

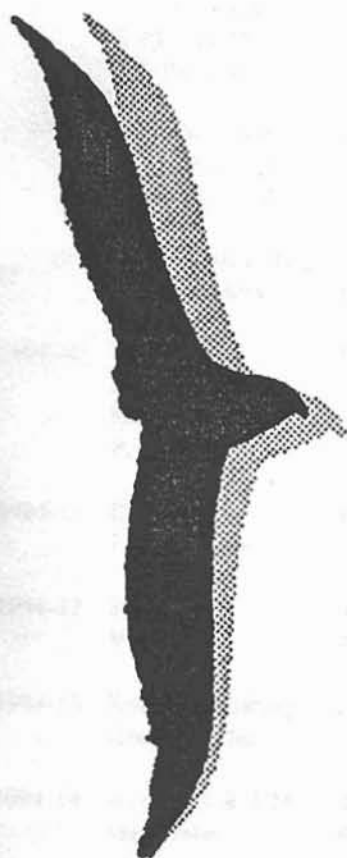
# Serie research memoranda

## Estimating the Effect of Training on Individual Durations of Search

Anders Holm

Research Memorandum 1996-33

July 1996



|a|l|e|r|t|

applied  
labour  
economics  
research  
team



# Estimating the Effect of Training on Individual Durations of Search

Anders Holm\*

Tinbergen Institute

Keizersgracht 482

1017 EG Amsterdam

The Netherlands<sup>†</sup>

Phone: + 31 20 551 35 00

Fax: + 31 20 551 35 55

E-mail: anders@tinbinst.nl

and

Institute for Local Government Studies, Denmark

Nyropsgade 37

1602 Copenhagen V

Denmark

---

\*I would like to thank Professor Richard Davies, Dr. Gerard J. van den Berg, Professor Erling B. Andersen, Dr. Torben Tranæs, Dr. Peter Jensen, Dr. Karsten Albæk and Dr. Lars Muus for valuable comments and hints. I would also like to thank Mikkel Baadsgaard for valuable research assistance and finally the Institute of Local Government Studies, Denmark and the European Commission for providing a research grant (TMR grant no. erb4001gt954632).

<sup>†</sup>Address for correspondence

## Abstract

In this paper a statistical model combining duration of search for an apprentice vacancy and choice of entering a training programme aiming at helping students finding an vacancy is presented. Using conventional duration models yields a negative effect on durations of search from the programme. Allowing for correlated unobserved heterogeneity in a joint statistical model for choice of training and duration of search renders this effect insignificant. However, a somewhat more complicated model yields a less clear picture.  
*keywords:* Training, Duration Data, Correlated Random Effects, Non-parametric mixture, Simulated Annealing, Experimental Data.

## 1 Introduction

This paper studies the effect of a training programme. The aim of the training programme is helping students finding an apprentice vacancy. In Denmark labour market training is in general provided at fixed time points during the duration of unemployment. Hence modelling choices of training by unemployed reduces to modelling discrete choices. The motivation for this analysis is an apparent paradox. Those students who do not accept the offer of training have, on average, shorter search spells for an apprentice vacancy than those who choose training. This effect prevails even when observed heterogeneity is accounted for. This phenomenon of an observed average negative effect of the training programme could be due to a self-selection effect as it is proposed in e.g. Willis and Rosen (1979). However it is plausible that in our case the training programme is chosen by individuals who on average have lower unobserved abilities in finding an apprentice vacancy than those who do not choose training. To investigate whether a potential self-selection effect is due to unobserved heterogeneity I propose a simultaneous statistical model for the choice of training and the impact of training on the duration of search spells for students searching for apprentice vacancies. The proposed model takes



into account the possible non-random allocation into the training programme by potential entrants.

As there are regional differences in the administration of the schemes, these regional differences can be interpreted as an experimental allocation of students into the training program. This establish an exclusion restriction in the model, i.e. a variable that should appear in the choice component and not in the duration component of the model. The exclusion restriction enables, from an identification point of view, the training variable to be treated as an endogeneous variable, as proposed in the joint model mentioned in the previous paragraph.

Even though the model is used in an application of somewhat limited interest, one can easily think of more general applications. With a small modification it is applicable to the general situation of unemployed searching for employment while searching who can be offered training. Another interesting application of the model could be the analysis of the effect of on the job training on employment spells. In general, it is useful in areas where correlated choices and duration are under investigation.

The micro-econometric literature on the labour market effects of training usually confines itself to linear models, as surveyed by Maddala (1983). In these models training influences wages, income or amount of unemployment during a year. If the error term in the linear model on the effect of training is correlated with the error term in an equation for selection into training, one obtains a simultaneous equation system, known in the literature as a 'self-selection' model. The novelty of this paper is thus to propose a duration model that deals with the problems of self-selection. Another recent paper in this area is Gritz (1993). In his model entry into training occurs randomly over time. In our case entry occurs at a fixed unemployment duration.

Section 2 of this paper shows the statistical model for homogeneous observations and the likelihood for data generated by heterogeneous observations is developed in section 3. Section 4 deals with estimation issues and proposes a method to estimate the quite complicated likelihood. Section 5 offers an application on a Danish data set and section 6 contains some concluding remarks.

## 2 A Specific Search Model with Optional Training

In this section I propose a search model for apprentice vacancies with an option for training while searching. The option is valid only at the outset of the search spell, because the model is designed to accommodate the situation where students are allowed admission into training only once. There is no admission into training at any point later in time<sup>1</sup>.

The training programme is supposed to help students find an apprentice vacancy. Thus the value of the training programme for the searching individual is to shorten search spells and perhaps provide the student with a student grant while training. The value of not choosing training is the salary in a potential job held while searching.

Let the random binary variable  $Y$  take the value 1 if the student chooses the training programme and 0 otherwise. Denote by  $V_y$ ,  $y \in \{0, 1\}$ , the value of choosing  $Y = y$ . Now  $Y = 1$  if  $V_1 - \zeta \geq V_0$ , where  $\zeta$  is a continuous random variable. The variable  $\zeta$  can be interpreted as a taste-variable, capturing the students taste related

---

<sup>1</sup>It is, however, easy to generalize the model to a situation where unemployed are searching for employment and during their search are offered training at a fixed point in time.

to the choice of  $y$ . We can write the probability of  $Y = y$  as:

$$P(Y = y) = F(V_1 - V_0)^y (1 - F(V_1 - V_0))^{1-y}, \quad (1)$$

where  $F$  is the cdf. of  $\zeta$ .

Now we want to put forward a joint probability of a choice,  $y$ , and the duration of search,  $t$ . The idea, of course, is to investigate the effects of the choice on the duration of search.

The hazard in a search spell  $[t, t + dt]$ , of choice  $y$  is:

$$\begin{aligned} \theta_y(t)dt &= P(\text{a vacancy is found in } [t, t + dt] \mid t, Y = y) \\ &\quad \times P(\text{the vacancy is accepted}). \end{aligned} \quad (2)$$

Rewriting (2) in terms of the survivor function  $G_y(t) = \exp\{-\int_0^t \theta_y(s)ds\}$  and combining with (1) yields the joint probability that an observed value of a search spell is larger than or equal to  $t$  and  $Y = y$ :

$$\begin{aligned} P(T \geq t, Y = y) &= P(T \geq t \mid Y = y)P(Y = y) \\ &= G_y(t)F(V_1 - V_0)^y (1 - F(V_1 - V_0))^{1-y}. \end{aligned}$$

This model is useful for two reasons. First it tells us something about the expected behavior of searching students. From (1) it follows that the arrival rate of apprentice vacancies on respectively off training does not solely determine the choice of training. choice specific benefit/search cost and the distribution of different types of apprentice vacancies on respectively off training also determine the choice. For example, some students might choose training, not because they expect their search spell to become shorter, but because they think they get access to more attractive apprentice vacancies while on training. Second, the model is useful because it tells us how to construct a likelihood for the model.

### 3 A General Model

I will generalize the model from the previous section to accommodate the situation of heterogeneous individuals. Some of these are observed to chose training before they start searching and some are not. As in the previous section there will be observations on the chooice and of a choice-specific search spell. We observe some regressors for each student, but we also assume that some important explanatory variables are not observed, causing unobserved heterogeneity in the data.

Application of a usual statistical model for durations (possibly accommodating non-observed heterogeneity) tends to yield biased estimates of the individual effects of a particular training programme as such a model assumes random allocation into the training programme. Estimates based on such models yield the average effect over observed individuals. The model that will be introduced below allows the selection into training to be decided by heterogeneous individuals for example determining their choice by comparing discounted lifetime earnings from choosing training and not choosing training. If some unobserved heterogeneity determines the choice of training as well as the duration of search, standard duration models will not yield an unbiased estimate of the individual effect of training, even if they allow for unobserved heterogeneity.

This far individuals have only been assumed heterogeneous in respect to the taste-variable  $\zeta$ . I shall now introduce further individual heterogeneity. The interpretation of the heterogeneity is that search income differs by individuals because each individual might face different labour market conditions, yielding different search incomes. Vacancy offer rates might depend on individual characteristics such as ability, age etc, and thereby become heterogeneous over individuals.

Consider the following reduced model determining the choice of training  $y$  and

the duration of the subsequent search spell  $t$ .

Let  $x_0$  be a vector of observed variables and  $\varepsilon_0$  a random effect representing unobserved heterogeneity; together they affect duration on choice 0. Define similarly for choice 1,  $x_1$  and  $\varepsilon_1$ . Let furthermore  $x_2$  be a vector of observed variables and  $\varepsilon_2$  a random effect representing unobserved heterogeneity that together determine the choice of  $y$ . Finally, let  $x = (x_0, x_1, x_2)$  and  $\varepsilon = (\varepsilon_0, \varepsilon_1, \varepsilon_2)$ .

Write the density function of  $t$  conditional on a choice of training and some observed variables and an unobserved random effect as:

$$g_y(t) = g_y(t \mid \beta_y x_y + \varepsilon_y, \eta_y), \quad (3)$$

where  $\eta_y$  is a vector of any parameters determining the shape of  $\theta_y$ , and where  $\beta_y$ ,  $y = 0, 1$ , are vectors of unknown regression parameters.

Next we write the probability of choosing  $y$ , conditional on some observed variables and an unobserved random effect:

$$P(Y = 1 \mid x_2, \varepsilon_2) = F(V_1 - V_0 \mid x_2, \varepsilon_2) = F(\beta_2 x_2 + \varepsilon_2), \quad (4)$$

where  $\beta_2$  is a vector of unknown parameters. Define  $z = \beta_2 x_2 + \varepsilon_2$  for notational convenience.

The equations (3) and (4) can be interpreted as a Taylor approximation of the underlying functional form of the structural model. This is what is normally being assumed in situations where reduced models for durations of unemployment are used to build likelihoods for spells of unemployment thought of being generated by an underlying theoretical search model, Lancaster (1979), Nickell (1979), Atkinson et al. (1984) and Naradranathan et al. (1985).

We assume independence between the regressors and the unobserved heterogeneity, i.e.,  $x \perp \varepsilon$ . Let  $\varepsilon \sim v(\varepsilon)$  where  $v(\cdot)$  is some unknown density<sup>2</sup>. Note that

---

<sup>2</sup>In the following we suppress any parameters appearing in  $v(\cdot)$  for notational simplicity.

even if  $\varepsilon$  was observed, we would never observe all three elements of  $\varepsilon$  simultaneously for any individual as each individual can only be observed either off training,  $y = 0$ , or on training,  $y = 1$ . Hence we need the marginal distribution of  $\varepsilon_0$  and  $\varepsilon_2$  and the marginal distribution of  $\varepsilon_1$  and  $\varepsilon_2$ . Therefore, we define  $\varepsilon^y = (\varepsilon_y, \varepsilon_2)$ ,  $v^y(\varepsilon^y) = \int v(\varepsilon) d\varepsilon_{1-y}$ ,  $y = 0, 1$ .

We can write the joint distribution of  $t$  and  $y$  conditional on  $x$  and  $\varepsilon$  as the product of the conditional distribution of  $t$  given  $y$ ,  $x_y$  and  $\varepsilon_y$ , (3), the conditional probability of  $y$  given  $x_2$  and  $\varepsilon_2$ , (4). That is:

$$\begin{aligned} & P(T \geq t, y \mid x, \varepsilon) \\ &= P(T \geq t \mid y, x_y, \varepsilon_y) \times P(Y = 1 \mid x_2, \varepsilon_2)^y \times (1 - P(Y = 1 \mid x_2, \varepsilon_2))^{1-y} \\ &= G_y(t) \times F(z)^y \times (1 - F(z))^{1-y}, \end{aligned}$$

where  $G_y(t) = \int_0^t g_y(s) ds$ . This is the joint distribution for a single individual of the observed duration given choice  $t$ , and the choice,  $y$ , given the three random effects  $\varepsilon$ , a set of observed vectors of explanatory variables  $x$  and vectors of unknown coefficients  $\beta = (\beta_0, \beta_1, \beta_2)$ .

Note how an observed value of training,  $y$ , affects observed durations for a single individual. This happens through the difference between the values of  $\beta_0 x_0$  and  $\beta_1 x_1$ ,  $\eta_0$  and  $\eta_1$  and the difference between values of  $\varepsilon_0$  and  $\varepsilon_1$ .

Denote each individual by a subscript ' $i$ ' and multiplying all individual joint distributions of  $t$  and  $y$  yields the sample likelihood conditional on  $\varepsilon_0$ ,  $\varepsilon_1$ , and  $\varepsilon_2$  with sample size  $n$ :

$$L_{cl}(\beta, \eta) = \prod_{i=1}^n g_0(t_i)^{1-y_i} \times g_1(t_i)^{y_i} \times F(z_i)^{y_i} \times (1 - F(z_i))^{1-y_i}, \quad (5)$$

which is denoted  $L_{cl}$ , where  $\eta = (\eta_0, \eta_1)'$ .

As  $\varepsilon$  is unobserved, inference on  $\beta$  and  $\eta$  must be based on the marginal likelihood, marginalized with respect to  $\varepsilon$ . That is, we must integrate (5) over the entire range of the multivariate support of  $\varepsilon$ . Denote this space  $\Omega$ . The marginal likelihood  $L_{ml}$ , is:

$$\begin{aligned} L_{ml}(\beta, \eta) &= \prod_{i=1}^n \int_{\Omega} g_0(t_i)^{1-y_i} \times g_1(t_i)^{y_i} \times F(z_i)^{y_i} \times (1 - F(z_i))^{1-y_i} dV(\varepsilon) \\ &= \prod_{i=1}^n \int_{\Omega^y} g_0(t_i)^{1-y_i} \times g_1(t_i)^{y_i} \times F(z_i)^{y_i} \times (1 - F(z_i))^{1-y_i} dV^y(\varepsilon^y), \end{aligned} \quad (6)$$

where  $\Omega^y$  is the support of  $\varepsilon^y$  and where the last equality follows because  $\varepsilon_{1-y}$  does not affect the (conditional) density of  $t$  given the choice of  $y$ , nor the (conditional) probability of  $y$ . The cdf. of  $\varepsilon$  is  $V$ .

In this set-up the effect from choosing training on the duration for a single individual is determined through different values of  $\eta_0$  and  $\eta_1$ , different values of the regressor parameters, i.e.  $\beta_0 x_0$  compared to  $\beta_1 x_1$ , and different values of the unobserved heterogeneity on each choice, that is  $\varepsilon_1$  and  $\varepsilon_0$ .

Note that even if the training programme does decrease search spells for comparable individuals this does not necessarily indicate that students off the training programme are worse off than those who are on the programme, as the differences in choices are due to their differences in the taste parameter,  $\zeta$ , which reflects differences in students' tastes regarding training.

In (6) the effect of training happens through many different parameters. This is perhaps not necessary and also a complication for the estimation of the model. We shall therefore propose a simplification of the model, where the entire effect of training is confined to a single parameter,  $\gamma$ . Assume that the observed part of the linear predictor in (3) can be written as  $\beta_1 x_1$  for  $y = 0$  and  $\beta_1 x_1 + \gamma$  for  $y = 1$ . Assume also that  $v^0(\varepsilon^0) = v^1(\varepsilon^1)$  and that  $g_0(\cdot \mid \cdot) = g_1(\cdot \mid \cdot) = g(t \mid \beta_1 x_{1,i} + \gamma y_i + \varepsilon_{1,i}, \eta_1)$

implying  $\eta_0 = \eta_1$ . Then (6) can be simplified to:

$$L_{ml}(\tilde{\beta}, \eta_1) = \prod_{i=1}^n \int_{\Omega^1} g(t_i) \times F(z_i)^{y_i} \times (1 - F(z_i))^{1-y_i} dV^1(\varepsilon^1), \quad (7)$$

where  $\tilde{\beta} = (\beta_1, \beta_2, \gamma)$ .

Note that in this model, the only effect on duration from choosing training is through a single parameter  $\gamma$ , as all other parameters remain unchanged for a single individual in the duration component of the likelihood no matter the choice of  $y$ . In this sense it resembles the usual linear selection model, introduced in the literature by Heckman (1979).

This far I have proposed a statistical model accommodating the search approach introduced in section 2. The model predicts the effect of choosing training on the duration of search spells and the models allow for observed as well as unobserved heterogeneity between individuals searching for an apprentice vacancy. The model estimates duration of search switching between two regimes, one with training and one without. I have furthermore presented a simplification of the model. In this model the only difference between the two regimes is a constant shift in the hazard of obtaining an apprentice vacancy.

## 4 A Semi-Parametric Estimation Method

Because of the unobserved heterogeneity, optimizing the likelihoods (6) and (7) involves the evaluation of a bivariate integral for each observation. Even with a parametric assumption about the joint distribution of the error terms, these likelihoods still represent a computational burden, if this assumption does not lead to any analytical solution of the integrals<sup>3</sup>.

---

<sup>3</sup>It is also hard to find bivariate distributions which lead to analytical solutions.

A general semi-parametric method for estimation of mixture models has been suggested by Kiefer and Wolfowitz (1956), and is further developed among others by Lindsay (1983). This method estimates the true error distribution by a non-parametric distribution. Besides lightening the computational burden it also offers a less restrictive approach in the sense that we do not have to specify an (arbitrary) functional form for the distribution of the unobserved heterogeneity.

We shall now follow this semi-parametric approach by proposing a bivariate non-parametric distribution as an approximation to the true error distributions  $v^0(.,.)$ ,  $v^1(.,.)$  and  $v(.,.)$ . This is of great convenience as nothing so far has been suggested about their parametric form.

We label this method 'semi-parametric' because in our case we still have to make parametric assumptions about the duration density and the choice distribution but not any assumptions about the density of the unobserved heterogeneity. We shall only show the likelihood for the restricted version of the model (7). The principle easily follows for the unrestricted model.

We replace the true error terms  $\varepsilon_1$  and  $\varepsilon_2$  by two finite dimensional vectors of mass-points  $u_1$  and  $u_2$  with dimension  $J$  and we replace the true error distribution by a vector of weights,  $p$ , also with dimension  $J$ . The semi-parametric likelihood  $L_{sp}^J$ , for given values of  $J$ , can now be written as:

$$L_{ml}(\tilde{\beta}, \eta_1) \approx L_{sp}^J(\tilde{\beta}, \eta_1, u_1, u_2, p, J) = \prod_{i=1}^n \sum_{j=1}^J g(t_i | \beta_1 x_{1,i} + \gamma y_i + u_{1,j}, \eta_1) \times F(\beta_2 x_{2,i} + u_{2,j})^{y_i} \times (1 - F(\beta_2 x_{2,i} + u_{2,j})^{1-y_i}) p_j, \quad (8)$$

where  $u_1 = (u_{1,1}, \dots, u_{1,J})'$ ,  $u_2 = (u_{2,1}, \dots, u_{2,J})'$ ,  $p = (p_1, \dots, p_J)'$ , and where  $0 \leq p_j \leq 1$ ,  $j = 1, \dots, J$ ,  $\sum_j p_j = 1$ <sup>4</sup>.

As  $J$  increases the approximation in (8) becomes more and more valid, assuming

---

<sup>4</sup>Meeting these constraints is dealt with in appendix 1.

appropriate choices of  $u_1$ ,  $u_2$  and  $p$ . The semi-parametric approach treats  $u_1$ ,  $u_2$ , and  $p$  as unknown coefficients that are to be determined by the data while maximizing  $L_{sp}^J$ .

Note that even if the two vectors of coefficients  $\beta_1$  and  $\beta_2$  contain an intercept term, these are absorbed into the location of the mass-points, see Davies (1993). This is because we cannot identify a separate mean in the error distributions, see Elbers and Ridder (1982).

We also need to know how to find optimal values of  $J$ :  $\hat{J}$ . This can be determined the following way: Let  $l_{sp,i}^{\hat{J}}$  be individual contributions to the log of the likelihood (8) using  $\hat{J}$  as values for  $J$  and define the following function  $D = \sum_{i=1}^n [\frac{l_{sp,i}^1}{l_{sp,i}^{\hat{J}}} - 1]$ , where the upper part of the fraction is individual contributions to  $\log L_{sp}^1$  using only one mass-point in the duration component and one mass-point in the choice component ( $J = 1$ ) and the lower part is individual contributions to  $\log L_{sp}^{\hat{J}}$  using  $\hat{J}$  mass-points for the non-parametric representation of the unknown bivariate mixing distribution  $\tilde{v}(\cdot, \cdot)$ .

By a theorem of Lindsay (1983), values of  $\hat{J}$  yielding  $D = 0$  provide an optimal mixing distribution in the sense that a mixing distribution including a higher number of mass-points will not yield a higher value of the semi-parametric likelihood. In the literature  $D(\cdot)$  is known as the Gateaux derivative, see Heckman and Singer (1984) or Lancaster (1990) for the usage in duration models with unobserved heterogeneity or e.g. Michel and Herget (1993) for a general mathematical framework.

Consistency of a non-parametric representation of an unknown distribution of heterogeneity is shown by Kiefer and Wolfowitz (1956) and consistency in the case of a Weibull duration model with a non-parametric error representation for omitted heterogeneity is shown by Heckman and Singer (1984).

Practical applications of the semi-parametric representation of a true bivariate error distribution are provided by Davies and Pickles (1987) and Davies (1993).

A computational problem when maximizing (8) remains. It turns out that the likelihood surface is non-convex and hence ordinary local optimizing algorithms such as the newton-raphson routine might reach a local rather than the global optimum. These routines reaches the local optimum in the direction of the largest values of the gradients given the starting values. Hence different starting values need not yield the same optimum for non-convex functions. For functions with this feature global optimizing routines are needed to ensure convergence to a global optimum. Such an algorithm could be the Simmulated annealing algorithm.

In this paragraph we explain this algorithm in some detail. A usefull full reference is e.g. Coranra et al. (1987). The Simulated annaeling algorithm routine search for a optimum in a pre-set range of the parameter space of the function:  $\Omega$ . Usually  $\Omega \subseteq R^k$ , where  $k$  is the dimension of the parameter space. The parameter vector as e.g. in (8)  $\lambda = \{\tilde{\beta}, \eta_1, u_1, u_2, p, J\}'$  is updated by the following equation:

$$\lambda^* = \lambda_j + r v_h e_h, h = 1, \dots, k,$$

where  $\lambda_j$  is the current value of the parameter vector,  $r$  is a random number in the range  $[-1, 1]$ ,  $v_h$  is the component of the step vector in the  $h$ 'te direction and  $e_h$  is the vector of the  $h$ 'te coodinate direction.

During optimization parameters are updated through three nested loops:

- the inner loop consists of drawing n number of r's say  $N_s$  times. For each seperate draw the value of the function under maximaztion,  $f(\lambda)$  is calculated using  $\lambda^*$  as argument. If  $f(\lambda^*) > f(\lambda_j)$  the new point is accecpeted. If  $f(\lambda^*) \leq f(\lambda_j)$  the new point is accepted with probablity  $p = \exp(\frac{f(\lambda^*) - f(\lambda_j)}{T})$ . This is called the metropolis step and allowes dimetrial steps. In practice a uniform random number

in the range  $[0, 1]$ ,  $p'$ , is generated and the new point is accepted if  $p' > p$ . The metropolis step in the algorithm is intended to allow movements out of a local optimum. In case of acceptance  $\lambda_j$  is replaced by  $\lambda^*$  and search is continued. In case of rejecting search is continued retaining  $\lambda_j$  as current value of the parameter vector.

- the intermediate loop changes the vector of steplengths  $v$  after  $N_s$  cycles in the inner loop. This can be done using various different techniques, see eg. Corana et al (1987). Usually reduction in steplengths is used to make the algorithm search in a still smaller area of the parameterspace as more step adjustments are made. During the two inner most loops of the algorithm the parameter vector yielding the highest value of the function under optimization is kept recorded.

- the outer loop adjust the temperature parameter  $T$  after say  $N_s \cdot N_t$  cycles of moves along each direction in the parameterspace and after  $N_t$  stepadjustments. After temperature reduction the parameter vector is reset to the value giving the current optimum of the function. By reducing temperature, detrimental steps becomes more unlikely, hence the behaviour of the algorithm becomes more and more like a local optimizing algorithm. The idea is of course that initially a search in the entire parameterspace is needed, but as temperature is lowered the hope is that the algorithm has settled near the global optimum. A proof of convergence to the global maximum is provided by Belisle (1992).

The simulated algorithm allows investigation of several local optima, without locking into the first one encountered. Hence it compromises between a time consuming random search through the entire parameter space and the efficiency of local optimizing algorithms.

In this section I have proposed a semi-parametric estimation method that replaces the bivariate integral in (6) and (7) by a summation over a non-parametric

representation of the unobserved heterogeneity in the likelihood. By this method no assumptions are needed for the density of the unobserved heterogeneity and hopefully, this method will prove computational feasible compared to the case of a parametric assumption about the unobserved heterogeneity. We are now equipped with a model and an estimation method for analyzing the effect of a training programme on search for apprentice vacancies.

## 5 An Application

In this section we shall assess the effect of the Danish training programme for students searching for an apprentice vacancy by using the methods described in the previous sections.

Vocational education in Denmark consists of two parts: first a public provided basic year of formal schooling and secondly an apprenticeship with an employer of two years duration. The students must find an apprentice vacancy by themselves. The apprentice vacancy involves training and payment of a salary. The training period also involves a final examination.

As all students leave schooling at the same date, there tends to be some clustering of searching students each year as a new cohort initiates searching. Adding to this number of 'new' students are of course students from previous cohorts who have not yet obtained an apprentice vacancy for completion of their vocational education.

In 1990 the Danish Ministry of Education decided to solve this problem for students who graduated from school in 1990 (22,254 students) and by January 1991 still had not obtained an apprentice vacancy (7,917 students). Students still searching at this point in time were offered complementary training. This is the

training programme under investigation in this paper.

Offering complementary training had two intentions. Besides offering the students the possibility of completing their education within the complementary training scheme, the scheme is also meant to increase the students' motivation for obtaining an apprentice vacancy. In fact, the students were obliged to accept any given apprentice vacancy if they found one and hence leave the complementary training scheme.

In this section we want to investigate whether the training programme has any effect on the duration on search for an apprentice vacancy. More precisely we want to test the hypothesis impeded in the following two questions:

Q1: 'Does the training programme have an impact on the duration of search spells for comparable individuals?'

Q2: 'Given that the answer to Q1 is affirmative, does the training programme decrease or increase search spells for comparable individuals?'

The data set that is going to be used is a survey data set containing individual information on search time, whether the students entered the training scheme, explanatory variables about the students' ability, age etc and information about the economic and social environment. The different variables are described in detail in appendix 2. The data set consists of 1205 observations

A less formal, but more comprehensive analysis using the data are found in Jensen et al. (1992). The data were sampled by AKF (Institute of Local Government Studies)<sup>5</sup> during September 1991.

Selection of individuals for the data set was done at the time the students were offered complementary training. At this point in time the students had already been

---

<sup>5</sup>In Danish the name is: AKF, Amternes og Kommunernes Forskningsinstitut.

searching for 5 months. Data was collected some 5-6 months later.

Although we know the entire search time for all students in the data set we only want to use search time from the time of the choice of entering the programme and onwards. This is done for simplicity.

We thus estimate the model conditional on the (common) left censoring time of the data. It is possible, however, at some computational and notational cost, to build a likelihood for durations with left censoring, see Lancaster (1979) and Nickell (1979). This is necessary when left censoring time varies among individuals, but in our case we can limit ourselves to an analysis conditional on the common left censoring time.

There is also right-censoring in the data. This is handled in terms of survivor functions and by censoring indicator variables. However, as all students have the same common left-censoring time and sampled approximately at the same date, all right-censored individuals have almost identical search-time and thus only contributes to the likelihood functions through different values of the explanatory values and different values of the choice of training.

First we shall apply the restricted model, (7), which consists of a component that determines a choice of training, a component that determines the duration on and off training and finally a bivariate error density.

The duration component is modeled in terms of a Weibull hazard<sup>6</sup> to allow for potential duration effect:

$$\theta(t_i) = \alpha \mu_i^\alpha t_i^{\alpha-1}, \quad (9)$$

where  $\mu_i = \exp(\beta_1 x_{1,i} + \gamma y_i + \varepsilon_{i,1})$ , and where  $\alpha$  is the Weibull parameter. Duration of a search spell given the choice thus depends on some coefficients  $\beta_1$ , which are not

---

<sup>6</sup>Alternative parametric assumptions about the distributions of spells on and off training where made, but the weibull hazards seems to gives the best description of the duration data.

changing over the choice of  $y$ , some observed regressors  $x_1$ , a parameter  $\gamma$ , which is the only parameter that differs by choice of training and a random effect,  $\varepsilon_1$ .

We model the choice component as a logistic choice model:

$$P(y_i = 1) = \frac{1}{1 + \exp(\beta_2 x_{2,i} + \varepsilon_{i,2})}, P(y_i = 0) = 1 - P(y_i = 1), \quad (10)$$

where  $y = \text{TRAIN}$  (training) and where the individual choice of training is dependent on some constants  $\beta_2$ , some observed regressors  $x_i$ , and some unobserved variables captured by the random effect  $\varepsilon_{i,2}$ . Note that  $y = 1$  only depends on entering the scheme; the effect of dropping-out from the scheme is not estimated in the model. This is avoided for reasons of simplicity.

The probability of choosing training, (10), and the density of spells, (9), together with an (unknown) distribution of  $\varepsilon_1$  and  $\varepsilon_2$  defines a 'selection' model as defined in (7). To estimate this model we will use the semi-parametric method of section 4.

Calculation of estimates of the parameters in the model and their standard errors were carried out by the simulated annealing algorithm and the final optimization is carried out by a local optimization algorithm, the DFPF algorithm, using numerical first derivatives and calculating the Hessian by the outer product of the numerical gradients, see Fletcher (1980). It should be noted that standard errors of the estimated parameters are only valid under the assumption that the distributions of the unobserved heterogeneity are finite discrete mixture distributions, see Titterton et al. (1985), with number of mass-points as indicated in the following tables containing the estimation results.

The results from estimating the restricted model are reported in table 1. The reported estimates are based on a non-parametric representation of the error using only 2 mass-points. This model is labeled as 'non-par' in table 1.

In table 1 we also show estimates obtained by excluding the error term in the

choice component, (10), and the duration component, (9). This model labeled 'no error' thus corresponds to the standard logit choice and Weibull duration models.

## 5.1 Results Using the Restricted Model

We start by commenting on the estimates for the choice component. Unless otherwise stated the comments concern the estimates for the model allowing for unobserved heterogeneity. Overall there is little difference between the two set of estimates reported in table 1, but some changes in the significance probabilities (the rise in significance probabilities for the estimated parameters reflects that more parameters have been estimated).

TABLE 1 about here

The variable DROP (previously dropped out from an apprenticeship) is positive indicating that students who previously have dropped out from an apprentice vacancy are more likely than others to enter the scheme.

The variable AGE is negative, which is sensible because younger students must be assumed to be more motivated for training than older students. Thus, AGE, might capture some effect from unobserved variables measuring motivation. This could indicate that the independence assumption between the error term and the regressors is invalid.

The significance of the REG3 dummy (regional indicator) indicates regional differences between this region and other regions. These differences are very plausible because the schemes are very different administered between regions, which yields regional differences in how attractive the training scheme is for the students to

enter. This allows an experimental interpretation of allocation into the training programme and also introduces an exclusion restriction in the model. Thus REG3 affects allocation into training, but not the outcome of the search process. By administering the training scheme differently across regions, in virtue, sets up the scheme as an experiment.

Both the included educational dummies (EDUC1 and EDUC3) are negative, indicating that student with 10 years of schooling (EDUC2=the excluded category) are more likely to enter the scheme than students with less than 10 years of schooling and students with academic qualifications. This is also plausible, because the former are usually considered less clear about their labour market career (which is why they have chosen an additional year of secondary schooling) and therefore are more likely to want guidance from the training scheme. The variable GRADE (average grade from final examination in secondary school) is negative indicating that students with the lower grades are more attracted to the scheme than others. DGRADE (dummy for unobserved value of GRADE) is also negative, indicating that students with unobserved value of GRADE enter the scheme more often than other students.

We now proceed to the duration equation, which contains the answers to Q1, whether the training programme has an effect, and Q2, whether this effect is positive or negative, obtained from the restricted model. We note that TRAIN (training indicator variable) is significant at a 10% level for the models where unobserved heterogeneity is not taken into account, suggesting that the answer to Q1, whether there is an effect from the training programme, is positive. As TRAIN is negative there is evidence that the answer to Q2, whether training has a decreasing or increasing effect on the duration of search spells, is that training decreases the hazard, and hence increases the duration of search spells.

But moving from the model with no error to the model with the bivariate non-parametric representation of the error terms there is a rise in the coefficient for training (TRAIN) towards zero. In fact, TRAIN becomes insignificant at a 10% level in this model. Thus controlling for self-selection bias in terms of correlated unobserved heterogeneity the negative effect from the training programme diminishes. This suggests that the answer to Q1 is that training does not affect search spells, when unobserved heterogeneity is taken into consideration.

Table 1 does therefore indicate that there is a negative correlation between the unobserved propensity to enter the programme and the unobserved ability to obtain an apprentice vacancy, such that the most able students are outside the training programme, and thereby mistakenly yielding a negative effect of training when unobserved heterogeneity is not taken into account. That training is more attractive for less able students has already been suggested by the negative sign of GRADE in the choice equation.

The effect of the GRADE variable on durations of search also changes when allowing for unobserved heterogeneity. It drops in size and significance, indicating that when unobserved heterogeneity is allowed for, GRADE loses impact on the duration of search spells.

It seems that the Weibull parameter  $\alpha$  increases, which is to be expected, as unmeasured heterogeneity tends to under-estimate the (potential) duration effects.

A final comment on table 1 is that there seems to be a general downward change in all coefficients in the duration model moving from the model with no heterogeneity to the model allowing for heterogeneity. This is not in line with the results in Lancaster and Nickell (1980), where it is stated that the regressor parameters are downward biased in a duration model that does not take into account unobserved heterogeneity. However, estimations on simulated data showed the same pattern as

the results reported in table 1. It remains to further research to investigate this phenomenon.

## 5.2 Application of the Unrestricted Model

Now we will try to relax the assumption of uniform regressor coefficients in the duration on and off the programme. We will thus estimate the unrestricted model (6), where the regressor coefficients and error terms in the duration component of the model vary by choice of training.

Duration of a search spell given the choice of training thus depends on some choice specific coefficients  $\beta_y$ , which are changing over choice of  $y$ , some observed regressors  $x_y$  which in this application are uniform by choice (i.e.  $x_0 = x_1$ ), and some choice specific unobserved regressors captured by the error term  $\varepsilon_{i,y}$ . That is, according to (6) we specify choice specific hazards as:

$$\theta_0(t_i) = \alpha_0 \mu_{i,0}^{\alpha_0} t_i^{\alpha_0-1}, \quad (11)$$

$$\theta_1(t_i) = \alpha_1 \mu_{i,1}^{\alpha_1} t_i^{\alpha_1-1}, \quad (12)$$

where  $\mu_{i,y} = \exp(\beta_y x_i + \varepsilon_{i,y})$ ,  $y = 0, 1$  and where  $\alpha_y$  is a choice specific Weibull parameter.

The effects of the training scheme are now determined by differences in the choice specific regressors  $\beta_y$  (including a possible constant term) and choice specific heterogeneity  $\varepsilon_y$ . We retain the choice component (12) of the model as it is identical for both models (6) and (7). Again we apply the semi-parametric estimation method of section 4.

The full model for a single individual now consists of a choice specific duration component, (11) or (12), a choice component, (10), and bivariate error densities of

$\varepsilon^0$  and  $\varepsilon^1$ , depending on the individual choice of mode.

### 5.3 Results Using the Unrestricted Model

Table 2 reports parameter estimates for the choice and hazard model for individuals on the scheme while table 3 reports similar estimates for individuals off the scheme. Both tables reports estimates of the components of the (two) non-parametric error distributions as well as the values of the log-likelihood for both models. As can be seen from the table both bivariate error-distributions are estimated using only 2 mass-points.

As for table 1, we report results both for the models with (two) bivariate non-parametric representations of the error terms (one for each choice of  $y$ ), which again is labeled 'non-par.' and a model where neither of the duration components nor the choice component allow for an error term. These results are once again labeled 'no-error'.

TABLE 2 about here

TABLE 3 about here

Overall we obtain similar results using the unrestricted model compared to the results obtained by using the restricted model. The constant term is larger for individuals off the scheme, that is, the unrestricted model estimates a negative partial effect of the programme just as we saw the restricted model in table 1 did. However, this effect diminishes somewhat allowing for correlated heterogeneity. The coefficient for GRADE (average grade from final examination in secondary school)

is smaller for persons off the programme and thus indicating that GRADE is of less importance for persons off the programme. This effect is enlarged allowing for correlated heterogeneity as the coefficients for GRADE diminishes in unrestricted model for persons off the programme while it only changes a little in the unrestricted model for persons on the programme.

Adversively the coefficient for AGE is smaller for persons on the training programme than for persons off the programme (and negative for both choices). Hence age does not have the same negative impact on the hazard on than off the programme. One effect of the programme thus seems to be offsetting a possible negative screening effect on older students.

An interesting difference between the estimation results in table 1 and table 2 is that in the unrestricted model the Weibull parameters are higher for persons outside the programme, than for persons on the scheme. This might be due to a screening effect such that individuals who have been on the training programme for a long period has lower hazard than comparable students off the programme. This effect is enlarged in the model allowing for unobserved heterogeneity.

Note finally that the parameters in the choice component in the unrestricted model for persons on the programme are rescaled due to a very large estimated variance for the unobserved heterogeneity. This estimated variance is large because the two mass points are located widely apart and both associated with a relatively large weight (60

Overall the unrestricted model does indicate a negative effect of the training programme even when unobserved heterogeneity is taken into account. This is somewhat in disagreement with the results obtained from the restricted model reported in table 1. But the unrestricted model yields a somewhat insignificant fit on many of the parameters in model and it is thus questionable whether the data

supports estimation of the unrestricted model at all.

In this section we have applied the restricted and the unrestricted model of section 3 to a data set of students searching for an apprentice vacancy and while doing that they are offered training.

The main findings of this application is that whereas training has a significant negative effect using standard duration models, but allowing for bivariate unobserved heterogeneity in estimating choice of training and duration of search we obtain only a insignificant negative effect from training. This result is supported by the fact that the students who enter the programme have lower grades from secondary school than those who choose not to enter the programme.

## 6 Discussion

In this paper I have proposed a search model for agents deciding whether to undertake training while searching for an apprentice vacancy. I have furthermore proposed a likelihood function which accommodates correlated unobserved heterogeneity in both the propensity to choose training and in the probability of finding an apprentice vacancy. A semi-parametric approach was suggested to estimate the effect and distribution of unmeasured regressors.

The proposed models used on a data-set of students searching for an apprenticeship. The models were used because the students were offered training during their search. This training should help them find an apprentice vacancy faster. An immediate negative effect of the programme turned out to be insignificant when unobserved heterogeneity was taken into account.

Some computational problems in the proposed approach remain, however, as the

likelihood allowing for correlated heterogeneity is very flat, with many local maxima. There are furthermore a large number of nuisance parameters compared to the parameters of interest. One type of solution to these problems could be a decomposition of the model where one first estimates the probability of entering training, including the distribution of heterogeneity in this choice by the semi-parametric method. Next the estimated mass-points for this distribution is substituted into the model for durations, where they are used for estimating the distribution of heterogeneity in the duration model conditional on the distribution of heterogeneity in the choice model. This will not reduce the number of nuisance parameters, but perhaps break the optimization problem into two more tractable problems. Calculation of standard errors will also become difficult with this approach, unless some ancillary property can be shown.

## Appendix 1. Constraints on the Weights in the Semiparametric Estimating Approach

In section 4 we must ensure that  $0 \leq p_j \leq 1$ ,  $j = 1, \dots, J$ ,  $\sum_j p_j = 1$ . Therefore we invoke a transformation of some parameters that are to be substituted for the weights in the semiparametric likelihood (8). Denote the transformed parameters  $\psi_j$ ,  $j = 1, \dots, J-1$ , with  $\psi_J$  fixed at some arbitrary value,  $c$  say. Then we write the  $j$ 'th weight as:

$$p_j = f_j(\psi_1, \dots, \psi_{J-1} \mid \psi_J = c) = \frac{\exp(\psi_j)}{\sum_{j=1}^J \exp(\psi_j)}, j = 1, \dots, J.$$

Now  $f_j : R^{J-1} \rightarrow [0, 1]$ ,  $j = 1, \dots, J$  and  $\sum_j f_j = 1$ , such that the parameters that are going to be estimated, the  $(J-1)$   $\psi$ 's, can vary over the entire real line and still ensure that the calculated weights, the  $p_j$ 's, meet the constraints above.

## Appendix 2. Variables in the Training Data Set

|      |        |              |                                                                                                                                                                                                                                                      |
|------|--------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (1)  | ETH    | Ethnicity    | 1 if the respondent has a non-Danish name<br>0 otherwise.                                                                                                                                                                                            |
| (2)  | DROP   |              | 1 if the respondent previously has dropped<br>out from an apprenticeship, 0 otherwise.                                                                                                                                                               |
| (3)  | AGE    |              | The respondent's age measured in years.                                                                                                                                                                                                              |
| (4)  | REG    | Region       | 1 if the respondent lives in the greater<br>Copenhagen area<br>2 if the respondent lives in western Zealand<br>3 if the respondent lives on Funen<br>4 if the respondent lives in northern Jutland<br>5 if the respondent lives in southern Jutland. |
| (5)  | VOC    | Vocation     | 1 if the vocational schooling is in manufacturing<br>0 otherwise (trade and service).                                                                                                                                                                |
| (6)  | EDUC   | Education    | 1 if highest education is 7-9 years of primary<br>and secondary schooling<br>2 if highest education is 10 years of primary<br>and secondary schooling<br>3 if highest education involves some academic<br>qualifications.                            |
| (7)  | UPCT   | Unemployment | Unemployment ratio in the local area.                                                                                                                                                                                                                |
| (8)  | SEX    |              | 1 if the respondent is female, 0 otherwise.                                                                                                                                                                                                          |
| (9)  | GRADE  |              | Final average grade from examination in<br>secondary school.                                                                                                                                                                                         |
| (10) | DGRADE |              | Dummy variable taking the value 0 if GRADE is                                                                                                                                                                                                        |

- observed and 1 otherwise.
- (11) TRAIN Training 1 if the respondent enters the complementary training scheme.
- (12)  $T$  Search time in days for an apprentice vacancy, measured from the time the programme is offered.
- (13)  $d$  1 if the observed spell is complete, 0 otherwise.

### Appendix 3. Summary statistics

The table below gives mean and standard deviation of the variables used in the estimated models.

Mean and standard deviation  
for data used in estimations.

| Variable     | Mean   | Std. dev. |
|--------------|--------|-----------|
| DROP         | 0.12   | 0.33      |
| AGE          | 19.54  | 2.96      |
| REG3         | 0.10   | 0.29      |
| EDUC1        | 0.26   | 0.44      |
| EDUC2        | 0.09   | 0.29      |
| UPCT         | 11.10  | 2.62      |
| GRADE        | 7.85   | 1.02      |
| DGRADE       | 0.38   | 0.49      |
| TRAIN        | 0.58   | 0.49      |
| T            | 186.12 | 75.52     |
| d            | 0.33   | 0.47      |
| Observations | 1205   |           |

## References

- [1] Atkinson, A. B., J. Gomulka and J. Micklewright (1984). Unemployment Benefit, Duration and Incentives in Britain - How Robust is Evidence? *Journal of Public Economics*, 23, 3-26.
- [2] Belisle, C. J. P. (1992). Convergence Theorems for a Class of Simulated Annealing Algorithms on  $R^d$ . *Journal of Applied Probability*, 29, 885-895.
- [3] Corana, A., M. Marchesi, C. Martini and S. Ridella (1987). Minimizing Multimodal Functions of Continuous Variables with the "Simulated Annealing" Algorithm, *ACM Transactions on Mathematical Software*, 13, 262-280.
- [4] Davies, R. B. (1993). Non-parametric Control for Residual Heterogeneity in Modeling Recurrent Behavior, *Computational Statistics and Data Analysis*, 16, 143-160.
- [5] Davies, R. B. and A. R. Pickles (1987). A Joint Trip Timing, Store Choice Model Including Feedback and Inventory Effects and Non-parametric Control for Omitted Variables, *Transportation research*, 21A, 345-361.
- [6] Elbers, C. and G. Ridder (1982), True and Spurious Duration Dependence: The Identifiability of Proportional Hazard Model, *Review of Economic Studies*, XLIX, 403-409.
- [7] Fletcher, R. (1980). *Practical Methods of Optimization*. John Wiley and Sons, New York.
- [8] Gritz, R. M. (1993). The Impact of Training on the Frequency and Duration of Unemployment. *Journal of Econometrics*, 57, 21-51.

- [9] Heckman, J. (1979). Self-Selection as a Specification Error. *Econometrica*, 47, 153-162.
- [10] Heckman, J. and B. Singer (1984): A Method for Minimizing the Distributional Assumptions in Econometric Models for Duration Data, *Econometrica*, 52, 271-320.
- [11] Jensen, T. P., P. Gammeltoft, and A. H. Larsen. (1992). *Praktikpladser til alle?* AKF Forlaget. In Danish, with an English summary.
- [12] Kiefer, J. and J. Wolfowitz (1956). Consistency of the Maximum Likelihood Estimator in the presence of Infinitely Many Incidental Parameters. *Annals of Mathematical Statistics*, 27, 887-906.
- [13] Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment, *Econometrica*, 47, 939-956.
- [14] Lancaster, T. (1990). *The Econometric Analysis of Transitions Data*. Cambridge University Press.
- [15] Lancaster, T. and S. Nickell (1980). The Analysis of Re-employment Probabilities for the Unemployed. *Journal of the Royal Statistical Society*, series A, 143, 141-165.
- [16] Lindsay, B. G., (1983). The Geometry of Mixture Likelihoods: A general Theory. *The Annals of Statistics*, 11, 86-94.
- [17] Maddala, G. S. (1983). *Limited Dependent Variables in Econometrics*. Cambridge University Press.
- [18] Michel, A. N. and C. J. Herget (1993) *Applied Algebra and Functional Analysis*, Dover.

- [19] Naradranathan, W., S. Nickell and J. Stern (1985). Unemployment Benefit Revisited. *The Economic Journal*, 95, 307-329.
- [20] Nickell, S. J. (1979). The Effect of Unemployment and Related Benefits on the Duration of Unemployment. *The Economic Journal*, 89, 34-49.
- [21] Titterington, D. M., A. F. M. Smith and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley and Sons.
- [22] Willis, R. and S. Rosen (1979). Education and Self-selection, *Journal of Political Economy*, 87, 7-36.

**Table 1. Estimation results for the restricted model**  
Choice component

| Variables | no error | sd.error | non-par. | sd.error |
|-----------|----------|----------|----------|----------|
| constant  | 2.322    | 1.201    | -        | -        |
| DROP      | 0.671    | 0.293    | 0.659    | 0.294    |
| AGE       | -0.045   | 0.033    | -0.046   | 0.034    |
| REG3      | 0.262    | 0.111    | 0.274    | 0.123    |
| EDUC1     | -0.356   | 0.210    | -0.361   | 0.208    |
| EDUC3     | -0.566   | 0.322    | -0.575   | 0.321    |
| GRADE     | -0.143   | 0.125    | -0.140   | 0.126    |
| DGRADE    | -0.981   | 0.994    | -0.972   | 1.005    |

*Duration component*

| Variables   | no error | sd.error | non-par. | sd.error |
|-------------|----------|----------|----------|----------|
| constant    | -3.987   | 1.052    | -        | -        |
| TRAIN       | -0.217   | 0.128    | -0.123   | 0.092    |
| DROP        | -0.035   | 0.365    | -0.023   | 0.156    |
| AGE         | -0.047   | 0.031    | -0.028   | 0.021    |
| UPCT        | -0.035   | 0.027    | -0.022   | 0.018    |
| GRADE       | 0.186    | 0.099    | 0.122    | 0.068    |
| DGRADE      | 1.438    | 0.795    | 0.972    | 0.545    |
| $\alpha^1)$ | 1.118    | 0.080    | 1.670    | 0.126    |

*Error component*

| Variables | no error | non-par. |
|-----------|----------|----------|
| $u_{1,1}$ | -3.987   | -3.819   |
| $u_{1,2}$ | -        | 1.026    |
| $u_{2,1}$ | 2.322    | 2.345    |
| $u_{2,2}$ | -        | 1.718    |
| $p_1$     | 1.000    | 0.974    |
| $p_2$     | -        | 0.026    |
| log L     | -1209.1  | -1192.3  |

Table 2. Estimation results for the unrestricted model.  
Only for individual on the scheme

| <i>Choice component</i>   |          |          |          |          |
|---------------------------|----------|----------|----------|----------|
| Variables                 | no error | sd.error | non-par. | sd.error |
| constant                  | 2.325    | 1.086    | -        | -        |
| DROP                      | 0.671    | 0.293    | 2.100    | 1.340    |
| AGE                       | -0.045   | 0.032    | -0.173   | 0.099    |
| REG3                      | 0.261    | 0.112    | 6.181    | 4.880    |
| EDUC1                     | -0.356   | 0.211    | -2.961   | 2.740    |
| EDUC3                     | -0.566   | 0.321    | -3.092   | 2.889    |
| GRADE                     | -0.143   | 0.117    | -0.252   | 0.362    |
| DGRADE                    | -0.981   | 0.934    | -1.498   | 2.817    |
| <i>Duration component</i> |          |          |          |          |
| Variables                 | no error | sd.error | non-par. | sd.error |
| constant                  | -4.742   | 1.441    | -        | -        |
| DROP                      | -0.231   | 0.303    | -0.297   | 0.313    |
| AGE                       | -0.022   | 0.043    | -0.019   | 0.049    |
| UPCT                      | -0.031   | 0.038    | -0.041   | 0.041    |
| GRADE                     | 0.189    | 0.133    | 0.171    | 0.136    |
| DGRADE                    | 1.476    | 1.066    | 1.333    | 1.096    |
| $\alpha^1)$               | 1.098    | 0.109    |          |          |
| <i>Error component</i>    |          |          |          |          |
| Variables                 | no error |          | non-par. |          |
| $u_{1,1}$                 | -4.742   |          | -4.579   |          |
| $u_{1,2}$                 | -        |          | -3.075   |          |
| $u_{2,1}$                 | 2.325    |          | 9.663    |          |
| $u_{2,2}$                 | -        |          | -1.685   |          |
| $p_1$                     | 1.000    |          | 0.608    |          |
| $p_2$                     | -        |          | 0.392    |          |
| log L                     | -819.9   |          | -818.8   |          |

Table 3. Estimation results for the unrestricted model.  
Only for individual off the scheme

| <i>Choice component</i>   |          |          |          |          |
|---------------------------|----------|----------|----------|----------|
| Variables                 | no error | sd.error | non-par. | sd.error |
| constant                  | 2.325    | 1.086    | -        | -        |
| DROP                      | 0.671    | 0.293    | 0.668    | 0.294    |
| AGE                       | -0.045   | 0.032    | -0.047   | 0.034    |
| REG3                      | 0.261    | 0.112    | 0.280    | 0.124    |
| EDUC1                     | -0.356   | 0.211    | -0.354   | 0.213    |
| EDUC3                     | -0.566   | 0.321    | -0.574   | 0.323    |
| GRADE                     | -0.143   | 0.117    | -0.141   | 0.125    |
| DGRADE                    | -0.981   | 0.934    | -0.971   | 0.992    |
| <i>Duration component</i> |          |          |          |          |
| Variables                 | no error | sd.error | non-par. | sd.error |
| constant                  | -3.452   | 1.561    | -        | -        |
| DROP                      | 0.237    | 0.319    | 0.179    | 0.183    |
| AGE                       | -0.066   | 0.045    | -0.033   | 0.025    |
| UPCT                      | -0.043   | 0.036    | -0.022   | 0.021    |
| GRADE                     | 0.177    | 0.147    | 0.106    | 0.089    |
| DGRADE                    | 1.369    | 1.189    | 0.825    | 0.722    |
| $\alpha^1)$               | 1.166    | 0.123    | 2.02     | 0.214    |
| <i>Error component</i>    |          |          |          |          |
| Variables                 | no error | non-par. |          |          |
| $u_{1,1}$                 | -3.452   | -3.498   |          |          |
| $u_{1,2}$                 | -        | 1.191    |          |          |
| $u_{2,1}$                 | 2.322    | 2.371    |          |          |
| $u_{2,2}$                 | -        | 1.163    |          |          |
| $p_1$                     | 1.000    | 0.978    |          |          |
| $p_2$                     | -        | 0.022    |          |          |
| log L                     | 736.2    | 720.8    |          |          |