Empirical Economics (2003) 28:515–533

EMPIRICAL
ECONOMICS

Estimating the employment effects of education for disabled workers in Norway

Arild Aakvik[1]

Department of Economics, University of Bergen, Fosswinckelsg. 6, N-5007 Bergen, Norway.
(E-mail: arild.aakvik@econ.uib.no)

First version: Jan 2001/Final version received: April 2002

Abstract. In this paper we evaluate the effectiveness of educational programmes used as an employment strategy for disabled workers in Norway. To obtain these estimates we follow the employment career of a sample of participants in educational programmes and nonparticipants three years after they had left the vocational rehabilitation benefit system. We specify an employment outcome model that includes both unobserved heterogeneity and selection bias due to correlation between unmeasured factors and a person's training status. Even though participants in educational programmes have employment rates that are around eight percentage points higher than those who did not participate in such programmes, econometric selection models produce a training effect for education not significantly different from zero.

Key words: programme evaluation, education, disability, employment

JEL classification: I210, J240, J310

1. Introduction

Concern about the wide gap in educational level and general skills between partly disabled and non-disabled workers has resulted in an increasing use of ordinary education as a vocational rehabilitation strategy.[2] The main purpose

---

[1] The data were provided by The Norwegian Social Data Service in Bergen. I would like to thank Dag Kiberg for preparing the raw data. All the data are gathered from sources at The Directorate of Labour, The Social Insurance Organization and Statistics Norway. None of the mentioned institutions are in any circumstances responsible for the analysis or for the conclusions drawn from it.

[2] The level of education of ordinary unemployed is also substantially higher compared to vocational rehabilitation clients. Only 50 percent of the vocational rehabilitation (VR) clients have high school or more, while the same number is 75 percent for ordinary unemployed. Around 85 percent of the labour force in Norway have high school or more.

of both educational and other training programmes for partly disabled workers is to enhance participants' human capital and productive skills and, in turn, increase their employment prospects. Although education is the most commonly used vocational rehabilitation programme in Norway, little research has been done to evaluate its effectiveness and economic impact.[3] We focus on this deficiency and examine the effects of general training (education) on employment among partly disabled workers.

Our longitudinal data allow us to use several different approaches to compare the employment outcomes of participants and the employment outcomes of eligible but nonparticipating individuals used to proxy the outcome of participants had they not participated. We specify employment outcome models without invoking distributional assumptions about omitted variables, and model potential correlation between omitted variables and education using an index-sufficient representation.

Our data consist of a random sample of 1506 persons who entered the vocational rehabilitation benefit scheme in 1989 and left it before 1991. Besides the benefit given to disabled people who have previously been employed, the Norwegian vocational rehabilitation system offers active training to employ persons with various disabilities. Members of our training group consist of individuals who participated in educational training programmes. These clients are integrated into ordinary classes in the public school system, or attend classroom training aimed specifically towards unemployed people. The comparison group consists of individuals who received rehabilitation benefit, and were directed to and had applied for training, but did not participate in active VR training.[4,5] The outcome for untrained persons is used to proxy the outcome for participants had they not participated in the programme. We focus on the average treatment effect on the treated, sometimes denoted TT, since this is usually the most interesting policy parameter. This treatment parameter focuses on the effect of training on the trainees. To obtain the estimates of the effect of education we follow the employment career of participants and nonparticipants for four years after they leave the vocational rehabilitation system.

In non-experimental evaluation studies, persons who receive education are not necessarily comparable to nonparticipants even in the hypothetical case where the participants did not participate in the programme. Using nonparticipants to proxy participants in the case where participants had not been educated runs the risk of producing selection bias in estimates of the effect of education due to differences in observed and unobserved factors between the two groups. The solution to selection bias due to observable characteristics (in the data) is to include these variables in regression analysis. With longitudinal

[3] For the US economy, Dean and Dolan (1992) analyse the effectiveness of *higher* education for persons with work disabilities. Some evaluation studies on the overall effectiveness of vocational rehabilitation programmes using a comparison group approach has been done for the US economy, for instance Nowak (1983), Worrall (1988), Dean and Dolan (1991). None of these studies look at the effects of education on the movement in and out of employment.

[4] The decision to accept a person for training, given that the person has applied, is mainly made by case workers. However, there can be a substantial time lag between acceptance and the startup of training, giving room for "drop outs" due to lack of motivation, employment, or other reasons.

[5] Our comparison group consists of persons not participating in a training programme more broadly defined than education. We do not know how many persons in the comparison group applied for educational programmes.

data and a binary outcome variable this corresponds to estimating a random-effects probit model where dependence in successive employment probabilities for each person is internalized. In this model unobserved heterogeneity is assumed to be normally distributed and independent of explanatory variables.

Selection bias due to unobserved variables implies that at least one individual-specific factor not observed in the data is correlated with a person's training status and employment outcome. One potential form of selection bias due to unobserved variables may be that trainees have higher motivation or ability than nonparticipants, and that motivation or ability is an important factor for the employment outcome after training. It may also be the case that motivated people have higher opportunity cost and thus are more eager to get employed without participating in an employment programme.

Several different approaches to correct for selection bias have been discussed in the literature. Heckman (1979) proposed a two-stage procedure that can be used to correct for non-random selection into training programmes. The original model relies on the assumption of a bivariate normal distribution between the error terms. In this paper we reject the assumption of normality, and are thus forced to use other, less developed, methods to estimate the effect of educational programmes. Recent developments in models of selection bias are related to the classical econometric selection model of Heckman, but are based on non-parametric and semi-parametric estimation of the selection equation and the outcome equation. These estimation techniques offer more flexibility in analysing critical parts of a regression relationship, for instance non-normality, although at the cost of poorer efficiency compared to a correctly specified parametric model. However, a parametric model runs a greater risk of model bias due to misspecification. Flexible estimators require fewer modeling assumptions. See, for instance, Vella (1998) for a recent overview of selection models using semi-parametric estimation strategies, and Ichimura and Todd (2001) on how to implement non-parametric and semi-parametric estimators.

In this paper we estimate the selection equation using a kernel regression method. To capture potential selection bias due to correlation between unobserved variables and a person's training status we include a smoothed single-index variable estimated from the selection equation in the outcome regression. The employment equation is estimated using a semi-parametric approach based on a mixture representation of unobserved heterogeneity.

Descriptive statistics reveal that participants in educational programmes have substantially higher success rates in the labour market compared to the employment outcome of eligible individuals not participating in active training. Mean employment rates for persons participating in educational programmes are around eight percentage points higher than employment rates experienced by nonparticipants. However, we find that persons participating in training programmes are more likely to be employed even without the training programme than persons not participating in such programmes. Adjustment for observed discrepancies between trainees and nonparticipants reduces the estimated average training effect on the treated to less than three percentage points. In our selection model we find that the introduction of unobservables correlated with the education dummy reduces the estimated training effect even more, indicating creaming also on unobservables, although the magnitude is very small and imprecisely estimated.

We organize the rest of the paper in the following way. In the next section we introduce the data and institutional settings. In section 3 we discuss the selection process, and estimate a simple probit model to see if trainees are statistically different on observed variables from members of the comparison group. In section 4 we first present some simple cross-sectional and longitudinal estimates of the effects of education on the employment probability, where we control for observed differences between trainees and members of the comparison group. In section 5 we relax the assumption that the effects of omitted variables are randomly distributed and independent of background variables. Section 6 summarizes our findings.

2. Data and institutional settings

The available data are drawn from several independent databases. The data contain information on socioeconomic background variables and labour market participation for a random sample of the Norwegian population. The present available observation period is five years, from the beginning of 1989 to the beginning of 1994.

Vocational rehabilitation is a social insurance scheme, where individuals receive a benefit after a period, usually of 52 weeks, on sickness leave. Not everyone who receives this benefit participates in active vocational training programmes. We use an internal comparison group, that is, individuals who were directed to and applied for a training programme, but did not receive active training as part of their rehabilitation effort, either due to self-selection out of the programme or due to supply side restrictions. Bell et al. (1995) offer several arguments in favour of internal comparison groups, see also Heckman et al. (1998). Internal comparison groups are often favoured since they usually mimic the treatment group better than other types of constructed comparison groups. In internal comparison groups, nonparticipants are located in the same labour market as participants. Heckman et al. (1998) have shown that failure to match within local labour markets can be an important source of bias in evaluation studies.

Our sample consists of persons who had a documented medical diagnosis making them eligible for vocational rehabilitation, and who applied for participation in a training programme in 1989. We focus on the outcome of individuals discharged from vocational rehabilitation before 1991. We can observe retrospectively which clients applied for and participated in training, and what type of training they received. Individuals who participated in other types of training than education, such as sheltered work, wage subsidies, employment schemes in the public sector, etc., are left out from the sample since we choose to focus on the effect of educational programmes.[6] For a

[6] It might be the case that people who have participated in an educational programme as part of their rehabilitation effort are more likely to continue their education with their own funds after the end of vocational rehabilitation. It is difficult to estimate the effect of education as part of the VR programme if some persons continue with their education elsewhere, since we are not able to identify the separate effect of education as part of the VR effort and their self-financed education. If individuals continue education elsewhere, the estimated effect of education is likely to be downward biased.

detailed analysis of the selection into different training programmes and the effect of education compared to work related training, see Aakvik and Kjerstad (2002).

The decision to accept a person into a training programme is mainly taken by case workers at the Employment Office and local managers of vocational rehabilitation centres. This decision is usually based on subjective judgement regarding employment prospects. Guidelines emphasize that an "evaluation of the clients' total situation in each case should be considered when a participation decision is made. Main inclusion criteria are health, age, personal characteristics, social conditions, education and labour market experience." The candidates for training participation may themselves also influence the participation decision by supplying subjective information to the programme administrator and case workers.

Disabled individuals participating in educational programmes are usually integrated into regular classes run by the public school system. The vocational rehabilitation service varies greatly in substance and duration across clients, reflecting a diverse clientele and the broad orientation of vocational rehabilitation. Unfortunately, we are not able from the data to distinguish among the different types of education and classroom training that each individual receives. Thus education is a dummy variable taking the value one if a person participates in an educational programme, and zero otherwise. All expenses are covered by the social security system. The vocational rehabilitation benefit, which amounts to about 60 percent of previous income, ceases upon return to work.

Before proceeding, we examine the characteristics of participants in educational programmes and nonparticipants in our sample. Table 1 contains descriptive statistics on central variables in our empirical analysis.[7]

A simple measure of the effect of education as a re-employment strategy for disabled workers would be to compare employment histories of participants and nonparticipants. Table 1 shows that members of the treatment group have higher employment rates compared to nonparticipants for all subsequent years. Individuals participating in educational programmes have a 7 percentage point higher employment rate than nonparticipants one year after having completed training. The employment rate among those participating in education programs increases by three percentage points from 1991 to 1992 and is constant from 1992 to 1993.[8] Nonparticipants have the same pattern, but with only a two percentage point increase.

Our simple measure of treatment effects discussed above is hampered by the lack of controlling for observed and possible unobserved individual characteristics. Failure to control for such differences leads to bias in the estimate of training impacts, see for instance Heckman and Hotz (1989). Table 1 reveals several discrepancies in observed characteristics between trainees and comparison-group members. For instance, women are over-represented

[7] Missing observations are not a major problem in our analysis. The same persons are followed over time. We can thus fill in for many missing observations using data for different years. Less than 1 percent of the observations in the sample were deleted due to missing observations.

[8] A chi-square test of no training effect reveals a test statistic significant at the 1 percent level, indicating that the training effect is positive. However, this test assumes that the treatment group and the comparison group are drawn randomly from a population, which is not likely to be the case in nonexperimental data.

Table 1. Sample characteristics of participants and nonparticipants

	Educational programmes	Non-participants
Number of individuals	277	1229
Employment rates, percent, 1991	45	38
Employment rates, percent, 1992	48	40
Employment rates, percent, 1993	48	40
Male, percent	45	52
Age, 1989	31.6	35.7
Work experience in years, 1989	8.4	10.2
Work experience in years, 1993	10.4	12.1
Mean income of spouses in NOK, 1989	59432	58540
Mean income of spouses in NOK, 1993	79973	71049
Percent with children	30	28
Married, percent, 1989	32	41
Unemployment rate, 1989[a]	3.60	3.39
Unemployment rate, 1993	4.40	4.38
Education in years, 1989	10.58	9.94

[a] We have calculated the unemployment rate as the total number of unemployed individuals multiplied by 100 and divided by the total number of individuals in the age interval between 16 and 67 in the municipality (kommune). In official statistics the number of unemployed individuals is divided by the number of individuals in the labour force, producing higher rates than those reported here.

in educational programmes. Furthermore, participants in educational programmes are more than 4 years younger than members of the comparison group. They also have more years of schooling in the beginning of our observation period, but less experience in the labour market.

The revealed differences in descriptive statistics between the two groups are expected to influence the impact of training. In the next section we analyse the selection process more formally using a probit model.

3. Selection of participants into educational programmes

We analyse the selection of participants into educational programmes using a simple probit model. The selection of individuals into training may be a complicated process influenced by the client's own decisions, institutional eligibility criteria, subjective judgements of local programme administrators or case workers, and by random components such as "first-come, first-serve". Given the relatively vague criteria for selection and the large degree of discrepancy among case workers we would expect unobserved (to the econometrician) variables to have a strong influence on the selection outcome and, potentially, the employment outcome as well. An ideal situation for analysing the selection process would require a lot of information on case workers, health indicators and local labour markets, which is usually not available.

Whatever the nature of the selection process into training, it can be described in terms of an index function. Let d_i^* be an index, which is assumed to be a linear function of observed (z_i) and unobserved (v_i) variables. The probit model takes the following form

$$\begin{aligned} d_i^* &= \gamma' z_i + v_i \quad v_i \sim N[0,1] \\ d_i &= 1 \quad \text{if } d_i^* > 0, \text{ and 0 otherwise,} \end{aligned} \tag{1}$$

where d_i is a dichotomous variable taking the value one if person i is a participant in an educational programme, and zero if the person is a member of the comparison group. z_i is a vector of explanatory variables, including age, work experience, gender, years of education prior to training, income of spouse, local unemployment rate, medical diagnosis, as well as a constant term. All the variables are measured in 1989. γ is a set of parameters that reflect the effect of changes in background variables on the training index. v_i is assumed to be a standard normally distributed error term leading to a probit model,

$$\text{Prob}(d_i = 1) = \text{Prob}(v_i > -\gamma' z_i) = \Phi(\gamma' z_i), \tag{2}$$

where Φ denotes the standard normal cumulative distribution function (c.d.f.).

The probit model estimates the effect of background variables on the index function. Discussing coefficients in terms of marginal effects is often more interesting. The effect of a unit change in a continuous variable z_{ik} on the probability of being a training participant, where z_{ik} is the kth element of the p-vector of explanatory variables z_i, can be calculated by the following equation

$$\frac{\partial E(d_i|z_i)}{\partial z_{ik}} = \frac{\partial \Phi(\gamma' z_i)}{\partial z_{ik}} = \phi(\gamma' z_i)\gamma_k, \tag{3}$$

where ϕ is the standard normal probability density function (p.d.f.). Taking the sample average, i.e. $\frac{1}{n}\sum_{i=1}^{n} \phi(\gamma' z_i)\gamma_k$ gives the reported marginal effects. Thus the marginal effects are calculated as the mean of the analytic derivative calculated for each person (mean derivatives). Marginal effects for dummy variables are calculated as the change in $\Phi(\cdot)$ when the dummy variable goes from 0 to 1 for all observations. The average of these numbers gives the reported marginal effects for dummy variables.

We report the coefficients and standard errors for the selection regression of equation (1) in Column 1 in Table 2, while marginal effects are reported in Column 2.

The Chi-squared test shows that the selection model is significant as measured against the same model, with no explanatory variables. This means that trainees in educational programmes differ significantly from nonparticipants with respect to important observable variables. We have also included 13 dummy variables that indicate the medical diagnosis of each person in the sample. Most of these were not statistically different from zero in the regression. To save space we have omitted these numbers from the tables.

Individual characteristics such as youth, higher level of education (before training), and more working experience significantly increase the probability of participating in educational training programmes. So does a higher level of

Table 2. Selection of participants into educational programmes

	Probit estimates 1		Marginal effects 2
Constant	−1.145**	(0.433)	−0.284
Work experience	0.015*	(0.009)	0.033
Spouse's income/100000	0.061	(0.111)	0.003
Married	−0.193	(0.134)	−0.045
Age	−0.028**	(0.007)	−0.006
Male	−0.263**	(0.092)	−0.063
Children	−0.108	(0.098)	−0.024
Education	0.124**	(0.024)	0.033
Unemployment	0.048*	(0.025)	0.012
Prob $(d_i = 1 \mid z_i)$	16.2%		
Number of observations	1506		
Log-Likelihood	−660		
Chi-squared	119		

Note: Numbers in parentheses are standard errors. ** Indicates significant at the 1 percent level, while * indicates significant at the 5 percent level, both for a two-sided test of population coefficients equal to zero. To save space we have omitted 13 dummies indicating individual medical diagnosis. We have tried specifications with interaction terms and with some variables entered non-linearly. None of these specifications turned out to improve the fit of the model or to give new insight.

unemployment in the local district. Finally, the chances of participating are higher for women than for men.[9]

4. Models for the effect of education on employment

4.1. Controlling for observed differences between the treatment and comparison group

The effect of training can generally be decomposed into an effect on wages, working hours, employment, and time of employment after training. In Norway, employment is the clearly defined objective of vocational rehabilitation, pointed out in several governmental reports. Card and Sullivan (1988), and LaLonde (1995) also point out that most of the earnings gains reported in the literature follow from higher employment rather than increased wages. In this paper we focus on employment as our measure of success (or lack therof) of training. We start with estimating simple cross-sectional and longitudinal

[9] The selection process into vocational rehabilitation programmes is different from the selection process into labour market programmes for ordinary unemployed individuals. Unemployed individuals may opt to collect a training allowance if they are not entitled to unemployment benefits. A person is entitled to unemployment insurance if she has been attached to the labour market for some time prior to the occurrence of unemployment. Unemployed persons marginally attached to the labour market are thus likely to be over-represented on training programmes since they receive a training allowance during the training period. This type of self-selection is less likely to be the case in VR training programmes, since most participants rely on benefit schemes other than the training allowance. See the discussion in Aakvik (2000).

probit models of the effects of education as an employment strategy for disabled workers. These estimates will give us a benchmark to which later models will be compared. Later we specify models that are flexible in capturing the effects of heterogeneity in unobserved variables, and that control for correlation between omitted variables and a person's training status.

As a starting point, we specify models that adjust for observed differences that exist between the trainees and members of the comparison groups. Here we assume that the employment outcome of trainees and nonparticipants are stochastically independent of a person's training status given that we control for differences in observed variables.

Longitudinal data introduce the problem of serial correlation in the subsequent employment probabilities over time for each person. Employment rates are assumed to be a function of explanatory variables with regression coefficients that vary from one individual to the next. We assume that this variability reflects individual heterogeneity due to omitted variables, rather than state dependence. The structural random-effects probit model for longitudinal data may be written

$$y_{it}^* = \beta_t' x_{it} + \mu_t d_i + \varepsilon_{it}$$

$$y_{it} = 1 \quad \text{if } y_{it}^* > 0, \text{ and 0 otherwise,} \tag{4}$$

where y_{it} is the observed dichotomous response variable taking the value one if person i is employed at time t, and zero if the person is not employed. A person is defined as employed if she has an employment spell of more than 60 days in a given year. The job must at least be a regular part time job. x_{it} is a vector of explanatory time varying and time invariant variables including a constant term, d_i is a dummy variable taking the value one if a person has participated in an educational programme, and zero otherwise, and μ is the training effect. β is a parameter vector that reflects the effects of the changes in background variables on the employment index. ε_i is a standard normally distributed error term.

If ε_{it} is an independent standard normal variable, the longitudinal structure of the data is irrelevant, and we may apply a standard pooled probit model to estimate the effect of training. To allow for time-varying coefficients we also estimate a cross-sectional version of the probit model. The cross-sectional probit models are more general than the panel data model we estimate in this paper since in the panel data probit model with and without random effects we assume that the regression coefficients are fixed over time. Lechner (1995) provides a framework for testing the assumption of the programme effect being time constant.

The longitudinal data have an inherent problem of correlation between the dependent variable over time for each individual due to omitted variables. Let $\varepsilon_{it} = c_i + u_{it}$, where c_i captures individual specific unobserved characteristics, and u_{it} is meant to account for the purely stochastic aspects of the model. Since u_{it} is stochastic, we have

$$\mathrm{Cov}(u_{it}, x_{it}) = \mathrm{Cov}(u_{it}, d_i) = \mathrm{Cov}(u_{it}, c_i) = E(u_{it}) = 0$$

$$\mathrm{Cov}(u_{it}, u_{is}) = \sigma_u \quad \text{if } t = s \quad \text{and} \quad \mathrm{Cov}(u_{it}, u_{is}) = 0 \quad \text{if } t \neq s. \tag{5}$$

These are assumptions often made in panel data models. Where our assumptions differ from the usual ones is in the individual component. A first approach starts with the assumption that c_i is a random variable with mean zero and variance σ_c^2. In this model selection is assumed to be captured by observed variables, that is, selection bias due to correlation between unobserved variables and training status is absent. This model is used as a benchmark to which other specifications can be compared. c_i is removed from the model by integrating the likelihood function over the normal probability density function. It is assumed that there is a positive correlation between any two measurements for the same individual. The within person correlation takes the following simple form

$$\text{Corr}(c_i + u_{it}, c_i + u_{is}) = \rho = \sigma_c^2/(\sigma_u^2 + \sigma_c^2), \tag{6}$$

where $t \neq s$. For $t = s$ the correlation is one. This is the Heckman and Willis (1976) uniform correlation model for which Butler and Moffitt (1982) have made an efficient computer algorithm. The estimate of the serial correlation component ρ gives a test for the appropriateness of the random-effects model versus the specification that does not model the within-subject correlation, and is reported together with the coefficient estimates. For different panel probit models and their properties see, for instance, Bertschek and Lechner (1998).

The econometric framework leads to different probit models where the effect of training can be evaluated on employment outcomes. The treatment effect for a given individual is the difference in employment outcome in the treated state and the employment outcome in the untreated state. However, this person-specific treatment effect is a counterfactual, and can rarely be estimated. Instead, it is more common to work with averages.[10] We write the probability of employment as, $\text{Prob}(y_{it} = 1) = \Phi(\beta_t' x_{it} + \mu_t d_i)$. The average effect of treatment (ATE) within this framework is simply

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n}[\Phi(\beta_t' x_{it} + \mu_t) - \Phi(\beta_t' x_{it})] \tag{7}$$

where n is the total sample used in this paper. It should be noted that this parameter is not the training effect for a random person in the general population. However, it is the training effect on employment for a random person in our sample. Since we use an internal comparison group, this parameter is not very different from the effect of treatment on the treated (TT). The TT parameter is calculated using only training participants in a similar form as in equation (7). Standard errors for the treatment effects are calculated using bootstrap.

4.2. *Empirical results*

We report the employment regressions in Table 3.1, where we control for observed differences between trainees and members of the comparison group.

[10] For a thorough discussion of different treatment parameters see, for instance, Heckman (1997). An econometric framework for estimating distributional treatment effects on discrete outcomes is given in Aakvik et al. (2002).

Table 3.1. Probit regressions of employment

	Probit model 1991 1	Probit model 1992 2	Probit model 1993 3	Pooled probit model 1991–1993 4	Random effects probit model 1991–1993 5
Constant	1.245*	1.334*	1.278*	1.158**	1.415*
	(0.507)	(0.531)	(0.551)	(0.301)	(0.790)
Work experience	0.150**	0.138**	0.133**	0.139**	0.374**
	(0.020)	(0.020)	(0.019)	(0.011)	(0.032)
Work exp. squared/10	−0.027**	−0.019**	−0.015*	−0.020**	−0.053**
	(0.007)	(0.007)	(0.007)	(0.004)	(0.011)
Spouse's income (log)	0.027**	0.032**	0.028**	0.029**	0.062**
	(0.009)	(0.009)	(0.008)	(0.005)	(0.013)
Married	0.067	0.005	0.015	0.029	0.165
	(0.101)	(0.100)	(0.099)	(0.058)	(0.140)
Age	−0.136**	−0.131**	−0.123**	−0.125**	−0.224**
	(0.026)	(0.027)	(0.027)	(0.015)	(0.041)
Age squared/100	0.111**	0.088**	0.073**	0.084**	0.010*
	(0.031)	(0.032)	(0.032)	(0.018)	(0.005)
Male	−0.514**	−0.448**	−0.494**	−0.486**	−1.348**
	(0.080)	(0.079)	(0.080)	(0.046)	(0.135)
Education	0.087**	0.085**	0.096**	0.089**	0.212**
	(0.021)	(0.021)	(0.021)	(0.012)	(0.035)
Children	−0.181*	−0.248**	−0.233**	−0.221**	−0.435**
	(0.087)	(0.087)	(0.087)	(0.050)	(0.142)
Local unemployment	−0.013	−0.022	−0.004	−0.013	−0.002
	(0.039)	(0.036)	(0.037)	(0.021)	(0.046)
Training effect:					
Education (d_i)	0.062	0.045	0.038	0.048	0.039
	(0.039)	(0.036)	(0.037)	(0.051)	(0.142)
Correlation (ρ)					0.836**
					(0.036)
Training effect (percent)[a]	2.4	1.8	1.5	1.9	1.3
	(0.018)	(0.016)	(0.013)	(0.019)	(0.015)
Number of observations	1506	1506	1506	4518	4518
Log-Likelihood	−916	−910	−893	−2728	−1991
Chi-squared statistics	190	223	260	660	1474

Note: Numbers in parentheses are standard errors. ** Indicates significant at the 1 percent level, while * indicates significant at the 10 percent level, both for a two-sided test of population coefficients equal to zero.

[a] The training effect reported here is the effect of treatment on the treated (TT). Standard errors are calculated using bootstrapping.

Members of the comparison group are used to proxy the outcome of participants had they not participated. Columns 1–3 report the estimates of the effect of training in 1991, 1992 and 1993, respectively, using cross-sectional estimation techniques. In Column 4 we use a standard probit model on the pooled data to estimate the effect of training, where we assume that ε_{it} is an independent standard normal variable. In Column 5 we report the results from the random-effects probit model stated in equation (4), incorporating serial correlation in subsequent employment outcomes. Marginal effects are

Table 3.2. Marginal effects in probit regressions of employment

	Probit model 1991 1	Probit model 1992 2	Probit model 1993 3	Pooled probit model 1991–1993 4	Random-effects probit model 1991–1993 5
Constant	0.481	0.521	0.498	0.451	0.469
Work experience	0.056	0.052	0.051	0.053	0.121
Work exp. squared/10	−0.010	−0.009	−0.007	−0.009	−0.017
Spouse's income (log)	0.011	0.013	0.012	0.012	0.019
Married	0.028	0.003	0.007	0.012	0.056
Age	−0.052	−0.051	−0.047	−0.048	−0.074
Age squared/100	0.042	0.034	0.028	0.032	0.034
Male	−0.198	−0.173	−0.190	−0.189	−0.442
Education	0.034	0.034	0.037	0.035	0.071
Children	−0.065	−0.093	−0.088	−0.083	−0.139
Local unemployment	−0.006	−0.007	−0.004	−0.006	−0.002
Training (d_i) (ATE)	0.026	0.019	0.016	0.021	0.014

reported in Table 3.2. In the cross-sectional regressions in Columns 1–3 we estimate employment in each year as a function of background variables including a dummy for participation in educational programmes. The dependent variable is coded 1 if a person was re-employed in a specific year, and zero otherwise. A pooled probit model and a random-effects probit model are used to estimate the effect of training on employment simultaneously on all three years following the end of training.

Our cross-sectional probit models of employment effects of education show point estimates varying from 2.4 percentage points in 1991 to 1.5 percentage points in 1993. We cannot reject the null hypothesis that the program effects are jointly zero. These treatment effects are the effect of treatment on the treated (TT). The average treatment effects (ATE) reported in Table 3.2 are slightly higher than the estimated TT-parameters. This suggests that the case workers are not doing a very good job of sorting persons with the largest expected gain into the program. Although the estimated effects of education are positive, they are not statistically different from zero. Estimating a random-effects probit model, where data from 1991 to 1993 are pooled, improves the fit of the model, but the training dummy remains insignificantly different from zero.

Other variables behave much as expected, and most of them are significantly different from zero. The effect of work experience on the probability of employment is increasing and concave, and higher spousal income increase the probability of getting a job. The effect of age is similar to other Norwegian studies of vocational rehabilitation, where younger individuals have better chances of getting a job, everything else equal; see Aakvik and Risa (1998). Women have better chances of being employed compared to men, mainly because the relatively high fraction of part-time employment among women in our sample.[11] The effect of education is positive and significant on the prob-

[11] Using only full time employment as our success measure does not change the results much except for the gender variable. Male clients have better chances of getting a full time job than female clients.

ability of employment. Having a child reduces the employment probabilities. Higher unemployment rates in local districts also reduce the probability of employment. However, this coefficient is not statistically significantly different from zero.

The within-person correlation in the dependent variable stated in equation (5) is significant and estimated to equal 0.84, as reported in Table 3. This is due to relatively stable employment relationships given that a person gets a job. In a random-effects model where we used only 1991 and 1993 data, within-person correlation was reduced to 0.45. Removing the 1992 data had no impact on the estimated coefficients or standard errors. However, the change in the estimated correlation when the 1992 data are dropped is not fully consistent with the simple random effects probit model, where it is assumed the correlation is time constant.

Comparing the results of Table 2 and Table 3 shows strong indications that the observed selection process is consistent with a hypothesis of "creaming", in the sense that persons most likely to be employed are over represented in training programmes. For instance, young, highly educated female individuals with high work experience are over-represented as programme participants. These individuals also have significantly higher employment rates whether they participate in educational programmes or not. An exception is the local unemployment rate, which increases the probability of participating in training, but decreases the probability of employment. The relatively strong drop in the difference in employment rates after adjustment for observed characteristics also indicates that creaming takes place in the Norwegian vocational rehabilitation sector.

5. Unobserved heterogeneity and selection bias

5.1. Nonparametric specification of unobserved heterogeneity in employment outcomes

Table 2 revealed a strong dependence between the education dummy variable (d_i) and several background variables. This is selection on observable variables, and is eliminated by including the vector of observed variables in the employment regressions, see Goldberger (1972). However, the models reported in Table 3.1 may be flawed by not allowing for a flexible specification of unobserved population heterogeneity. The random-effects model reported in Table 3 assumes that the effect of unobservables is normally distributed.

Economic theory has in most cases little to say about the functional form of individual-specific effects. The choice of the normal distribution of c_i in panel data models is usually based on convention. Heckman and Singer (1984) show that structural parameter estimates are sensitive to the specification of the error term distribution within the framework of a single-duration Weibull model. To capture potential non-normality we allow c_i to follow a discrete distribution with a small number of mass points. Contrary to the standard random-effects probit model presented in Table 3.1, the mass point approach does not assume any parametric known distribution for c_i. Both the mass points and the associated probabilities are parameters of the probit likelihood function, so the procedure jointly estimates the distribution of unobservables and the structural parameters of interest. The nonparametric estimation of the unknown distribution of unobservables has its origins in Kiefer and Wolfowitz (1956), and was extended by Laird (1978) and Lindsey (1983a, 1983b). For a description of the maximizing algorithm, see Cosslett (1983). For a single-spell duration model, see Heckman and Singer (1984).

Let the probit model for our data be

$$\mathrm{Prob}(y_{it} = 1 \mid x_{it}, d_i, c_i) = \Phi(\beta' x_{it} + \mu d_i + c_i), \tag{8}$$

where the x_{it}-vector does not include a constant term. The heterogeneity term c_i is assumed to be independent of observed background variables and training status.[12] We approximate the unknown distribution of c_i by a number k of mass points

$$\alpha_1, \ldots, \alpha_k \quad k \geq 1 \tag{9}$$

with

$$\mathrm{Prob}(c_i = \alpha_j) = \pi_j \quad j = 1, \ldots, k; \quad \sum_{j=1}^{k} \pi_j = 1 \tag{10}$$

where α_j are the mass points of the mixing distribution and π_j their corresponding probabilities. The marginal likelihood function of individual i can be written

$$\mathscr{L}_i = \sum_{j=1}^{n} \pi_j \prod_{t=1}^{T} [\Phi(\beta' x_{it} + \mu d_i + \alpha_j)]^{y_{it}} [1 - \Phi(\beta' x_{it} + \mu d_i + \alpha_j)]^{1-y_{it}}. \tag{11}$$

The model with one mass point is equal to the pooled probit model since the single mass point represents a constant term.

5.2. Selection bias due to correlation between unobserved variables and training status

Selection on unobservables is present if unmeasured characteristics of that individual affect both the training participation decision and the employment outcome of an individual. In our study we account for the selection process into training using an index-sufficient representation, where the mean selection bias is represented by an index function rather than by the full vector of background characteristics. The procedure we use here is motivated by the two-equation selectivity approach of Heckman (1976, 1979). Our approach, however, avoids relying on normality in the specification of the error terms in the selection and employment equations. In the first stage we estimate the conditional mean of the training variable using the method of kernels. The kernel procedure secures a smooth index. In the second stage we include the estimated

[12] Linear models with unobserved heterogeneity independent of the error term give unbiased estimates of structural parameters. It is the nonlinearity that creates bias in estimating structural parameters if c_i is omitted. See the discussion by Cameron and Heckman (1998).

single-index variable in the semiparametric employment outcome regression, which is used to estimate the training effect for participants.[13] A simple single-index training model can be written in the following form

$$y_{it}^* = \beta' x_{it} + \mu d_i + \lambda(d_i^*) + c_i + u_{it}, \tag{12}$$

where we have partitioned the error term into two components, one that is individual-specific not varying over time, and one purely random term that may vary over time. It is assumed that the correlation between a person's training status and unobservables arises through the error term in the selection equation. $\lambda(\cdot)$ is an unknown smooth function of the single-index d_i^*, see for instance Ahn and Powell (1993). The single index is the mean of the observable training outcome variable d_i given some vector of conditioning selection variables z_i, that is

$$d_i^* = d_i^*(z_i) = E(d_i | z_i). \tag{13}$$

It is assumed that u_{it} is a mean-zero random variable and that the distribution of unobserved heterogeneity c_i in the employment outcome can be approximated by a number of mass points of finite support. Substitution of (13) into (12) yields the following reduced form

$$y_{it}^* = \beta' x_{it} + \mu d_i + \theta(z_i) + c_i + u_{it}, \tag{14}$$

where $\theta(z_i) = \lambda(d_i^*(z_i))$.

The single-index variable d_i^* is estimated by nonparametric methods using a kernel regression estimator. Let h be the smoothing (bandwidth) parameter. The cross validation mean squared prediction error (CVMSPE) is a goodness of fit measure for the kernel method. We have based the choice of h on the maximum of CVMSPE, which gave a bandwidth parameter of 0.2. Let the conditional mean of the training dummy be $s_i = \gamma' z_i$.[14] The weighting function for each observation is

$$\pi_i(s) = K[(s_i - s)/h], \tag{16}$$

where $K(\cdot)$ is the kernel function. $K(\cdot)$ tends to zero as the distance between s and s_i increases, that is, less weight is given if two observations are far from each other. Given these weights, the smoothed selection regression function is computed as

[13] We also used other indexes, namely the propensity score estimated by a probit model, and the inverse Mills ratio for participants and nonparticipants also from the probit selection equation. We also explored the bivariate probit model. Results from these models are available upon request. However, we rejected the normality assumption in the probit selection equation using a conditional moment test. In this test the third and fourth moments are regressed on a constant term and the score of the likelihood function. Results from the normality tests are available from the author.

Aakvik et al. (2002) use a full information maximum likelihood procedure where unobserved heterogeneity in both the selection equation and the employment equation is formulated in terms of a mixture distribution, which avoids relying on normality assumptions.

[14] We use Manski's maximum score estimator of γ; see Manski (1985). The score estimator gives consistent estimates of the parameter vector.

$$\hat{d}_i^*(s) = \left[\sum_{i=1}^{n} \pi_i(s) \cdot d_i\right] \cdot \left[\sum_{i=1}^{n} \pi_i(s)\right]^{-1}. \tag{17}$$

Our final estimator in equation (14) is a hybrid between a parametric and a non-parametric selection model, where the probit specification represents the parametric part and the selection and heterogeneity the non-parametric part. The standard errors are estimated by bootstrap methods. Three methods are explored; normal approximation, percentile, and bias-corrected. The three methods gave approximately the same results, and we report the results using the normal approximation with 100 replications in the text.

The z_i-vector in equation (13) consists of the same variables as the x_i-vector in equation (14). Thus the classical identification problem applies. In particular, without valid instruments the identification rests on a functional form assumption. In this paper, the decision to go into training and the recording of the employment outcomes take place at different periods. Several background variables are time-varying. Time varying-variables can be used as exclusion restrictions to identify the structural parameters, offering a solution to a problem that plagues many datasets; see, for instance, Heckman et al. (1998).[15]

5.3. *Empirical results*

We have estimated the binary probit employment model with discrete mixture distribution and correction for selectivity using a single-index sufficient representation. In Columns 1 and 2 in Table 4 we show the results for the one mass point model (pooled probit). Column 1 reports the estimates without the index, while the estimates including the index are reported in Column 2. In Columns 3 and 4 we show the results without and with the selection correction term for the model with two mass points. We have calculated the training effects in percentage points and reported them at the end of the table. The asymptotic distribution and the rate of convergence are not known for the discrete mixture model, although some results are available, see for instance Cameron and Heckman (1998). The model did not converge without starting values. As starting values we used the pooled probit estimates.

It is clear from Table 4 that additional mass points improve the model fit significantly. The likelihood function with one mass point is -2728, while the model with two masses has a likelihood of -2053. We are not able to find maximum points for the likelihood function that are robust to different starting values for more than two mass points. However, many studies deem

[15] There is a close connection between the Local Average Treatment Effect (LATE) parameter developed in Imbens and Angrist (1994), and the types of instruments used in the identification of treatment effects; see Vytlacil (2002). Using discrete instruments, LATE is defined as the effect of treatment for those who are triggered by the instrument to go into training (treatment effects for 'compliers'). In the continuous case, LATE is the effect for those who are just indifferent to going into training. This effect has to be simulated for all levels of the selection index, see for instance Aakvik et al. (2002). The effect of treatment on the treated (TT) that is estimated in this paper is the integral of the LATE up to the level where the unobservables make the individual indifferent to entering into training, while ATE is the integral over the full support of the selection index, see for instance Aakvik et al. (2002).

Table 4. Probit regressions of employment with a discrete mixture distribution and correction for selectivity bias

	1a	1b	2a	2b
Work experience	0.139**	0.136**	0.174**	0.172**
	(0.012)	(0.013)	(0.020)	(0.023)
Work experience^2/10	−0.020**	−0.020**	−0.032**	−0.033**
	(0.003)	(0.003)	(0.008)	(0.009)
Spouse's income/100000	0.029**	0.028**	0.033**	0.024**
	(0.005)	(0.004)	(0.010)	(0.009)
Married	0.029	0.038	−0.008	0.199
	(0.055)	(0.056)	(0.110)	(0.111)
Age	−0.125**	−0.121**	−0.067**	−0.050*
	(0.023)	(0.013)	(0.026)	(0.025)
Age squared/10	0.008**	0.008**	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.003)
Male	−0.486**	−0.454**	−0.604**	−0.450**
	(0.045)	(0.049)	(0.085)	(0.096)
Education	0.089**	0.074**	0.092**	0.065*
	(0.013)	(0.014)	(0.027)	(0.031)
Children	−0.221**	−0.210**	−0.247*	−0.146
	(0.048)	(0.050)	(0.100)	(0.103)
Local unemployment	−0.013**	−0.009	−0.026	−0.022
	(0.020)	(0.019)	(0.035)	(0.036)
Index	–	−0.056	–	−0.050
		(0.051)		(0.057)
α_1	1.158**	1.104**	−1.500**	−1.856**
	(0.305)	(0.301)	(0.580*)	(0.588)
α_2	–	–	1.225*	0.886**
			(0.571*)	(0.589)
Mass probability			0.565	0.557
Training: Education (d_i)	0.048	0.061	0.133	0.063
	(0.050)	(0.052)	(0.105)	(0.056)
Training Effect (percent)	1.9	0.2	2.1	0.2
	(0.019)	(0.002)	(0.018)	(0.002)
Number of obs.	4518	4518	4518	4518
Log-Likelihood	−2728	−2726	−2053	−2052

Note: Column 1a is a probit model with one mass point and no selection correction. Column 1b is a one mass point model with selection correction. Table 2a is a probit model with two mass points and no selection correction while Column 2b is a two mass point model with selection correction. Numbers in parentheses are standard errors from bootstrap using the normal approximation with 100 replications.

two mass points sufficient to characterize unobserved heterogeneity; see for instance Cameron and Heckman (1998) and Card and Sullivan (1988).

The effect of selection bias is negative but these effects are estimated with large standard errors, and are relatively small in magnitude. The effect of the index variable is almost of the same magnitude as the training effect, reducing the training effect almost to zero. The effect of training without including the index is also relatively small, and is not statistically significantly different from zero at conventional significance levels. We experimented with different specifications by adding indexes of higher order. However, the estimated coefficients on these terms were very small and not significantly different from zero.

6. Concluding remarks

Our main objective in this paper is to evaluate what effect education as a vocational rehabilitation programme for disabled persons has in terms of increasing employment rates. Descriptive statistics reveal that the employment rate among persons participating in educational programmes was more than eight percentage points higher than employment rate for eligible individuals not participating in active training. However, many factors can render this simple estimator of the training effect invalid. First, we find evidence of differences in observed characteristics between vocational rehabilitation clients participating in educational programmes and VR clients not participating in any programme. Individual characteristics such as youth, higher level of education (before training), and more work experience significantly increase the probability of participating in educational training programmes. So does higher level of unemployment in the local district. Finally, the chances of participating are higher for women than for men.

All the individual characteristics that increase the probability of participating in educational training have positive effects on the probability of being employed after the training period. This fact is consistent with the hypothesis of creaming in the sense that persons participating in training programmes are those most likely to obtain a job after training. The local unemployment rate contradicts the creaming hypothesis in the sense that individuals living in areas with a relatively high unemployment rate are more likely to participate in training programmes, but are less likely to get a job. However, this coefficient is not statistically different from zero in the employment outcome.

Adjustment for observed discrepancies between trainees and nonparticipants reduces the estimated training effects to below three percentage points. We rejected the assumption of normality for the error terms. Thus, standard econometric selection models relying on joint normality of the error terms cannot be used. We specify an alternative model that uses non-parametric and semi-parametric estimation techniques. From these models we find that selection on unobservables accords with the creaming hypothesis, although the magnitude is low.

References

Aakvik A (2000) Assessing the effects of labor market training in Norway. Unpublished manuscript, Department of Economics, University of Bergen

Aakvik A, Heckman JJ, Vytlacil EJ (2002) Training effects on employment when the training effects are heterogeneous: An application to Norwegian vocational rehabilitation programs. Journal of Econometrics (forthcoming)

Aakvik A, Kjerstad E (2002) Skill formation among vocational rehabilitation clients – public policy vs private incentives. Education Economics (forthcoming)

Aakvik A, Risa AE (1998) Success through selection in Norwegian rehabilitation programs? Unpublished manuscript, Department of Economics, University of Bergen

Ahn H, Powell JL (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. Journal of Econometrics 58:3–29

Bell SH, Orr LL, Blomquist JD, Cain GG (1995) Program applicants as a comparison group in evaluating training programs. W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan

Bertschek I, Lechner M (1998) Convenient estimators for the panel probit model. Journal of Econometrics 87:329–371

Butler JS, Moffitt R (1982) A computationally efficient quadrature procedure for the one factor multinomial probit model. Econometrica 50:761–764

Cameron SV, Heckman JJ (1998) Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. Journal of Political Economy 106:262–333

Card D, Sullivan D (1988) Measuring the effect of subsidized training programs on movements in and out of employment. Econometrica 56:497–530

Cosslett SR (1983) Distribution-free maximum likelihood estimator of the binary choice model. Econometrica 51:765–782

Dean DH, Dolan RC (1992) Efficacy of higher education for persons with work disabilities. Economics of Educational Review 11:51–60

Dean DH, Dolan RC (1991) Fixed-effects estimates of earnings impacts for the vocational rehabilitation program. Journal of Human Resources 26:380–391

Goldberger AS (1972) Selection bias in evaluating treatment effects: Some formal illustrations. Discussion paper 123–72. Institute for Research on Poverty, University of Wisconsin

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

Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47:53–161

Heckman JJ (1997) Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. Journal of Human Resources 32:441–462

Heckman JJ, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental data. Econometrica 66:1017–1098

Heckman JJ, Hotz VJ (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. Journal of the American Statistical Association 84:862–74

Heckman JJ, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52:271–320

Heckman JJ, Willis R (1976) Estimation of a stochastic model of reproduction: An econometric approach. In: Terleckyj N (ed) Household production and consumption. New York: National Bureau of Economic Research

Ichimura H, Todd P (1998) Implementing nonparametric and semiparametric estimators. Manuscript prepared for Handbook of Econometrics Volume 5

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 62:467–475

Kiefer J, Wolfowitz J (1956) Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics 27:887–906

Laird NM (1978) Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association 73:805–811

LaLonde RJ (1995) The promise of public sector-sponsored training programs. Journal of Economic Perspectives 9:149–168

Lechner M (1995) Some specification tests for probit models estimated on panel data. Journal of Business and Economic Statistics 13:475–488

Lindsey BG (1983a) The geometry of mixture likelihoods: A general theory. The Annals of Statistics 11:86–94

Lindsey BG (1983b) The geometry of mixture likelihoods: The exponential family. The Annals of Statistics 11:783–792

Manski CF (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. Journal of Econometrics 27:313–334

Nowak L (1983) A cost effectiveness evaluation of the Federal/State vocational rehabilitation program – using a comparison group. The American Economist 27:23–29

Vella F (1998) Estimating models with sample selection bias: A survey. Journal of Human Resources 33:127–169

Vytlacil, EJ (2002) Independence, monotonicity, and latent index models: An equivalence result. Econometrica 70:331–341

Worrall JD (1988) Benefit and cost models. In: Berkowitz M (ed) Measuring the efficiency of public programs. Costs and benefits in vocational rehabilitation. Temple University Press, Philadelphia.