QUANTITATIVE METHODS IN ECONOMICS
Vol. XI, No. 1, 2010, pp. 100-109

ASSESSING THE IMPACT OF TRAINING ON UNEMPLOYMENT DURATION USING HAZARD MODELS WITH INSTRUMENTAL VARIABLES[14]

Joanna Landmesser
Department of Econometrics and Statistics
Warsaw University of Life Sciences
e-mail: Joanna_landmesser@sggw.pl

Abstract: The goal of the study is to prove the effectiveness of training programs directed to the unemployed on the local labor market in Poland. We estimate a semiparametric hazard model to assess the impact of training on the individual's unemployment duration. To resolve the potential sample selection problem, the participation in a training program is instrumented using a probit model. The main question of this paper is whether the training significantly raises the transition rate from the unemployment into the employment state.

Key words: program's evaluation, instrumental variable method, hazard models

INTRODUCTION

In 2002, the government of Poland implemented a special program, whose main objective was the vocational activation of people belonging to risk groups in local job markets. The program is administered by the Polish Ministry of Labor and Social Policy and funded by the Labor Fund. The expenditure on the active labor market policy has increased, especially on trainings, apprenticeships and vocational training at the workplace. Program beneficiaries are selected from the unemployed workers who register in the state labor offices.

As far as the impact of the active labor market policy in Poland is concerned, literature is modest. The effectiveness of the policy has been studied by the World

[14] This work was supported by grant No. N N111 209436 from Polish Ministry of Science and Higher Education.

Bank in 1997, as part of a project covered by Czech Republic, Poland, Hungary and Turkey, and by Kluve, Lehmann and Schmidt [Kluve et al. 2000] and also by Puhani [Puhani 1998]. The recent program has been evaluated in 2008 by the Polish Ministry of Labor and Social Policy. The research indicated that only trainings and business incentives increased the chances of finding jobs, whereas programs such as intervention and public works were ineffective [Bukowski 2008]. The part of this research was a microeconometric analysis based on the logit model and propensity score matching.

The selection bias problem, which is crucial for good evaluations of the training's program, was considered by Landmesser [Landmesser 2010]. This evaluation employed matching methods to find a control group for the group of trainees, and it assessed the impact of the vocational training on the unemployment duration using a hazard model. A positive effect of training on reemployment probabilities was found. Although the study was carefully implemented, the method used to control for endogeneity was in this case rough and the matching was imprecise. For every treated one, only one untreated one, that resembled it as much as possible in terms of observable pre-training characteristics, was selected. As a result, too many individuals were excluded from the pooled sample and the method was not enough sufficient for evaluating the impact of the program. The matching methods are not robust against "hidden bias" arising from unobserved variables that simultaneously affect assignment to treatment.

In this article we would like to use an alternative method for evaluating the impact of the vocational training on the duration of unemployment. The goal of the study is to prove the effectiveness of training programs directed to the unemployed on the local labor market in Poland. We identify the effectiveness with the impact of training on chances of finding jobs. To resolve the potential sample selection problem, the participation in a training program is instrumented using a probit model. Then, a semiparametric hazard model is estimated to assess the impact of training on the length of the employment search. We investigate whether the training significantly raises the transition rate from the unemployment into the employment state in the short- and the long-run.

In our research study we try to analyze the situation on the local labor market. Therefore, the study is based on the data obtained from the District Labor Office in Słupsk in Poland from 2000 to 2007.

METHODOLOGICAL CONSIDERATIONS

The aim of the evaluation of training program effects is to assess the difference between the level of the outcome variable (i.e. duration of the unemployment period) at time t for a given participant having received training and the level of that variable at time t for the same individual without participation in the training program [Hujer et al. 1999]. Let Y_i^1 be the unemployment duration after training, and Y_i^0 - the unemployment duration without training. The effect of training for individual *i* is then defined as $Y_i^1 - Y_i^0$. But it is impossible to observe individual treatment effect since we do not know the outcomes for untreated observations when it is under treatment, and for treated when it is not under treatment. We can only observe either Y_i^1 or Y_i^0, never both (compare with [Lalonde, 1986], [Dehejia, Wahba, 1999]).

If the group of treated and the group of untreated are random samples from the population, the outcomes are independent of treatment: $Y^0, Y^1 \perp P$. In such a case, the average treatment effect could be obtained by comparing the expected level of the outcome for the two groups.

For the non-experimental data sets like ours, the independence assumption is not valid and we have to cope with sample selection problem [Heckman, et al. 1998]. The comparison between the outcomes of the two groups requires some assumptions. The conditional independence assumption states that conditional on the relevant covariates X, the outcomes are independent of treatment variable P:

$$Y^0, Y^1 \perp P \mid X \tag{1}$$

Consider the simple linear model

$$Y = \beta X_Y + \alpha P + u \tag{2}$$

The error term *u* embodies all omitted (observed and unobserved) factors that determine *Y*. If the assumption (1) is not fulfilled, there may be a correlation between the treatment variable *P* and *u*. The variable *P* is then endogenous and OLS gives biased estimates of parameters (selection bias). The solution to the problem, for instance, is to estimate simultaneously the equation for treatment and then the outcome equation. It is also possible, to use instrumental variable (IV) methods to handle endogenous treatment variable. In the IV approach, the participation is substituted with a variable *IP* (an instrument) that is correlated with participation *P* but not with error term *u*. If we denote $\widetilde{Z} = (X_Y, IP)$ and $\widetilde{X} = (X_Y, P)$ the IV estimator for linear model equals $(\widetilde{Z}^T \widetilde{X})^{-1} \widetilde{Z}^T Y$ ([Bowden, Turkington 1984], [Bijwaard 2008]).

To tackle the problem of sample selection in this study, we are substituting participation in the training program *P* with a variable that is correlated with participation but not with error term *u*. We implement probit regressions for men and women separately to analyze the determinants of participation in the training course. For individual *i* the past participation in the training program P_i is defined as:

$$P_i = \begin{cases} 1 & \text{if } P_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The latent variable P_i^* is defined as a function of a vector of individual-level variables X_{Yi}, a vector of exogenous variables X_{Pi} and an error component v_i:

$$P_i^* = \gamma X_{Yi} + \delta X_{Pi} + v_i \tag{4}$$

We propose the propensity to participate in a training course as a suitable instrumental variable.

The outcome variable, we are interested in, is the duration of time an individual spends in the state of being unemployed. Therefore, in the next step we analyze the impact of training program on the length of the employment search. An appropriate approach, which considers right censoring of unemployment spells, and which controls characteristics of individuals that influence the unemployment duration, is the use of hazard models (see, e.g. [Kalbfleisch, Prentice, 1980], [Hosmer, Lemeshow, 1999], [Cameron, Trivedi, 2005]).

In the terminology of survival analysis, the survivor function $S(t)$ is the probability that the length of the unemployment after training exceeds a time point t and is defined by

$$S(t) = \Pr[T > t] = 1 - F(t) \tag{5}$$

where T is a random variable, which represents the duration in the unemployment state with a density function $f(t)$.

Given $S(t)$, the hazard function $h(t)$ denoting the chance of leaving the unemployment state at time t among the individuals who were not yet employed at that time is

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d(\log S(t))}{dt} = \lim_{dt \to 0} \frac{\Pr[t \le T < t + dt \mid T \ge t]}{dt} \tag{6}$$

In other words, the hazard function $h(t)$ is the limit of the probability that the unemployment episode is completed during the interval $[t, t+dt]$, given that it has not been completed before time t, for $dt \to 0$. The hazard rate – the value of hazard function – describes the intensity of transition from one state to another.

The survivor curve can be specified as a function of individual characteristics for unemployed people and the program participation, so that $h = h(t; X_Y, P)$. The widely applied semiparametric method of analyzing the effect of covariates on the hazard rate is the Cox's proportional hazard model [Cox 1972]. In Cox model we have:

$$h(t; X_{Yi}, P_i) = h_0(t) \exp(\beta X_{Yi} + \alpha P_i) \tag{7}$$

Cox proposed a partial maximum likelihood estimation of this model. The model is estimated non-parametrically and there is no need to make assumptions about the baseline hazard $h_0(t)$. It can be stated that in the Cox model the hazard functions for two individuals *i* and *j* are multiplicatively related, that is their ratio is constant. One subject's hazard is a multiplicative replica of another one. If parameter α is positive, the individual receiving a training is likely to find employment before the individual who received no training.

CHARACTERISTICS OF THE DATA SET

The data used in our analysis concern the unemployed registered in the District Labor Office in Słupsk in Poland in the period from January 2000 to August 2007. The selected sample consists of 3513 persons, who were registered as unemployed at least for one day. On the basis of the history of events for each person registered in the labor office we can state the period of time a person was looking for a job or the period of time during which an unemployed is actually looking for a job (in days). The time spent in the unemployment state is called a spell. The spell is completed when the event occurs (finding a job). Otherwise, unemployment spells are treated as right censored. While our data basis contains multiple spells for 3513 persons we have got 6198 episodes. Descriptive statistics for the resulting spell data set can be found in Table 1.

Table 1. Descriptive statistics for the data set

number of:		mean duration in days:	
individuals	3513		
spells	6198	all spells	349,45
censored spells	870	censored spells	714,54
spells of trainees	625	spells of trainees	404,19
spells of non-trainees	5573	spells of non-trainees	343,31

Source: own computations

The participation in a vocational training seems to increase the unemployment duration. However, a simple comparison between the averages has to be done carefully since it is subject to potential selection effects.

To model the participation in a vocational training the set of covariates for the hazard models includes dummy variables capturing the whole, the short-run and the long-run effect of the participation in the training:

tr – participation in a vocational training during the last 3 years prior to the unemployment beginning,

trs – participation in a vocational training during the last 12 months prior to the unemployment beginning, *trl* – participation in a vocational training between 13 and 36 months prior to the unemployment beginning.

EMPIRICAL RESULTS

To tackle the problem of sample selection, we estimate probit regressions for the whole sample and for men and women separately to analyze the determinants of the participation in the training program. The empirical studies on training participation suggest that important determinants of training are: age, sex, caring for children, belonging to minority groups, educational degrees or occupational status (e.g. [Blundell et al. 1994], [Hujer et al. 1999]). Our empirical findings show that the only significant variables for the participation in training are:
period – dummy variable: 1 for the time period 2004-2007 and 0 for the time period 2000-2003 (with this variable we prove the training course availability for the unemployed in the time span),
edu1 – dummy variable: 1 if individual has incomplete primary, primary, lower secondary or basic vocational education level,
edu2 – dummy variable: 1 if individual has general secondary, vocational secondary or post-secondary education level,
edu3 – dummy variable: 1 if individual has tertiary education level,
language – dummy variable: 1 if individual declares any foreign language skills.
The hypothesis that age, sex or the marital status has influence on training participation could not be confirmed. The results of probit models estimation for men and women separately are given in Table 2.

Individual participation in the vocational training is influenced by the variable *period*, which confirms that the variable could be a valid instrumental variable. Individuals who have primary or secondary education levels, in comparison with the tertiary education level, or individuals who declare any foreign language skills tend to participate more.

Table 2. Results of probit models estimation for participation in training during the last 3 years (variable *tr*), last 12 months (*trs*) and between 13 and 36 last months (*trl*)

Covariates	Whole sample		Men		Women	
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Probit regression for variable *tr*						
period	0.329	0.000	0.333	0.000	0.339	0.000
edu1	0.555	0.000	0.794	0.004	0.314	0.085
edu2	0.470	0.001	0.519	0.066	0.462	0.006
language	0.232	0.001	0.230	0.010	0.204	0.058
cons	-2.563	0.000	-2.727	0.000	-2.479	0.000
	No. of obs. = 6,198 Pseudo R2 = 0.029		No. of obs. = 3,247 Pseudo R2 = 0.032		No. of obs. = 2,951 Pseudo R2 = 0.034	
Probit regression for variable *trs*						
period	0.272	0.002	0.233	0.056	0.330	0.011
edu1	0.546	0.006	0.699	0.055	0.397	0.115
edu2	0.421	0.033	0.205	0.598	0.504	0.030
language	0.400	0.000	0.404	0.001	0.371	0.012
cons	-3.005	0.000	-3.092	0.000	-2.973	0.000
	No. of obs. = 6,198 Pseudo R2 = 0.037		No. of obs. = 3,247 Pseudo R2 = 0.043		No. of obs. = 2,951 Pseudo R2 = 0.049	
Probit regression for variable *trl*						
period	0.314	0.000	0.344	0.000	0.284	0.014
edu1	0.460	0.009	0.729	0.037	0.188	0.397
edu2	0.420	0.017	0.616	0.084	0.335	0.103
language	0.073	0.374	0.079	0.453	0.030	0.821
cons	-2.625	0.000	-2.830	0.000	-2.501	0.000
	No. of obs. = 6198 Pseudo R2 = 0.021		No. of obs. = 3247 Pseudo R2 = 0.026		No. of obs. = 2951 Pseudo R2 = 0.019	

Source: own computations using Stata Statistical Software

Now we consider the impact of training on the length of the unemployment duration. The estimated hazard models as a determinant for the probability of leaving the unemployment state will comprise usual socio-demographic characteristics of individuals and variables capturing the effect of participation in training. The additional new covariates are dummies:
age 25 – with 1 if individual is 25 or younger,
age 2640 – with 1 if individual is 26 or older, but younger than 41,
age41 – with 1 if individual is 41 or older,
marr – with 1 if individual is married,
town – with 1 if the place of residence is town,
disabled – with 1 if individual is disabled,
benefit – with 1 if individual receives unemployment benefit.

The results of Cox regressions are given in Table 3. We estimated two types of models: models A with the covariate *tr* for investigation of effects of any training in the past, and models B with covariates *trs* and *trl* for investigation of effects of training in the short- and the long-run.

Table 3. Results of Cox models estimation for participation in training

Models A								
Covariates	Naive				Control			
	Men		Women		Men		Women	
	HR	P>|z|	HR	P>|z|	HR	P>|z|	HR	P>|z|
age25	1,530	0,000	0,882	0,033	1,556	0,000	0,885	0,037
age2640	1,190	0,000	0,984	0,769	1,224	0,000	0,985	0,772
marr	1,316	0,000	0,910	0,035	1,352	0,000	0,914	0,048
edu1	0,699	0,000	0,540	0,000	0,479	0,000	0,526	0,000
edu2	0,844	0,046	0,704	0,000	0,646	0,000	0,662	0,000
town	1,109	0,007	1,053	0,232	1,110	0,007	1,050	0,258
disabled	0,548	0,000	0,788	0,039	0,552	0,000	0,786	0,037
benefit	0,629	0,000	0,630	0,000	0,636	0,000	0,635	0,000
*tr, tr**	1,447	0,000	1,216	0,097	1,827	0,000	1,169	0,157
	No. of obs. = 3,247 ln L = -20,872.4		No. of obs. = 2,951 ln L = -17,076.1		No. of obs. = 3,247 ln L = -20,859.6		No. of obs. = 2,951 ln L = -17,076.4	
Models B								
Covariates	Naive				Control			
	Men		Women		Men		Women	
	HR	P>|z|	HR	P>|z|	HR	P>|z|	HR	P>|z|
age25	1,529	0,000	0,881	0,031	1,541	0,000	0,848	0,006
age2640	1,190	0,000	0,982	0,734	1,219	0,000	0,980	0,713
marr	1,315	0,000	0,910	0,036	1,351	0,000	0,919	0,063
edu1	0,698	0,000	0,539	0,000	0,503	0,000	0,556	0,000
edu2	0,844	0,045	0,702	0,000	0,711	0,005	0,710	0,000
town	1,109	0,007	1,052	0,239	1,105	0,010	1,038	0,388
disabled	0,546	0,000	0,786	0,038	0,553	0,000	0,801	0,056
benefit	0,629	0,000	0,630	0,000	0,637	0,000	0,633	0,000
*trs, trs**	1,367	0,045	1,544	0,014	1,499	0,002	1,846	0,000
*trl, trl**	1,492	0,000	1,042	0,791	1,262	0,152	0,436	0,001
	No. of obs. = 3,247 ln L = -20,872.2		No. of obs. = 2,951 ln L = -17,074.7		No. of obs. = 3,247 ln L = -20,859.0		No. of obs. = 2,951 ln L = -17,069.4	

Source: own computations using Stata Statistical Software; HR – hazard rates

The first columns under the “Naive” heading were obtained by using the hazard function $h(t; X_{Yi}, P_i) = h_0(t)\exp(\beta X_{Yi} + \alpha P_i)$, where P_i denotes the participation in the training program, and contains estimates for hazard rates. The columns under the “Control” heading were obtained by using instead $h(t; X_{Yi}, P_i) = h_0(t)\exp(\beta X_{Yi} + \alpha IP_i)$, where instrument $IP_i = P_i^*$ denotes the index value obtained from the estimation (4) of the probit model. The index values are the right-hand sides of probit equations less the residuals (not the expected probabilities computed using the normal distribution) [Wodon, Minowa 2001].

The age coefficients imply that men older than 41 are at a disadvantage to find a job. Married women are at the disadvantage at the job market, but for the married men the effect is the opposite. Primary or secondary education level has a significant negative effect on the opportunity to break unemployment. The disabled have a significant lower reemployment chance. There is a greater tendency to leave the unemployment state if the registered person receives no unemployment benefit.

The naive estimates indicate that every training during the last three years reduces the length of unemployment. Training in the past has a positive effect on reemployment probabilities; the transition rate for men increases by about 45%, and for women increases by about 22%. There are positive impacts of training in the short-run for both men and women and in the long-run for men only.

When we use the index values from probit models instead of variable *tr* (see estimates in models A under the "Control" heading), we can detect still positive impacts in the case of both men and woman. These effects are greater for men and are smaller for women (although for women not statistically significant). In the short-run we still observe strong positive impacts on employment for both men and women (see models B under the "Control" heading). The recent training seems to provide the unemployed with modern knowledge which positively distinguishes them from the other unemployed when searching for a job. Surprisingly, in the long-run this effect is statistically insignificant for men; for women the impact is significantly negative.

CONCLUSIONS

The goal of our study was to prove the effectiveness of training programs directed to the unemployed on the local labor market. We estimated hazard models to assess the impact of vocational training on the duration of unemployment spells. To resolve the potential sample selection problem, the participation in a training program was instrumented using a probit model. The IV procedure provides an answer to the question to what extent the effect of the training program was a result of the effectiveness of the program and to what extent it was due to the fact that program participants had different characteristics than other unemployed. The empirical results obtained confirm that training courses proved not fully effective.

In particular, our results indicate that, mostly, there is a positive effect of the training in the past on reemployment probabilities. In the short-run this positive effect is much bigger and statistically significant for both men and women.

Surprisingly, we detect a negative impact of training for women in the long run. Such an effect can be called the stigmatization effect. The long-run effect for men is statistically insignificant. Other results obtained for the whole sample - not presented in the article - also show the lack of training impact in the long-run. This is called a deadweight loss effect and it occurs when a training participant would have reached the same result without participating in the program (it will be the case when company hires a subsidized employee but would also do so, if there is no subsidy) [Bukowski 2008].

REFERENCES

Bijwaard G. (2008) Instrumental Variable Estimator for Duration Data, Tinbergen Institute Discussion Paper, No. 032/4.

Blundell R., Dearden L., Meghir C. (1994) The Determinants and Effects of Work Related Training in Britain, Working Paper, Institute for Fiscal Studies, London.

Bowden R.J., Turkington D.A. (1984) Instrumental Variables, Cambridge University Press, Cambridge.

Bukowski M. (ed.) (2008) Employment in Poland 2007: Security on Flexible Labour Market, Ministry of Labour and Social Policy, Warsaw.

Cameron A.C., Trivedi P.K. (2005) Microeconometrics: Methods and Applications, Cambridge University Press, New York.

Cox D.R. (1972) Regression Models and Life Tables (with Discussion), Journal of the Royal Statistical Society, Series B 34.

Dehejia R.H., Wahba S. (1999) Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, Journal of the American Statistical Association, 94, pp. 1053-1062.

Heckman J.J., Ichimura, H., Smith, J.A. and Todd, P. (1998) Characterising Selection Bias Using Experimental Data, Econometrica, 66(5), pp. 1017-98.

Hosmer D., Lemeshow S. (1999) Applied Survival Analysis: Regression Modeling of Time to Event Data, John Wiley and Sons, New York.

Hujer R., Maurer K.-O., Wellner M. (1999) Estimating the Effect of Vocational Training on Unemployment Duration in West Germany – A Discrete Hazard-Rate Model with Instrumental Variables, Jahrbücher für Nationalölkonomie und Statistik, Vol. 218, No. 1.

Kalbfleisch J., Prentice R. (1980) The Statistical Analysis of Failure Time Data, John Wiley and Sons, New York.

Kluve J., Lehmann H., Schmidt Ch.M. (2000) Disentangling Treatment Effects of Polish Active Labor Market Policies: Evidence from Matched Samples, Centre for Economic Reform and Transformation, Discussion Paper No. 2000/07, Edinburgh.

Lalonde R.J. (1986) Evaluating the Econometric Evaluations of Training Programs, American Economic Review, 76(4), pp. 604–620.

Landmesser J. (2010) Ocena skuteczności szkoleń dla bezrobotnych z obszarów wiejskich, in: Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu, Zeszyt 5, Tom XII, pp. 109-114.

Puhani A.P. (1998) Advantage through Training? A Microeconometric Evaluation of the Employment Effects of Active Labour Market Programmes in Poland, ZEW Discussion Paper No. 98-25.

Wodon Q., Minowa M. (2001) Training for the Urban Unemployed: A Reevaluation of Mexico's Training Program, Probecat, World Bank Economists' Forum, Washington D.C., pp. 197-215.