 (6), 1255–1294.
- Jones, M.K., R.J. Jones, P.L. Latreille, and P.J. Sloane**, “Training, Job Satisfaction, and Workplace Performance in Britain: Evidence from WERS 2004,” *Labour*, 2009, 23 (Special Issue), 139–175.
- Kluve, J.**, “The Effectiveness of European Active Labour Market Programmes,” *Labour Economics*, 2010, 17 (6), 904–918.
- Lalive, R., J.C. van Ours, and F. Zweimüller**, “The Impact of Active Labour Market Programmes on the Duration of Unemployment in Switzerland,” *Economic Journal*, 2008, 118 (525), 235–257.
- Lechner, M., R. Miquel, and C. Wunsh**, “The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany After Unification,” *German Economic Review*, 2008, 8 (4), 468–509.
- Montizaan, R., F. Corvers, and A. De Grip**, “The effects of pension rights and retirement age on training participation: evidence from a natural experiment,” *Labour Economics*, 2010, 17 (1), 240–247.
- Mroz, T.A. and T.H. Savage**, “The Long-Term Effects of Youth Unemployment,” *Journal of Human Resources*, 2006, 41 (2), 259–293.
- Mundlak, Y.**, “On the Pooling of Time Series and Cross Section Data,” *Econometrica*, 1978, 46 (1), 69–85.
- OECD**, *Employment Outlook* 2010.
- Pavlopoulos, D., R. Muffels, and J.K. Vermunt**, “Training and Low-pay Mobility: The Case of the UK and the Netherlands,” *Labour*, 2009, 23 (Special Issue), 37–59.

- Picchio, M.**, “Temporary Contracts and Transitions to Stable Jobs in Italy,” *Labour*, 2008, 22 (Special Issue), 147–174.
- **and J.C. van Ours**, “Market Imperfections and Firm-Sponsored Training,” *Labour Economics*, 2011, 18, forthcoming.
- Ridder, G.**, “An Event History Approach to the Evaluation of Training, Recruitment and Employment Programmes,” *Journal of Applied Econometrics*, 1986, 1 (2), 109–126.
- Sanders, J. and A. de Grip**, “Training, Task Flexibility and the Employability of Low-Skilled Workers,” *International Journal of Manpower*, 2004, 25 (1), 73–89.
- Schnalzenberger, M. and R. Winter-Ebmer**, “Layoff Tax and Employment of the Elderly,” *Labour Economics*, 2009, 16 (6), 618–624.
- Sianesi, B.**, “Differential Effects of Active Labour Market Programs for the Unemployed,” *Labour Economics*, 2008, 15 (3), 370–399.
- Stewart, M.**, “The Interrelated Dynamics of Unemployment and Low-Wage Employment,” *Journal of Applied Econometrics*, 2007, 22 (3), 511–531.
- van den Berg, G.J. and G. Ridder**, “Attrition in Longitudinal Panel Data and the Empirical Analysis of Dynamic Labour Market Behaviour,” *Journal of Applied Econometrics*, 1994, 9 (4), 421–435.
- **and M. Lindeboom**, “Attrition in Panel Survey Data and the Estimation of Multi-State Labor Market Models,” *Journal of Human Resources*, 1998, 33 (2), 458–478.
- van Ours, J.C. and L. Stoeldraijer**, “Age, Wage and Productivity in Dutch Manufacturing,” *De Economist*, 2011, forthcoming.
- Vandenbergh, V.**, “Introduction to De ECONOMIST Special Issue on “Ageing Workforces”,” *De Economist*, 2011, forthcoming.
- Wooldridge, J.M.**, “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity,” *Journal of Applied Econometrics*, 2005, 20 (1), 39–54.