UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Retaining through Training Even for Older Workers

Matteo Picchio [a, c, f]

Jan C. van Ours [a, b, d, e, f]

[a] CentER, ReflecT, and Department of Economics, Tilburg University, The Netherlands
[b] University of Melbourne, Australia
[c] SHERPPA, Ghent University, Belgium
[d] CESifo, Germany
[e] CEPR, United Kingdom
[f] IZA, Germany

October 2011

2011/748

D/2011/7012/53

Retaining through Training Even for Older Workers*

Matteo Picchio[a,c,f,†] and Jan C. van Ours[a,b,d,e,f]

[a] *Tilburg University, CentER, ReflecT, The Netherlands*
[b] *University of Melbourne, Australia*
[c] *Ghent University, Sherppa, Belgium*
[d] *CESifo, Germany*
[e] *CEPR, United Kingdom*
[f] *IZA, Germany*

October 24, 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 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

*Data from the European Community Household Panel Survey 1994–2001 are used with the permission of Eurostat (contract ECHP/2010/16). The results and conclusions are those of the authors and not those of Eurostat, the European Commission, or any of the authorities whose data have been used. 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. Matteo Picchio acknowledges financial support by Fonds Wetenschappelijk Onderzoek (FWO). We also wish to thank the participants to the ESEM conference in Oslo (08/2011), to the XXVI National Conference in Labour Economics in Milan (09/2011), and to the EALE conference in Cyprus (09/2011) for their comments and suggestions.

†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). The effects of ALMP on unemployed workers and welfare recipients have been evaluated in many studies. Kluve (2010) for example presents an overview study of 137 program evaluations from 19 countries. The effect of training on employment prospects of unemployed workers is often found to be mild. Either training does not increase job finding rates significantly or even modest negative effects are found.

Whereas the effect of training of unemployed workers on job finding rates has been studied quite frequently, the effect of training of employed workers on job separation rates is rarely investigated. 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.[1] Our paper focuses on the Dutch labour market. The Dutch labour market is of special interest since in recent years it has one the highest levels of firm-provided training[2] and one of the lowest unemployment rates in Europe. The Netherlands are also interesting because whereas training is often found to be influenced by product market competition (see e.g. Bassannini and Brunello, 2011), we show in a companion paper that in the Netherlands training is affected by labour market imperfections but not by product market competition (Picchio and van Ours, 2011).

We contribute to the small literature on the employment effects of on-the-job training. The lack of evaluation studies in this area is related to the lack of suitable data. Whereas for unemployed workers a training course is well-defined because out of the regular routine of the worker, a training course for an employed workers is often not very well defined as it is part of the work on the job. This implies that often neither the start, the finish nor

[1]Other studies focus on the effects of on-the-job training on wages and productivity. Bartel (1995) for example shows that, at firm level, wages and productivity are positively affected by on-the-job training. 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).

[2]According to the Eurostat Continuing Vocational Training Survey, in 1999 among the EU25 countries the Netherlands ranked third in terms of hours of firm-provided training per 1,000 hours worked (behind Denmark and Sweden), and third in terms of cost of firm-provided training as a percentage of total labour cost (behind the UK and Denmark). More statistics on training can be found at http://epp.eurostat.ec.europa.eu/portal/page/portal/education/data/database.

the nature of the training is documented. Furthermore, when data on on-the-job training is available, it is only rarely the case that these are panel data of any considerable length.

Our empirical analysis is based on data from the European Community Household Panel (ECHP). We investigate whether firm-provided training enhances the probability of retaining workers into the workforce. The ECHP data are quite unique in the sense that they contain panel information of sufficient length which can be exploited to distinguish between correlation and causal effects.[3] Indeed, from an econometric viewpoint, it is challenging to disentangle the pure effect of training from the spurious one determined by individual unobserved heterogeneity. Unobserved heterogeneity like motivations, labour market attachment, and innate ability might indeed jointly determine the likelihood of training participation and the labour market performances. Using techniques to control for the endogeneity of training participations, we explicitly model the interrelated 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. 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 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.

---

[3]Unfortunately, the data collection was discontinued in 2001. So far, ECHP data have only been used to analyse training participation (Arulampalam et al., 2004; Bassannini and Brunello, 2008) or the effects of training on wages and employment security as perceived by the worker (Bassannini, 2004).

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 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 1995 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.[4] Table 1 clarifies the structure of our data.

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 (at the survey time) 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.[5]

[4]We 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.

[5]We build the non-employment indicator on the basis of variables PE003 and PE004 of the ECHP

Table 1: The Structure of the Unbalanced Panel

Years of observation (the initial year $t = 0$ is not included)	Individual records Absolute frequencies	Individual records Relative frequencies	Total records Absolute frequencies	Total records 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

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 $t-1$ has a 88.3% probability of nonemployment at t against 1.9% and 5.5% of people at work at time $t-1$ with and without firm-provided training, respectively. The nonemployment probability is therefore lower for those who attended some firm-provided training in the past than those who did not. 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 expected, 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 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,

survey. The firm-provided training indicator is built on the basis of variables PT001, PT002, and PT017. Question PT001 asks people whether they attended training or education since January of the last year. From question PT002, we can identify work-related training from formal education. Finally, the answers to question PT017 provide information on whether the training course was paid or organized by the firm.

Table 2: Raw Conditional and Unconditional Nonemployment Probabilities

	Employment status at $t-1$			
	Not employed	Employed with firm-provided training	Employed without firm-provided training	Total
Employment status at t				
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

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

	Employment status at $t-1$			
	Not employed	Employed with firm-provided training	Employed without firm-provided training	Total
Employment status at t				
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)

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.[6] 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 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.

[6] In 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
Not-employed	.308	.462	.000	1.000
Employed with firm-provided training	.041	.198	.000	1.000
Employed without firm-provided training	.651	.477	.000	1.000
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 (years)	43.498	10.047	26.000	64.000
Potential experience (years)	18.102	13.662	.000	52.000
Bad health[(a)]	.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 NT		33,348		
Number of individuals N		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.

Table 5: Summary Statistics of Employees

	Mean	Std. Dev.	Minimum	Maximum
Employed with firm-provided training	.059	.236	.000	1.000
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 (years)	41.364	8.902	26.000	64.000
Potential experience (years)	20.441	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 NT		23,080		
Number of individuals N		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.

3 Econometric Modelling

3.1 Dynamic Probit Models

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 employment 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 $i = 1, \ldots, N$ and $t = 1, \ldots, T$

$$\begin{aligned} y_{it} &= \mathbb{1}[y_{it-1}\delta_1 + w_{it-1}\gamma_1 + \mathbf{x}'_{it}\beta_1 + a_{1it} + u_{1it} > 0] && (1) \\ w_{it} &= \mathbb{1}[y_{it-1}\delta_2 + w_{it-1}\gamma_2 + \mathbf{z}'_{it}\beta_2 + a_{2it} + u_{2it} > 0] \quad \text{if } y_{it} = 0, && (2) \end{aligned}$$

where:

- $\mathbb{1}[\cdot]$ is the indicator function;

- y_{it} is a binary variable equal to 1 if individual i was not at work at time t and 0 otherwise;
- w_{it} is equal to 1 if employee i attended firm-provided training courses since the beginning of the previous year and 0 otherwise;
- $\mathbf{x}_{it}$ 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;
- (a_{1it}, a_{2it}) is the 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 Mroz and Savage (2006) to understand the effect of youth unemployment on subsequent labour market performances, 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, a_{1it}, 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, a_{2it}, 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 and across equations due to the presence of the unobserved determinants a_{1it} and a_{2it}. As in Mroz and Savage (2006), these unobserved components are specified, for $j = 1, 2$, as $a_{jit} = c_{ji} + \eta_{jit}$, where c_{ji} is the time-invariant (permanent) fixed-effect and η_{jit} is the unobserved transitory factor. The permanent component c_{ji} captures those characteristics that do not vary over time, such as innate ability and intelligence. The transitory factor η_{jit} is instead time-varying unobserved heterogeneity like motivations and labour market attachments. It allows for contemporaneous correlation at each point in time between employment status and training participation which is not captured by the individual permanent component. Unconditional on a_{1it} and a_{2it} 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.

3.2 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 nonparametric 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.[7] Formally, the parametric specification of the unobserved heterogeneity terms is assumed to be,

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

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

[7] An 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 it is computationally less demanding. Note that Arulampalam and Stewart (2009) and Akay (2011) show that two estimators provide similar results, especially when the panel is moderately long (longer than $T = 5$).

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. 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 with a finite and (*a priori*) unknown number M of points of support. The probabilities associated to the mass points sum to one and, $\forall\, m = 1, \ldots, M$, are denoted by $p^m \equiv \Pr(v_1 = v_1^m, v_2 = v_2^m)$ and specified as logistic transforms:

$$p_m = \exp(\lambda_m) \Big/ \sum_{g=1}^{M} \exp(\lambda_g) \quad \text{with} \quad m = 1, \ldots, M \quad \text{and} \quad \lambda_M = 0.$$

Similarly, the transitory component $\boldsymbol{\eta}_{it} \equiv (\eta_{1it}, \eta_{2it})$ is assumed to be a random draw from a discrete distribution function with Q support points and probability weights denoted by $r^q \equiv \Pr(\eta_1 = \eta_1^q, v_2 = v_2^q)$, $\forall q = 1, \ldots, Q$. As a pre-specified low number of points of support may result in substantial bias, we choose (M, Q) points of support that minimize the Akaike Information Criterion (AIC).

3.3 The Likelihood Function and Average Partial Effects

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

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

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

Since the permanent component $\mathbf{v}_i$ and the transitory component $\boldsymbol{\eta}_{it}$ are not observed and are random terms from bivariate distributions, they 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 distributions are estimated by ML jointly with all the other parameters. On the basis of the model in equations (5) and (6) and the assumptions on the distribution of $\mathbf{v}_i$ and $\boldsymbol{\eta}_{it}$, the contribution to the likelihood function of individual i is given by

$$\mathscr{L}_i = \sum_{m=1}^{M} p^m \prod_{t=1}^{T} \sum_{q=1}^{Q} r^q$$
$$\times \Big\{ \Phi\big[(2y_{it}-1)(y_{it-1}\delta_1 + w_{it-1}\gamma_1 + \mathbf{x}_{it}'\beta_1 + \bar{\mathbf{x}}_i'\alpha_1 + y_{i0}\theta_1 + w_{i0}\psi_1 + v_{1i}^m + \eta_{1it}^q)\big]$$
$$\times \;\Phi\big[(2w_{it}-1)(y_{it-1}\delta_2 + w_{it-1}\gamma_2 + \mathbf{z}_{it}'\beta_2 + \bar{\mathbf{z}}_i'\alpha_2 + y_{i0}\theta_2 + w_{i0}\psi_2 + v_{2i}^m + \eta_{2it}^q)\big]^{(1-y_{it})} \Big\},$$

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 contributions, i.e. $\ell = \sum_{i=1}^{N} \ln(\mathscr{L}_i)$.[8]

Once the model is estimated, we compute predicted probabilities and average partial effects (APEs) which are focal to quantify the impact of firm-provided training on employability. There are different ways in which the marginal effect of y_{it-1} or 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 ($\bar{\mathbf{x}}$) and at the mean of the transitory component $\bar{\bar{\eta}}_1$, 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:

$$\widehat{\pi}_1 = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\widehat{p}^m\Phi(\qquad \bar{\mathbf{x}}'\widehat{\beta}_1 + \bar{\mathbf{x}}_i'\widehat{\alpha}_1 + y_{i0}\widehat{\theta}_1 + w_{i0}\widehat{\psi}_1 + \widehat{v}_{1i}^m + \bar{\bar{\eta}}_1); \tag{7}$$

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

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

We obtain the APEs by taking the difference between these quantities.[9] Two APEs will

[8] We use the Matlab minimizer *fminunc* with analytic first derivatives to obtain the ML estimates.

[9] Standard errors of the predicted probabilities and of the APEs are estimated by bootstrapping the results

be particular useful for discussion in Section 4: $\widehat{\pi}_2 - \widehat{\pi}_1$ and $\widehat{\pi}_2 - \widehat{\pi}_3$. 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.

3.4 Identification

This study deals with employment status and firm-provided training. It treats them as outcome variables which evolve with an endogenous pattern through the individuals' labour market career. The model is designed to recover the potentially endogenous interrelated dynamics of these outcomes, which are also determinants of later outcomes. Identification of the main causal effects crucially depends on the capacity to control for different sources of endogeneity: (i) outcome variables are determinants of later outcomes; (ii) contemporaneous correlation between the outcome variables; (iii) presence of permanent and transitory unobserved components.

We exploit different identification sources. First, the sequencing of the training realizations and of the employment status realization is such that training participation in the period before the interview date might be considered as predetermined with respect to the employment position at the interview date. In other words, the two equations would not be simultaneous. Thanks to the recursive structure of the model, identification would be attained without exclusion restrictions as in Biewen (2009). However, there might be contractual arrangements between firms and workers and delayed compensation schemes. For example, firms might be more likely to train those workers that they seek to retain. If so, the employment at the interview time and training participation in the previous time period would be simultaneous and the predeterminedness assumption would fail.

Then, there might be a second source of information playing a role in providing us with identification of the simultaneous equations model. As shown by Bhargava (1991), to the extent that some of the time-varying variables are strictly exogenous, the imposed stability of the structural parameters over time jointly with the time-variation of such exogenous variables will provide a multiplicity of instruments associated with each exclusion restriction that can be used to attain overidentification in controlling for endogenous

(individual-cluster bootstrap with replacement).

determinants.[10]

Thirdly, some household characteristics (namely the number of household members, an indicator for people living in a couple, a dummy for the presence of kids, and its interaction with the female indicator) will be included in the nonemployment equation but excluded from the firm-provided training equation. We assume that these regressors, conditional on (un)observables, determine the employment probability but not the probability of receiving firm-provided training and they act therefore as excluded restrictions.[11]

Lastly, we have multiple observations per individual that is exploited to identify the unobserved heterogeneity distribution and therefore to control for selection on unobservables under mild parametric restrictions.

4 Estimation Results

4.1 Dynamic Unobserved Effects Probit

Equations (5) and (6) describe our econometric model in its more general formulation. Before presenting the estimation results of such a general model, we begin by showing the estimation results of models under stricter assumptions about the presence of individual heterogeneity and the distribution of the unobserved components. Tables 6 reports the estimation results of four different specifications of univariate models for the nonemployment equation, under the assumption that the nonemployment and the training equations are independent. Table 7 displays instead the estimation results of the bivariate dynamic unobserved effects probit model. 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 APEs that are of focal interest in this paper.

In specification 1, we include only the lagged outcome variables as explanatory variables and we estimate the model by pooled probit ML. The predicted nonemployment probabilities and APEs are equal to those reported in Table 2: there is a strong state dependence in nonemployment and firm-provided training increases the chances of being at work by about 3.6 percentage points. The raw evidence that nonemployment makes future nonemployment more likely, as well as the effect of training on nonemployment, could however be spurious. There might indeed be some individual characteristics determin-

[10]The time-variation of exogenous covariates has been exploited also to show nonparametric identification in mixed hazard duration models (Brinch, 2007). See also Honoré and Tamer (2006) and Heckman and Navarro (2007) for further identification results in panel dynamic discrete choice models.

[11]These exclusion restrictions are similar to those often used in the labour supply literature (e.g. Mroz, 1987).

ing both the employment status and the probability of receiving firm-provided training. For example, individuals that are more able, motivated, and attached to the labour market might be less likely to be out of the workforce and more likely to attend firm-provided training.

In specification 2, we add a set of regressors to control for individual heterogeneity, we assume that there is no unobserved heterogeneity (UH), and we estimate the model by pooled probit ML. The inclusion of controls for observed individual heterogeneity removes some spurious components from the predicted employment probabilities. The probability of nonemployment given past nonemployment decreases from 88.3% to 82.7%. The nonemployment probability given past employment rises to 6.2%, if no training in the past, and to 2.9%, in case of training.

In specifications 3 and 4, we take into account the presence of correlated random effects, although without allowing yet for the presence of transitory components. In specification 3, we impose a Gaussian distribution on the residual permanent component v_{1i}, whilst in specification 4 we stick to a discrete distribution where the number of support points are chosen by following the AIC. Four points are worthy of mention. First, the estimation results are robust to the choice of the UH distribution. Second, the model with discrete UH is favoured according to standard information criteria. Third, firm-provided training significantly reduces future nonemployment probability by about 3 percentage points. Fourth, once the spurious effects due to permanent unobserved components are controlled for, we still find a large state dependence effect. Nevertheless, the predicted probability of nonemployment given past nonemployment decreases drastically to about 49%, 16.6 (13.3) percentage points higher than the nonemployment probability conditional on past employment with(out) training. These predicted probabilities suggest also that most of the state dependence in nonemployment is spurious and determined by unobserved permanent components.

Table 7 reports estimation results when the nonemployment equation and the training equation are allowed to be correlated through the unobserved heterogeneity determinants. The unobserved heterogeneity is assumed to be made up of a permanent component and a transitory component, as explained in Subsection 3.2. The numbers of points of support minimizing the AIC are $M = 3$ for the distribution of the permanent component $\mathbf{v}_i$ and $Q = 2$ for the one of the transitory component $\boldsymbol{\eta}_{it}$. Estimation results of the unobserved heterogeneity distributions are reported in Table 8.

The predicted probabilities and the estimated APEs from the bivariate model are qualitatively in line with those from the univariate model. From the quantitative point of view, the effect of training is bigger in size: an employee with a given set of observed and unobserved characteristics is 5.7 percentage points less likely to be out of the workforce at

Table 6: Univariate Dynamic Unobserved Effects Probit Model for the Nonemployment Probability

	Specification 1 No (un)observed heterogeneity			Specification 2 No unobserved heterogeneity			Specification 3 Correl. random effects dynamic probit with Gaussian UH			Specification 4 Correl. random effects dynamic probit with discrete UH		
Variable	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.
Nonemployment$_{t-1}$	2.790	***	.021	2.480	***	.031	1.567	***	.049	1.533	***	.040
Firm-provided training$_{t-1}$	-.483	***	.081	-.360	***	.084	-.376	***	.111	-.367	***	.113
Female	–		–	.116	***	.027	.210	***	.052	.221	***	.056
Education - Reference: Education ISCED 0-2												
Education ISCED 5-7	–		–	-.168	***	.033	-.289	***	.058	-.279	***	.058
Education ISCD 3	–		–	-.116	***	.025	-.265	***	.045	-.244	***	.044
Age - Reference: [26,35] years old												
(35, 49] years old	–		–	.163	***	.031	.138	***	.048	.117	**	.046
(49, 64] years old	–		–	.527	***	.039	.547	***	.068	.558	***	.071
Work experience/10	–		–	-.517	***	.030	1.312	***	.235	1.215	***	.185
Work exper. squared/1000	–		–	.112	***	.008	.226	***	.038	.236	***	.038
Bad health	–		–	.347	***	.026	.228	***	.046	.233	***	.043
No. household members	–		–	-.032	**	.013	.055		.043	.061		.041
Kids<12 years	–		–	.085	**	.042	.234	*	.120	.230	*	.128
Kids<12 years$*$Female	–		–	.179	***	.046	.224		.145	.225		.149
Living in a couple	–		–	.129	***	.044	-.551	***	.132	-.550	***	.139
ln(household net income)	–		–	-.001		.008	.075	***	.017	.073	***	.015
Time indicators – Reference: 1996												
1997	–		–	.009		.040	-.166	***	.048	-.153	***	.047
1998	–		–	-.044		.036	-.392	***	.056	-.368	***	.055
1999	–		–	-.043		.036	-.546	***	.066	-.515	***	.063
2000	–		–	-.118	***	.038	-.782	***	.078	-.747	***	.070
2001	–		–	-.072	*	.039	-.914	***	.092	-.870	***	.081
Constant	-1.598	***	.014	-1.483	***	.056	-1.570	***	.100	–		–
Initial conditions												
Nonemployment$_0$	–		–	–		–	1.672	***	.097	1.926	***	.108
Firm-provided training$_0$	–		–	–		–	-.165		.131	-.158		.128
Predicted probability $\widehat{\pi}_1$	.055	***	.001	.062	***	.002	.337	***	.015	.353	***	.013
Predicted probability $\widehat{\pi}_2$	.019	***	.004	.029	***	.006	.308	***	.018	.323	***	.017
Predicted probability $\widehat{\pi}_3$	.883	***	.003	.827	***	.006	.483	***	.009	.489	***	.009
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.036	***	.004	-.033	***	.006	-.029	***	.009	-.030	***	.009
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.865	***	.005	-.798	***	.008	-.175	***	.019	-.166	***	.018
NT (N)	33,348 (7,257)			33,348 (7,257)			33,348 (7,257)			33,348 (7,257)		
Log-likelihood	-8,443.0			-7,793.0			-7,444.3			-7,298.2		
No. of parameters	3			21			31			37		
AIC/N	2.325			2.156			2.060			2.022		
Pseudo-R^2	.590			.622			.638			.646		

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Individual time averages of the time-varying covariates are included in specifications 3 and 4 but not reported for the sake of brevity. The standard errors of the predicted probabilities and the APEs are obtained by bootstrapping the resuls 500 times (individual-cluster bootstrap with replacement).

Table 7: Bivariate Dynamic Unobserved Effects Probit Model with Discrete UH

	Specification 5					
	Nonemployment equation			Firm-provided training equation		
Variable	Coeff.		S.E.	Coeff.		S.E.
Nonemployment$_{t-1}$	1.567	***	.040	-.120		.103
Firm-provided training$_{t-1}$	-.738	***	.122	.735	***	.059
Female	.212	***	.053	-.007		.044
Education - Reference: Education ISCED 0-2						
Education ISCED 5-7	-.271	***	.055	-.133	**	.059
Education ISCD 3	-.239	***	.041	.001		.046
Age - Reference: [26,35] years old						
(35, 49] years old	.108	**	.045	-.146	***	.048
(49, 64] years old	.536	***	.066	-.219	**	.099
Work experience/10	1.209	***	.175	-.173		.319
Work exper. squared/1000	.216	***	.038	-.039		.055
Bad health	.229	***	.043	.067		.061
No. household members	.056		.041	–		–
Kids<12 years	.229	*	.128	–		–
Kids<12 years$*$Female	.212		.149	–		–
Living in a couple	-.543	***	.139	–		–
$\ln$(household net income)	.075	***	.015	-.036	**	.017
Permanent contract	–		–	.040		.062
Part-time job	–		–	-.369	***	.083
Public employment	–		–	.115		.071
Occupation indicators – Reference: High-skilled white collar worker						
Blue collar	–		–	-.240	***	.085
Low-skilled white collar	–		–	-.215	***	.064
Sectoral indicators – Reference: Services						
Agriculture	–		–	.026		.382
Industry	–		–	-.236	***	.086
Unknown sector	–		–	-.328	***	.081
Job tenure indicators – Reference: 0-4 years						
Unknown	–		–	-.396	***	.143
5-9 years	–		–	-.283	***	.065
10-14 years	–		–	-.250	***	.090
15 years or more	–		–	-.287	**	.120
Initial conditions and cross-equations correlation						
Nonemployment$_0$	1.783	***	.100	.688	***	.124
Firm-provided training$_0$	-.127		.144	.353	***	.070
$\widehat{\rho}_v$	.477	***	.182			
Predicted probability $\widehat{\pi}_1$	.348	***	.016	–		–
Predicted probability $\widehat{\pi}_2$	.292	***	.022	–		–
Predicted probability $\widehat{\pi}_3$	.495	***	.009	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.057	***	.013	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.203	***	.023	–		–
NT (N)	33,348 (7,257)					
Log-likelihood	-11,914.8					
No. of parameters	99					
AIC/N	3.311					
Pseudo-R^2	.538					

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level. Time dummies, firm size indicators, and individual time averages of the time-varying covariates are included in the model specification but not reported for the sake of brevity. The standard errors of the predicted probabilities, APEs, and $\widehat{\rho}_v$ are obtained by bootstrapping the resuls 500 times (individual-cluster bootstrap with replacement).

Table 8: Estimation Results of the UH Distributions of Model Specifications 3, 4, and 5.

	Nonemployment equation Coeff.		S.E.	Firm-provided training equation Coeff.		S.E.	Probability weight
				Specification 3			
$\widehat{\sigma}_v$	.896	***	.025	–		–	
				Specification 4			
$\widehat{v}_1^1$	-3.281	***	.412	–		–	p^1=.089
$\widehat{v}_1^2$	-2.445	***	.197	–		–	p^2=.452
$\widehat{v}_1^3$	.211		.181	–		–	p^3=.070
$\widehat{v}_1^4$	-.922	***	.139	–		–	p^4=.389
				Specification 5			
Permanent component							
$\widehat{v}_j^1$	-2.626	***	.136	-2.429	***	.191	p^1=.566
$\widehat{v}_j^2$	-.979	***	.119	-2.008	***	.204	p^2=.431
$\widehat{v}_j^3$	2.098	**	1.016	1.585		4.726	p^3=.003
Transitory component							
$\widehat{\eta}_j^1$	.000		–	.000		–	r^1=.704
$\widehat{\eta}_j^2$	.828	***	.102	.691	***	.132	r^2=.296

Notes: * Significant at 10% level; ** significant at 5% level; *** significant at 1% level.

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 34.8% to 29.2%, i.e. by 16.1%. As shown in the estimation result of specification 1 in Table 6, the corresponding figure from raw data is 65.5%. This points out that more than three fourths of the relative raw effect of firm-provided training on employability is spurious and determined by observed and unobserved characteristics.

Looking at the impact of exogenous variables on the nonemployment probability, women are less likely to be at work. The probability of being out of the workforce is, *ceteris paribus*, increasing with age and 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, 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 worker are less likely to receive training. There is therefore some evidence that education and firm-provided training are not complementary assets in the Netherlands, in contrast to the findings in Blundell et al. (1999) for the US and the UK. Part-time workers are less likely to get firm-provided training. High skilled white collar workers are more likely to receive training, suggesting that tasks and human capital formation are complementary assets. Finally, firm-provided training is more present in the services sector and among newly hired workers.[12]

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% against an European average of 39.1%.[13] 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.[14] 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 are unlikely 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 often a one-way street out of the labour force or into long-term unemployment.[15] Employers are indeed reluctant to hire an older worker because of the pay-productivity gap and because of the possible

[12] Firm size indicators are included in the specification of the training equation (and not reported in Table 7) but they are not (jointly) significantly different from zero.

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

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

[15] 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.

obsolescence of general human capital.[16] 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 subsection, 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 9 reports the estimation results of the coefficients and APEs of primary interest. 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 6.8 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 very close 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 Dutch public sector 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

[16] See 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 9: Bivariate Dynamic Unobserved Effects Probit Model with Age Interactions

	Nonemployment equation			Firm-provided training equation		
Variable	Coeff.		S.E.	Coeff.		S.E.
Nonemployment$_{t-1}$	1.255	***	.062	-.125		.141
Firm-provided training$_{t-1}$	-.703	***	.182	.766	***	.075
Nonemployment$_{t-1}$*Age $(35, 49]$	.309	***	.068	-.005		.183
Nonemployment$_{t-1}$*Age $(49, 64]$	.859	***	.076	-.078		.350
Firm-provided training$_{t-1}$*Age $(35, 49]$	-.185		.248	-.012		.095
Firm-provided training$_{t-1}$*Age $(49, 64]$	-.099		.309	-.359	**	.180
Age $(35, 49]$	-.018		.050	-.152	***	.051
Age $(49, 64]$	.128	*	.075	-.213	**	.103
Predictions and APEs if age $[26, 35]$						
Predicted probability $\widehat{\pi}_1$	.340	***	.016	–		–
Predicted probability $\widehat{\pi}_2$	.285	***	.023	–		–
Predicted probability $\widehat{\pi}_3$	.457	***	.010	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.055	***	.016	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.172	***	.023	–		–
Predictions and APEs if age $(35, 49]$						
Predicted probability $\widehat{\pi}_1$	.339	***	.015	–		–
Predicted probability $\widehat{\pi}_2$	.270	***	.028	–		–
Predicted probability $\widehat{\pi}_3$	.488	***	.010	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.069	***	.021	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.218	***	.030	–		–
Predictions and APEs if age $(49, 64]$						
Predicted probability $\widehat{\pi}_1$	.352	***	.015	–		–
Predicted probability $\widehat{\pi}_2$	.284	***	.039	–		–
Predicted probability $\widehat{\pi}_3$	.562	***	.013	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.068	*	.035	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.278	***	.046	–		–
NT (N)			33,348 (7,257)			
Log-likelihood			-11,855.1			
No. of parameters			107			
AIC/N			3.297			
Pseudo-R^2			.540			

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 of the predicted probabilities and the APEs are obtained by bootstrapping the resuls 500 times (individual-cluster bootstrap with replacement).

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.[17] The first type of policy would 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 increasing the likelihood that they retain their employability.

4.3 Further Robustness Checks

We perform three 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; iii) of the distribution of the unobserved heterogeneity.

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 asses 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.[18] 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

[17] 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.

[18] The indicator for vocational training not provided by the firm is built on the basis of variables PT001, PT002, and PT017 of the ECHP data.

benchmark model, yielding the following simultanous three-equation model

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

where κ is the loading factor determining, together with v_{2i}, the points of support of the permanent component of the equation of training not provided by firms. The loading factor is used to simplify the specification of the distribution of the permanent unobserved heterogeneity and reduce the computational complexity in estimating the model. The likelihood function of the benchmark model can be trivially extended to the three-equation case. Table 10 reports the estimation results of the three-equation model. All the estimation results of the nonemployment equation and of the firm-provided equation are in line with those reported in Table 7.

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 and thereby removed from the model specification. Table 11 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. 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 -5.5 percentage points in the probability of being nonemployed (-34%). If not at work at $t-2$, the APE is equal to -6.3 percentage points (-25%). Hence, the estimated APE $\widehat{\pi}_2 - \widehat{\pi}_1$ of our benchmark model does not seem to be sensitive to the specification of the dynamic 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.

Finally, we estimate linear probability models. As pointed out by Stewart (2007), they can indeed be viewed as semiparametric since they do not impose parametric restrictions

Table 10: Dynamic Unobserved Effects Probit Model with 3 Simultaneous Equations

Variable	Nonemployment equation Coeff.		S.E.	Firm-provided training equation Coeff.		S.E.	Other vocational training equation Coeff.		S.E.
Nonemployment$_{t-1}$	1.610	***	.039	-.059		.101	.114		.101
Firm-provided training$_{t-1}$	-.504	***	.119	.802	***	.057	.360	***	.085
Other vocational training$_{t-1}$	.002		.092	.253	***	.092	1.128	***	.082
Female	.204	***	.051	-.011		.042	.045		.061
Education - Reference: Education ISCED 0-2									
Education ISCED 5-7	-.258	***	.053	-.118	**	.057	-.080		.075
Education ISCD 3	-.228	***	.040	.013		.045	-.061		.059
Age - Reference: [26,35] years old									
(35, 49] years old	.106	**	.044	-.140	***	.047	-.009		.060
(49, 64] years old	.547	***	.064	-.221	**	.097	-.138		.116
Work experience/10	1.148	***	.180	-.305		.319	.114		.421
Work exper. squared/1000	.205	***	.038	-.041		.056	.038		.076
Bad health	.224	***	.043	.060		.062	-.148	*	.085
No. household members	.057		.040	–		–	–		–
Kids<12 years	.218	*	.129	–		–	–		–
Kids<12 years$*$Female	.218		.149	–		–	–		–
Living in a couple	-.523	***	.138	–		–	–		–
ln(household net income)	.072	***	.015	-.033	*	.017	-.016		.021
Permanent contract	–		–	.085		.062	-.197	**	.083
Part-time job	–		–	-.346	***	.083	.227	**	.091
Public employment	–		–	.121	*	.072	-.211	**	.097
Occupation indicators – Reference: High-skilled white collar worker									
Blue collar	–		–	-.209	**	.085	-.328	***	.114
Low-skilled white collar	–		–	-.191	***	.064	.015		.081
Sectoral indicators – Reference: Services									
Agriculture	–		–	.055		.378	.663	**	.321
Industry	–		–	-.220	**	.086	-.070		.122
Unknown sector	–		–	-.321	***	.082	-.181		.111
Job tenure indicators – Reference: 0-4 years									
Unknown	–		–	-.307	**	.142	-.171		.158
5-9 years	–		–	-.274	***	.066	.087		.091
10-14 years	–		–	-.233	**	.092	.165		.152
15 years or more	–		–	-.266	**	.122	.215		.206
Initial conditions									
Nonemployment$_0$	1.700	***	.095	.498	***	.108	.516	***	.119
Firm-provided training$_0$	-.101		.124	.333	***	.066	.080		.107
Other vocational training$_0$	.150		.126	.101		.078	.645	***	.107
Predicted probability $\widehat{\pi}_1$	.343	***	.015	–		–	–		–
Predicted probability $\widehat{\pi}_2$	.301	***	.021	–		–	–		–
Predicted probability $\widehat{\pi}_3$	.500	***	.010	–		–	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.042	***	.013	–		–	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.199	***	.025	–		–	–		–
NT (N)				33,348 (7,257)					
Log-likelihood				-14,646.1					
No. of parameters				162					
AIC/N				4.081					
Pseudo-R^2				.498					

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 of the predicted probabilities and APEs are obtained by bootstrapping the resuls 239 times (individual-cluster bootstrap with replacement).

Table 11: Dynamic Unobserved Effects Probit Model with Lag of Order Two

	Nonemployment equation			Firm-provided training equation		
Variable	Coeff.		S.E.	Coeff.		S.E.
Nonemployment$_{t-1}$	1.909	***	.036	-.120		.128
Nonemployment$_{t-2}$	.774	***	.042	.150		.109
Firm-provided training$_{t-1}$	-.535	***	.130	.763	***	.061
Predictions and APEs if the individual was employed at $t-2$						
Predicted probability $\widehat{\pi}_1$	.161	***	.041	–		–
Predicted probability $\widehat{\pi}_2$	.107	***	.041	–		–
Predicted probability $\widehat{\pi}_3$	.400	***	.038	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.055	***	.014	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.294	***	.028	–		–
Predictions and APEs if the individual was not employed at $t-2$						
Predicted probability $\widehat{\pi}_1$	.253	***	.040	–		–
Predicted probability $\widehat{\pi}_2$	.189	***	.046	–		–
Predicted probability $\widehat{\pi}_3$	.521	***	.044	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_1$	-.063	***	.016	–		–
APE: $\widehat{\pi}_2 - \widehat{\pi}_3$	-.332	***	.043	–		–
No. of individuals	7,257					
No. of observations	26,050					
Log-likelihood	-9,105.3					
No. of parameters	99					
AIC/N	2.537					
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 of the predicted probabilities and APEs are obtained by bootstrapping the resuls 500 times (individual-cluster bootstrap with replacement).

on the distribution of the unobserved heterogeneity component. In a linear framework, we can simply get rid of the fixed-effects by first-differencing. Then, we exploit further lags of the endogenous variables as instruments in an Arellano and Bond (1991) GMM estimation framework. The estimated effect of training on future nonemployment is equal to -0.029, very much in line with the estimated APEs reported in Tables 6 and 7. The estimated effect is however not significant.[19]

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 find that older workers who receive on-the-job training are more likely to retain 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 aging 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 which type of policy is needed to stimulate on-the-job training and whether or not our findings for older workers in the Netherlands are unique. Our findings may be related to Dutch labour market institutions or to favorable labor market conditions, i.e. low unemployment rates. We leave the analysis of the employment effect of on-the-job training in other countries and potential differences compared to the findings in this analysis to future research.

References

Akay, A., "Finite-Sample Comparison of Alternative Methods for Estimating Dynamic Panel Data Models," *Journal of Applied Econometrics*, 2011, *forthcoming*.

[19]The robust standard error is 0.080. The whole set of estimation results of linear probability models are not reported for the sake of brevity but are available upon request from the authors.

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.

Arellano, M. and S. Bond, "Some Tests of Specifications for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economics Studies*, 2 1991, *58*, 277–297.

Arulampalam, W. and M.B. Stewart, "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators," *Oxford Bulletin of Economics and Statistics*, 2009, *71* (5), 659–681.

__ , M.L. Bryan, and A.L. Booth, "Training in Europe," *Journal of the European Economic Association*, 2004, *2* (2-3), 346–360.

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.

Bassannini, A., "Improving Skills for More and Better Jobs? The Quest for Efficient Policies to Promote Adult Education and Training," *European Economy, Special Report*, 2004, *3*, 105–137.

__ and G. Brunello, "Is Training More Frequent When the Wage Premium Is Smaller? Evidence from the European Community Household Panel," *Labour Economics*, 2008, *15* (2), 272–290.

__ and __ , "Barriers to Entry, Deregulation and Workplace Training: A Theoretical Model with Evidence from Europe," *European Economic Review*, 2011, *forthcoming.*

Bassi, L.J., "Estimating the Effect of Training Programs with Non-Random Selection," *Review of Economics and Statistics*, 1984, *66* (1), 36–43.

Bhargava, A., "Identification and Panel Data Models with Endogenous Regressors," *Review of Economic Studies*, 1 1991, *58*, 129–140.

Biewen, M., "Measuring State Dependence in Individual Poverty Histories when There is Feedback to Employment Status and Household Composition," *Journal of Applied Econometrics*, 2009, *24* (7), 1095–1116.

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. Fougère, and A. Sérandon, "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories," *Review of Economic Studies*, 1997, *64* (6), 683–713.

Brinch, C.N., "Nonparametric Identification of the Mixed Hazards Model with Time-Varying Covariates," *Econometric Theory*, 2007, *23* (2), 349–354.

Chamberlain, G., "Panel Data," in Z. Grichiles 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.

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.

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.

— **and S. Navarro**, "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 2007, *136* (2), 341–396.

Honoré, B.E. and E. Tamer, "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 2006, *74* (3), 611–629.

Hyslop, D.R., "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 1999, *67* (6), 1255–1294.

Kluve, J., "The Effectiveness of European Active Labour Market Programmes," *Labour Economics*, 2010, *17* (6), 904–918.

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., "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 1987, *55* (4), 765–799.

— **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.

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* (4), 712–722.

Schnalzenberger, M. and R. Winter-Ebmer, "Layoff Tax and Employment of the Elderly," *Labour Economics*, 2009, *16* (6), 618–624.

Stewart, M., "The Interrelated Dynamics of Unemployment and Low-Wage Employment," *Journal of Applied Econometrics*, 2007, *22* (3), 511–531.

van Ours, J.C. and L. Stoeldraijer, "Age, Wage and Productivity in Dutch Manufacturing," *De Economist*, 2011, *159* (2), 113–137.

Vandenberghe, V., "Introduction to De Economist Special Issue on "Ageing Workforces"," *De Economist*, 2011, *159* (2), 89–94.

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.