

Brännäs, Kurt

**Working Paper**

## Estimation in a Duration Model for Evaluating Educational Programs

IZA Discussion Papers, No. 103

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Brännäs, Kurt (2000) : Estimation in a Duration Model for Evaluating Educational Programs, IZA Discussion Papers, No. 103, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/20938>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 103

## **Estimation in a Duration Model for Evaluating Educational Programs**

Kurt Brännäs

January 2000

# Estimation in a Duration Model for Evaluating Educational Programs

**Kurt Brännäs**

*Umeå University, Sweden and IZA, Bonn*

Discussion Paper No. 103  
January 2000

IZA

P.O. Box 7240  
D-53072 Bonn  
Germany

Tel.: +49-228-3894-0  
Fax: +49-228-3894-210  
Email: [iza@iza.org](mailto:iza@iza.org)

This Discussion Paper is issued within the framework of IZA's research area *Project Evaluation*. Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent, nonprofit limited liability company (Gesellschaft mit beschränkter Haftung) supported by the Deutsche Post AG. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public. The current research program deals with (1) mobility and flexibility of labor markets, (2) internationalization of labor markets and European integration, (3) the welfare state and labor markets, (4) labor markets in transition, (5) the future of work, (6) project evaluation and (7) general labor economics.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

## ABSTRACT

### **Estimation in a Duration Model for Evaluating Educational Programs<sup>\*</sup>**

The paper proposes two estimation approaches for duration models that are subject to right censored observations and selection effects. Main focus is on accelerated duration models and the estimators that are of the limited information type, i.e. they are not based on a fully specified selection model. The estimators are instrumental variable adaptations to the Powell symmetrically trimmed least squares and the Buckley-James estimators for right censored data. In a small Monte Carlo study they are found to perform well. Empirical results are presented for total time outside the unemployment state for participants in the Swedish Adult Education Initiative and labour market training. The latter ones are included as controls and are found to have significantly longer times outside the unemployment state. There are strong indications of selection effects.

JEL Classification: C1, C2, I2, J6

Keywords: Buckley-James, Powell, censoring, selection, instrumental variable, adult education initiative

Kurt Brännäs  
Department of Economics  
Umeå University  
SE-901 87 Umeå  
Sweden  
Email: kurt.brannas@econ.umu.se

---

<sup>\*</sup> The author wishes to thank IZA and its members for both financial support and a stimulating work environment during his sabbatical visit. My colleagues Roger Axelsson and Olle Westerlund are thanked for getting me started on the problem, providing the dataset, and some very useful remarks. A previous version of this paper was presented at the IZA-CEPR Workshop "Evaluation of Labor Market Programs" in Berlin, November 1999. P. Jensen and R. Hujer are thanked for helpful remarks.

## ABSTRACT

### **Estimation in a Duration Model for Evaluating Educational Programs \***

The paper proposes two estimation approaches for duration models that are subject to right censored observations and selection effects. Main focus is on accelerated duration models and the estimators that are of the limited information type, i.e. they are not based on a fully specified selection model. The estimators are instrumental variable adaptations to the Powell symmetrically trimmed least squares and the Buckley-James estimators for right censored data. In a small Monte Carlo study they are found to perform well. Empirical results are presented for total time outside the unemployment state for participants in the Swedish Adult Education Initiative and labour market training. The latter ones are included as controls and are found to have significantly longer times outside the unemployment state. There are strong indications of selection effects.

JEL Classification: C1, C2, I2, J6

Keywords: Buckley-James, Powell, censoring, selection, instrumental variable, adult education initiative

Kurt Brännäs  
Department of Economics  
Umeå University  
SE-901 87 Umeå  
Sweden  
Email: kurt.brannas@econ.umu.se

---

\* The author wishes to thank IZA and its members for both financial support and a stimulating work environment during his sabbatical visit. My colleagues Roger Axelsson and Olle Westerlund are thanked for getting me started on the problem, providing the dataset, and some very useful remarks. A previous version of this paper was presented at the IZA-CEPR Workshop "Evaluation of Labor Market Programs" in Berlin, November 1999. P. Jensen and R. Hujer are thanked for helpful remarks.

## 1. Introduction

The present paper is focused on estimation techniques for evaluating an educational program in terms of a subsequent duration of, say, unemployment or employment. Our empirical work is focused on total time outside the unemployment state after participating in the Swedish Adult Education Initiative (AEI, Kunskaftslyftet). Exits to further studies as well as to employment are targeted by this program and the corresponding durations are therefore merged together. The AEI is a five year adult education program that started in July 1997 and targets primarily adult unemployed who lack a three year secondary school education. Employees with a short education can also be included. The AEI contains previous educational arrangements for adults and gives a financial framework for municipalities. Previous labor market training (LMT, Arbetsmarknadsutbildning) programs run by the Swedish Labor Market Board were at the same time planned to be reduced by volume. This did not materialize immediately, however. Axelsson and Westerlund (1999) give an extensive background description on programs and participants.

When future unemployment spells are accounted for in a decision, based on the net present value to enroll in a training program or not, there is an obvious dependence between unemployment spells and the decision. The mechanism is analogous to what we expect to find for post-program wage evaluations (see Appendix A). This is one route for motivating the selection problem for the duration model, and there can be others (e.g., Ham and LaLonde, 1996, Hujer et al., 1999, and Melkersson, 1999).

Note that evaluating programs using durations rather than income measures can usually be made much more easily as well as more promptly. Durations can frequently be obtained from administrative registers, while income measures may be obtained by questionnaires or by tax records with a delay of up to several years. Moreover, short post-program unemployment spells are likely to be of more policy interest than income enhancement effects.

Duration models are usually cast in either a hazard function formulation (the proportional hazard function appears to be the most widely used one) or in terms of a loglinear duration model. Our main emphasis is on the latter. The appeal of the loglinear model is that the model setup is closer to the Tobit model for which there is considerable prior research on the effects of selection or of censoring, and also that it a priori may be possible to get by without making excessive and explicit distributional assumptions. One such assumption is that of a proportional hazard. We will demonstrate a way to nonparametrically obtaining estimated hazard functions for the two groups (cf. Brännäs, 1992a).

In addition to the selection problem, censoring is a feature of most duration data sets that also needs to be accounted for. The paper discusses some adaptations to previously suggested estimation procedures to account for both selection and censoring rather than for merely censoring or merely selection. Two of the estimation procedures (adaptations to the Powell (1986) and Buckley and James (1979) estimators) are partly evaluated in a Monte Carlo study before we proceed to give some

empirical results related to the Swedish Adult Education Initiative.

We start in Section 2 by setting up the model framework and by discussing which measure is to form the basis for the evaluation. Section 3 discusses estimation in accordance with the two measures singled out in Section 2. Section 4 provides the Monte Carlo evidence on small sample properties of estimators and tests. The empirical results are given in Section 5. A discussion of some key issues raised in this paper is saved for the final section. Among such issues we demonstrate that the present modelling strategy remains a valid approach also for a multiple-spell model exercise.

## 2. Model Framework

We consider a loglinear duration model and a conventional selection model:<sup>1</sup>

$$y = \ln t = \mathbf{x}\boldsymbol{\beta} + \alpha d + \sigma\omega \quad (1)$$

$$d = \begin{cases} 1, & \delta = \mathbf{z}\boldsymbol{\gamma} + \xi > 0 \\ 0, & \delta \leq 0. \end{cases} \quad (2)$$

The exogenous variable vector  $\mathbf{x}$  is assumed to contain a constant term so that the intercept  $\beta_1$  contains any nonzero mean of  $\omega$ , which also is assumed to have unit variance. With this formulation we rule out time dependent explanatory variables, but are not restricted to only proportional hazard function models.<sup>2</sup> If the educational program is to be beneficial we expect  $\alpha$  to be negative, i.e. to correspond to a shorter unemployment duration. If instead  $t$  is the duration of employment, we expect  $\alpha$  to be positive.

With regard to the selection mechanism formulated in (2) it is of the conventional form, but could be generally defined. The simple analysis of Appendix A suggests that (2) is not the most clever specification. Estimation techniques that avoid fully accounting for the form of (2) by avoiding to estimate (1-2) jointly, should therefore be of great interest.

There are, at least, two alternatives for evaluating the effects of educational programs in this framework (e.g., Maddala, 1983, and Heckman, 1990). The  $\alpha$  parameter corresponds to the expected potential outcome of the program, i.e. how much an unemployed individual would gain if had he/she been placed into training.<sup>3</sup> A second alternative is to use the average impact of the training program on the log duration, i.e.

$$E(y|d = 1) - E(y|d = 0) = \alpha + \sigma[E(\omega|\xi > -\mathbf{z}\boldsymbol{\gamma}) - E(\omega|\xi \leq -\mathbf{z}\boldsymbol{\gamma})]. \quad (3)$$

---

<sup>1</sup>A multiple-spell version of the model is discussed in the final section.

<sup>2</sup>It is possible to introduce time dependence, dummy variables reflecting time intervals, etc. by letting the dependent variable in (1) be a logarithmic function of such variables. For the estimators considered in this paper this would, however, create problems to which there, at least, for now are no good solutions.

<sup>3</sup>This criterion becomes more involved if there are different shape parameters  $\sigma_1$  and  $\sigma_0$  corresponding to  $d = 1$  and  $d = 0$ .

This follows since

$$E(y|d = 1) = \mathbf{x}\boldsymbol{\beta} + \alpha + \sigma E(\omega|\xi > -\mathbf{z}\boldsymbol{\gamma}) \quad (4)$$

$$E(y|d = 0) = \mathbf{x}\boldsymbol{\beta} + \sigma E(\omega|\xi \leq -\mathbf{z}\boldsymbol{\gamma}). \quad (5)$$

To evaluate the measure in (3) we either have to make a distributional assumption or provide some nonparametric estimator for the conditional expectations. This reasoning carries over to the natural time scale  $t$ , on which the ratio  $E(t|d = 1)/E(t|d = 0)$  will be independent of the characteristics in the  $\mathbf{x}$  vector, while the difference  $E(t|d = 1) - E(t|d = 0)$  will not. Here,  $\exp(\alpha)$  corresponds to the former measure on the  $\ln t$  scale. For small  $\alpha$  the effect in percent is then approximately  $100\alpha$ . On the  $t$  scale  $\alpha > 1$  corresponds to an effect that prolongs the studied duration and  $\alpha < 1$  to a shortening effect. The ratio  $E(t_\alpha|d = 1)/E(t_0|d = 1) = \exp(\alpha)E(\exp(\sigma\omega)|d = 1)/E(\exp(\sigma\omega)|d = 0)$ , with  $t_0 = \exp(\mathbf{x}\boldsymbol{\beta}) \exp(\sigma\omega)$  and  $t_\alpha = t_0 \exp(\alpha)$ , corresponds to (3).

Depending on which measure is the one of ultimate interest it will have considerable impact on how we should estimate the unknown parameters of the models. The former measure,  $\alpha$ , makes semiparametric estimators based on relatively weak assumptions feasible.

If  $\tau = \exp(\omega)$  has hazard function  $\gamma_0(\tau)$  the hazard function corresponding to (1) can for given  $d$  be written

$$\lambda(t) = \gamma_0(t^\psi e^{-\psi(\mathbf{x}\boldsymbol{\beta} + \alpha d)}) \psi t^{\psi-1} \exp(-\psi(\mathbf{x}\boldsymbol{\beta} + \alpha d)),$$

where  $\psi = \sigma^{-1}$ . This hazard function will be proportional only under quite specific assumptions on the form of  $\gamma_0(\tau)$ . For a given density function of  $\tau$ ,  $g(\tau)$ , the density function of  $t$  given  $d$  is

$$f(t) = g(t^\psi e^{-\psi(\mathbf{x}\boldsymbol{\beta} + \alpha d)}) \psi t^{\psi-1} \exp(-\psi(\mathbf{x}\boldsymbol{\beta} + \alpha d)).$$

### 3. Estimation

In addition to selection, right censored durations are a rule rather than an exception. We need to account for the censored observations in order to avoid biased inferences. The assumption we make is the conventional one, i.e. that the censoring mechanism is independent of the duration  $t$  and to the right (Type 1 censoring or a weakly exogenous censoring mechanism). An observation is censored when  $y > c$ , where  $c$  is the possibly random censoring point, and we then set the indicator variable  $w = 0$  for that event. For noncensored observations,  $y \leq c$  and  $w = 1$ .

For inferential purposes one difficulty arises with giving a well-motivated bivariate distributional specification for  $(\omega, \xi)'$ . Such an assumption is needed for full maximum likelihood (ML) estimation of the two relationships (1)-(2) as well as for least squares (LS) estimation based on parametric assumptions. In the Tobit model context a bivariate normal assumption is the most widely used one, but alternatives such as a multivariate logistic distribution could be applied.



The general shape of the likelihood function is

$$\begin{aligned}
L &= \prod_{d=1} f(y|d=1)^w \Pr(y > c|d=1)^{1-w} \Pr(d=1) \\
&\quad \times \prod_{d=0} f(y|d=0)^w \Pr(y > c|d=0)^{1-w} \Pr(d=0). \tag{6}
\end{aligned}$$

Our main focus will be on estimators compromising between being based on as few assumptions on the distributions for  $\ln t$  and  $\delta$  as possible and being a tractable procedure. Therefore, the ML estimator will be given no further discussion in this paper, though ML gives all that is required for either measure. A least squares estimation procedure based on the bivariate normality assumption is given in Appendix B.

When the evaluation measure of key importance is  $\alpha$  (and we additionally wish to estimate the unknown vector  $\beta$ ) we consider two approaches to estimation that are both tractable and based on relatively weak assumptions. One approach is related to the trimmed mean and eliminates observations in the left tail of the  $y$  distribution to the same extent as the fraction that is missing due to censoring in the right tail. By this, symmetry is restored and the mean is a consistent estimator. This is the essence of the Powell (1986) symmetrically trimmed least squares estimator for censored data. In the second approach, we take the opposite stand and forecast the censored observations so as to obtain a full sample. An iterated use of forecasting and least squares estimation is the essence of the Buckley and James (1979) and Horowitz (1986) estimators.

These two approaches to estimation can both be extended to incorporate instrumental variables for dealing with the endogeneity of  $d$ , and they do not use explicit distributional assumptions in accounting for censoring. Note also that the selection equation is of no direct use for these estimators, which then are of a limited information type. The benefit of not accounting fully for the selection equation is a gain in robustness, but a potential loss of efficiency if the selection equation by coincidence would be correctly specified. The two estimators are introduced in an instrumental variable setting with equally many instruments/basic variables and unknown  $\beta$  and  $\alpha$  parameters. An extension into GMM estimators seems straightforward, but is not considered.

### 3.1 Limited Information Estimation

The extension of the Powell (1986) estimator to right censored data and to include instrumentation for  $d$  is straightforward. The key assumption for Powell's symmetrically trimmed estimator is that  $\omega$  has a symmetric distribution, while the estimator remains consistent and asymptotically normal also under heteroskedasticity in  $\sigma\omega$ .

Define  $\theta' = (\beta', \alpha)$ ,  $\mathbf{x}_i^* = (\mathbf{x}_i, d_i)$  and  $\hat{\mathbf{x}}_i = (\mathbf{x}_i, \hat{d}_i)$ , where  $\hat{d}_i$  is an instrumental variable for  $d_i$ . For instance,  $\hat{d}_i$  can be the predicted value from a first stage Probit model applied to (2) or from a regression of  $d$  on the explanatory variables and their cross-products.

The Powell estimator adapted to right censoring (cf. Brännäs, 1992b, ch. 3) and instrumental variables may be written

$$\hat{\boldsymbol{\theta}} = \left[ \sum_{i=1}^n \mathbf{1}(\mathbf{x}_i^* \hat{\boldsymbol{\theta}} < c_i) \hat{\mathbf{x}}_i' \mathbf{x}_i^* \right]^{-1} \sum_{i=1}^n \mathbf{1}(\mathbf{x}_i^* \hat{\boldsymbol{\theta}} < c_i) \hat{\mathbf{x}}_i' \max(y_i, 2\mathbf{x}_i^* \hat{\boldsymbol{\theta}} - c_i). \quad (7)$$

For uncensored observations,  $c_i$  may be set at some large value so that the indicator functions in practise are equal to one and the maximum value in (7) is equal to  $y_i$ . The estimator is iterated until convergence. Santos Silva (1998) recently proposed a modification for the Tobit model version that both reduces the problem with nonconvergent iterations and speeds up iterations substantially. As we only observe the smaller of  $y_i$  and  $c_i$  an arbitrarily set  $c_i$  would seem to rule out the usefulness of the Santos Silva (1998) approach in the present setting.

A consistent estimator of the covariance matrix is of the form

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = n^{-1} \mathbf{C}^{-1} \mathbf{D} \mathbf{C}^{-1},$$

where

$$\begin{aligned} \mathbf{C} &= n^{-1} \sum_{i=1}^n \mathbf{1}(2\mathbf{x}_i^* \hat{\boldsymbol{\theta}} - c_i \leq y_i < c_i) \hat{\mathbf{x}}_i' \mathbf{x}_i^* \\ \mathbf{D} &= n^{-1} \sum_{i=1}^n \mathbf{1}(\mathbf{x}_i^* \hat{\boldsymbol{\theta}} < c_i) \min(\hat{\omega}_i^2, (c_i - \mathbf{x}_i^* \hat{\boldsymbol{\theta}})^2) \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i. \end{aligned}$$

Santos Silva (1998) found no strong reason to improve on the covariance matrix estimator suggested by Powell (1986) for the Tobit model. The covariance matrix estimator given above therefore only adapts the original Powell's estimator to right censored data and for the instrumental variables.

Honoré and Powell (1994) provide an estimator that also uses trimming but applies to nonsymmetric distributions as well. In a simulation they found the Powell estimator to be relatively robust against nonsymmetric distributions, but that their estimator and in particular the Buckley and James (1979) estimator performed very well in terms of bias and MSE. Unfortunately, the Honoré and Powell (1994) estimator requires an observed censoring variable also when censoring is not registered. This is too strong an assumption for most purposes. For the adapted Powell estimator we chose to set the censoring values at some large value for noncensored observations. As the Honoré and Powell (1994) estimator is based on pairs of observations this may be a more sensitive setting. For this reason the estimator is left outside this study.

Buckley and James (1979, for right censored duration data) and Horowitz (1986, for left censored at zero - Tobit model data) propose estimators that resemble the EM-algorithm and for duration data estimate  $E(\omega | \omega > c - \mathbf{x}^* \boldsymbol{\theta})$  nonparametrically. In this case, there is no symmetry assumption on  $\omega$ . An extension to allow for instrumental variables appears straightforward here as well. This estimator does not

allow for heteroskedasticity, however. Brännäs and Laitila (1989) and Moon (1989) demonstrate considerable sensitivity in this respect. James and Smith (1984) proves consistency for a simple model and Wu and Zubovic (1995) study the fluctuations of the algorithm.

For given  $d_i$ ,

$$E(y_i|y_i > c_i) = \mathbf{x}_i\boldsymbol{\beta} + \alpha d_i + \sigma E(\omega_i|\omega_i > (c_i - \mathbf{x}_i\boldsymbol{\beta} - \alpha d_i)/\sigma),$$

where we employ the nonparametric Kaplan-Meier or product-limit estimator for the estimation of the conditional expectation in the right hand side expression. Let  $\hat{e}_i = y_i^* - \mathbf{x}_i\hat{\boldsymbol{\beta}} - \hat{\alpha}d_i$  with  $y_i^* = \min(y_i, c_i)$  be the residual. The product-limit estimator of the survivor function,  $1 - F(\cdot)$ , is

$$1 - \hat{F}(\omega) = \prod_{\hat{e}_{(i)} \leq \omega} \left(1 - \frac{m_{(i)}}{n_{(i)}}\right)^{\rho_{(i)}}, \quad (8)$$

where  $\hat{e}_{(1)} < \hat{e}_{(2)} < \dots < \hat{e}_{(n)}$  are the size ordered values on  $\hat{e}_i$ ,  $n_{(i)}$  is the number of residuals at risk at  $\hat{e}_{(i)-}$  (i.e. just to the left of  $\hat{e}_{(i)}$ ),  $m_{(i)}$  is the number of completed durations at  $\hat{e}_{(i)}$ , and  $\rho_{(i)} = 1$  if  $m_{(i)} > 0$  and  $\rho_{(i)} = 0$  otherwise. If the largest  $\hat{e}_i$  is censored, the remaining mass is assigned to  $\hat{e}_{(n)}$  by convention. At iteration  $r$  of the algorithm we have the following predictor for censored observations

$$\hat{y}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}}_r + \hat{\alpha}_r d_i + \sum_{j:\hat{e}_j \geq \hat{e}_i} \hat{f}(\hat{e}_j)\hat{e}_j/[1 - \hat{F}(\hat{e}_i)], \quad (9)$$

where  $\hat{f}(\hat{e}_j)$  is an estimator of  $f(\cdot)$  that can be obtained from  $\hat{F}(\cdot)$ . Instead of OLS estimation in the regression of  $y$  or  $\hat{y}$  (for censored observations) on  $\mathbf{x}$  and  $d$  as in the original Buckley and James (1979) estimator, we propose an instrumental variable estimator

$$\hat{\boldsymbol{\theta}}_{BJ} = (\hat{\mathbf{X}}'\hat{\mathbf{X}}^*)^{-1}\hat{\mathbf{X}}'\tilde{\mathbf{y}}, \quad (10)$$

where  $\tilde{\mathbf{y}}$  has elements  $y$  or  $\hat{y}$ . The estimation of the covariance matrix of  $\hat{\boldsymbol{\theta}}_{BJ}$  for the basic model is discussed in Brännäs (1992b). An extension to the present instrumental variable context is straightforward:

$$Cov(\hat{\boldsymbol{\theta}}_{BJ}) = \hat{\sigma}^2(\hat{\mathbf{X}}'\hat{\mathbf{X}}^*)^{-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}(\hat{\mathbf{X}}^*\hat{\mathbf{X}})^{-1}.$$

For  $\hat{\sigma}^2$  we use only residuals for full observations. Using a kernel estimator of  $f(\cdot)$  and some smoother of  $F(\cdot)$  may eliminate the potential cyclicity of the algorithm. The cyclicity arises due to the discrete jumps in  $\hat{f}(\cdot)$  and  $\hat{F}(\cdot)$ .

These estimators provide  $\hat{\alpha}$  estimates and correspond to the first measure. In Section 5 we will indicate a way to estimate the second measure of Section 2 as well. To obtain the second measure one could also use normal distribution results and a first stage Probit ML estimator, or say, run a polynomial regression of  $\hat{\omega}$  on  $\hat{\xi}$  which

then is used to approximate the conditional expectations. We will not pursue such ideas further in this paper.

Within the hazard function setting accounting for selection is quite straightforward. As one approach we suggest: start with the conventional likelihood equations based on a hazard function  $\lambda_0(t) \exp(\mathbf{x}^* \boldsymbol{\theta})$  for the parameter vector  $\boldsymbol{\theta}$ , i.e.

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \mathbf{x}_i^* (y_i - \Lambda_i(\boldsymbol{\theta})),$$

where  $\Lambda_i(\boldsymbol{\theta})$  is the integrated hazard function and may depend on other parameters than  $\boldsymbol{\theta}$  as well. Leave the residual part,  $y_i - \Lambda_i(\boldsymbol{\theta})$ , intact. In the instrumental variable part,  $\mathbf{x}_i^*$ , replace the  $d$  variable by  $\hat{d}$  or a whole set of instrumental variables (e.g., those included in  $\mathbf{z}$ ). Finally, use GMM based on the modified and possibly unaltered (for other parameters) likelihood equations instead of ML estimation to obtain estimates.

#### 4. Monte Carlo Study

This small Monte Carlo aims at: [1] studying the OLS (as a reference), the Powell, and the Buckley-James (the latter two with instrumental variables and accounting for censoring) estimators of the  $\alpha$  parameter in terms of their small sample performances, and [2] studying the power properties of conventional  $t$ -tests of  $\alpha = 0$  against  $\alpha \neq 0$ . We recognize from the outset that this study can be at most partial. The outcome could therefore screen out estimators/tests that underperform severely. A positive outcome cannot be overly stressed, however.

The study is based on the simple data generating process

$$\begin{aligned} y_i &= 1 + x_i + \alpha d_i + \omega_i \\ d_i &= \begin{cases} 1, & \delta_i = 2.5 - z_i + \xi_i > 0 \\ 0, & \delta_i \leq 0. \end{cases} \end{aligned}$$

We let  $\alpha = 0, 0.1, 0.2, 0.3$ , and  $\rho = \sigma_{\omega\xi} = -0.7, -0.35, 0, 0.35, 0.7$ . In the first experiment the distributions of  $\omega_i$  and  $\xi_i$  are both normal with zero means and unit variances, while in the second experiment  $\omega_i$  is a linear combination of the normal  $\xi_i$  and a standardized extreme-value variable  $\omega_i^*$ . The explanatory variables  $x_i$  and  $z_i$  are generated as independent from uniform  $(0, 5)$  distributions and kept constant during replications. Censoring is varied as  $c_i = \mu + \eta_i$ , where  $\eta_i$  is standard normal and  $\mu = 3.5$  (corresponding to 50-53 percent of censored observations) and 4 (38-41 percent censoring). For the second experiment the censoring percentages are marginally higher. The sample sizes are set to 500 and 5000, and 1000 replications are run in each cell. The nominal test size is set at 0.05. As an instrumental variable we use  $\hat{d}_i = z_i$ . The direct correlation between  $d$  and  $\hat{d}$  is about -0.77.

For the Powell estimator preliminary inspection reveals a cyclical problem; estimates at the iterative step  $k$  may be equal to estimates at  $k - 2$ ,  $k - 3$ , or so on. In the simulations we therefore check over 5 steps for identical estimates. If there

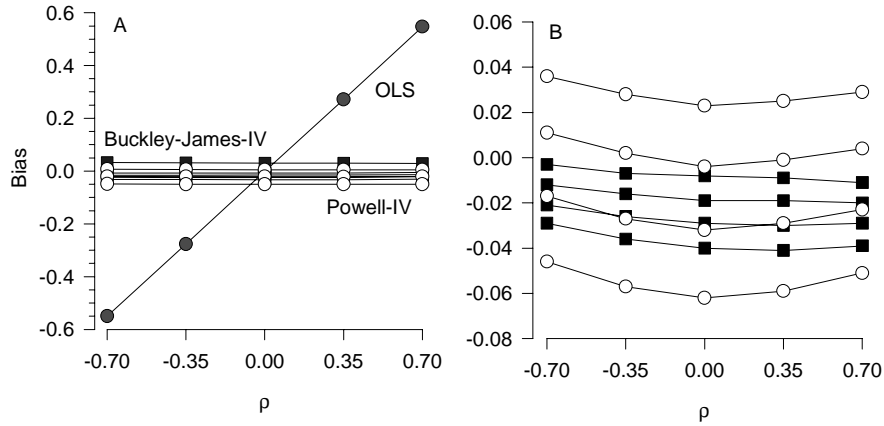


Figure 1: Biases of the OLS (only in left hand graph A) and the adapted Powell (black and square marker) and Buckley-James (white and circular marker) estimators for  $n = 500$  and 40 percent censoring. Graph A is for the normal distribution case and graph B for the extreme-value/normal distribution case.

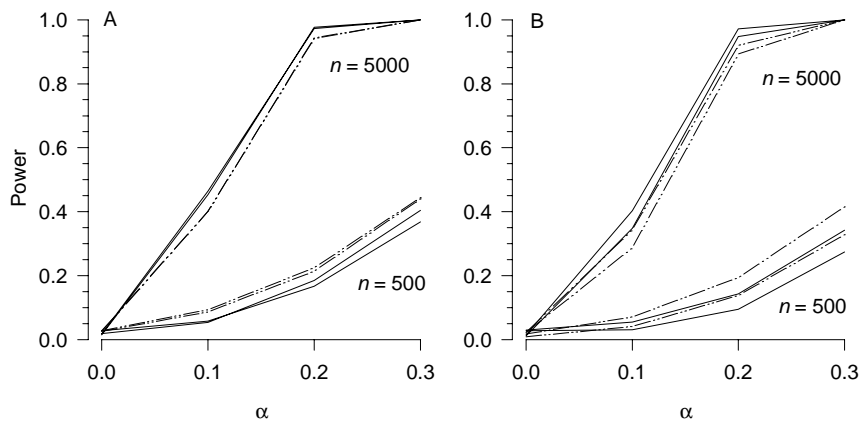


Figure 2: Power functions for adapted Powell (solid line) and Buckley-James (dot-dashed line) test statistics of  $\alpha = 0$  against  $\alpha \neq 0$  ( $n = 500$  and  $5000$ , 40 percent censoring, and  $\rho = 0$  and  $0.7$ ). Graph A is for the normal distribution case, and graph B is for the extreme-value/normal case.

are smaller than 0.00001 differences, we interrupt iterations and take the average of the final four iterations to be the estimates. The checking starts at the 6th iterative step. The inspection suggests that differences in estimates over the cycles are quite small and may go unnoticed with a less stringent convergence criterion than the used: no difference between successive iterations should exceed 0.00001. No cyclical problem was detected for the Buckley-James estimator.

In Figure 1 we report bias results of the OLS, Powell and Buckley-James estimators for  $n = 500$ . The left graph is for the normal distribution case and the right hand graph for the case of the linear combination of extreme-value and normal. The OLS estimator expectedly stands out with its huge bias that is invariant with respect to  $\alpha$  but increases linearly with the value of  $\rho$ .<sup>4</sup> In graph A both the adapted Powell and Buckley-James estimators have biases that only marginally become larger with  $\rho$ . These estimators also have marginally larger biases for larger values on  $\alpha$ . The bias of the Buckley-James estimator is slightly larger than that of the Powell estimator for  $\alpha = 0.3$  but smaller at  $\alpha = 0$ . All biases but those of the Buckley-James estimator for  $\alpha \leq 0.1$  are negative. For  $n = 500$ , the variance of the Powell estimator is 0.018 for 40 percent censoring and 0.015 for the Buckley-James estimator. For 50 percent censoring we have variances 0.020 and 0.015, respectively. For this censoring degree, the variances drop to 0.002 and 0.002-0.003, respectively, for  $n = 5000$ . The biases of the two adapted estimators are slightly higher for the higher censoring degree. In graph B we notice that the Buckley-James estimator has larger absolute bias than the Powell estimator for two cases ( $\alpha = 0$  and 0.3). It follows from this limited experiment that there are no huge differences in mean square error behavior between the adapted Powell and Buckley-James estimators.

Powell and Buckley-James estimators that do not account for selection have biases not too different from those of the OLS estimator. From this we conclude that, at least in this case, selection has a biasing effect that is more severe than has censoring. For  $n = 500$  and 40 percent censoring the Powell estimator required 6-7 iterative steps, while the Buckley-James estimator required 9-12 steps. The Powell estimator is much faster to calculate. In particular for  $n = 5000$  the repeated sorting required in the Buckley-James estimator makes simulating it quite time consuming.

The power functions for test statistics of  $\alpha = 0$  against  $\alpha \neq 0$  based on the adapted Powell and Buckley-James estimators are displayed in Figure 2 for the two sample sizes, 40 percent censoring,  $\rho = 0$  and 0.7, and for the two experiments (graph A and B, respectively). Obviously, there is no strong effect of  $\rho$  as lines are quite close in particular for the larger sample size. There is an expected strong sample size effect that also reverses the order between the two tests. The Powell test is more powerful for  $n = 5000$ , while the Buckley-James test is more powerful for the smaller sample size. The sizes of both test statistics are throughout significantly too small. For the Buckley-James estimator in the normal distribution experiment with 50 percent censoring the sizes are 0.059 and 0.060 for the two levels of  $\rho$ , respectively.

---

<sup>4</sup>Obviously, the biases, variances, etc. of the OLS estimator can be analytically determined under distributional assumptions. Under normality some of the results given in Appendix B will be useful.

Table 1: Variables and descriptive statistics ( $n = 15990$ ).

Variable	AEI		LMT	
	Mean	Standard-deviation	Mean	Standard-deviation
Insider time	87.9	60.5	187.6	107.6
Gender (Female=1)	0.72	0.45	0.42	0.49
Age	33.1	7.1	30.4	9.0
Children at home	1.08	1.18	0.63	1.06
Disability	0.07	0.26	0.12	0.32
Nationality (Swedish)	0.93	0.25	0.85	0.36
Nationality (Nordic country, non-Swedish)	0.03	0.16	0.02	0.13
Nationality (Europe, non Nordic)	0.02	0.13	0.08	0.27
Education (Primary school)	0.22	0.41	0.19	0.39
Education (Secondary school) <sup>a</sup>	0.09	0.29	0.29	0.46
Education (Post-secondary school)	0.06	0.23	0.15	0.36
Unemployment benefit (1997)	0.86	0.35	0.66	0.47
Unemployment, cash assistance (1997)	0.03	0.16	0.02	0.15
Unemployment rate (counties, 1998)	6.71	1.48	6.48	1.43
Days in event database (1995-96) <sup>b</sup>	402.1	267.2	413.7	258.9
Number of states in event data base (1995-96)	3.51	2.72	3.97	2.72
Ending training May–June 1998	0.83	0.38	0.31	0.46
number of observations	14 578		1 412	

Notes: Most of the data comes from the event history database Händel kept at the Swedish Labour Market Board for administrative purposes.

<sup>a</sup> Secondary school is the three year form, i.e. 2 year secondary serves as a base.

<sup>b</sup> As unemployed.

Neither is significantly different from the nominal size.

## 5. Empirical Results

The Swedish Adult Education Initiative (AEI) started in July 1997 and targets primarily adult unemployed who lack a three year secondary school education. Employees with a short education can also be included. Axelsson and Westerlund (1999) give an extensive background description on programs and participants.

The results presented in this section are based on a sample of 15 990 individuals who participated in AEI or labor market training (LMT) programs ending in either of the periods May – June 1998 (78 percent of the observations) or December 1997 – January 1998 (22 percent). Participant were unemployed before entering either program. The evaluation variable under study is the total time outside the active or passive measures (collectively labelled unemployment) of the Labor Market Board. The total time then contains time spent in employment (about 15 percent of the

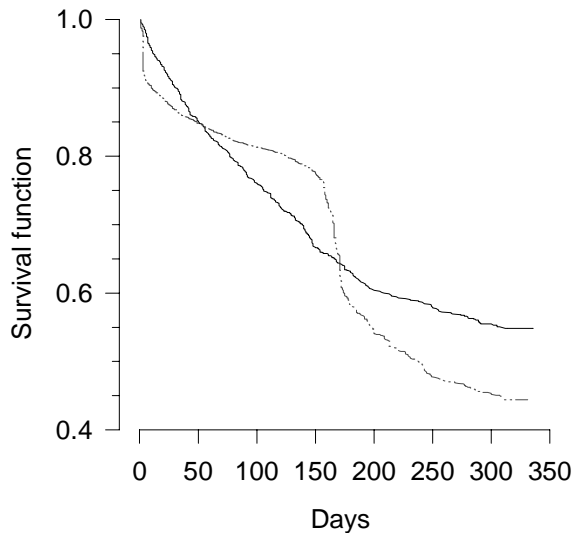


Figure 3: Kaplan-Meier estimated survival functions for post-program insider time in days for AEI (dot-dashed line) and LMT (solid line).

exits are to employment) as well as outside the labor force, e.g., in studies (about 73 percent). For the duration variable we choose to use the term insider time for brevity.<sup>5</sup>

In Table 1 we give a summary of the utilized variables together with descriptive statistics. In Figure 3 the survival functions for the AEI and LMT groups are exhibited. The censoring degree is as high as 77.4 percent of all durations. Obviously, this can be interpreted as a good signal as it indicates that about 77 percent of all individuals were occupied in studies or had employment at the time when durations were obtained. About 9 percent of the individuals come from LMT and among them about 59 percent had ongoing studies or employment. It should be observed that AEI programs are intended to prepare participants, e.g., for further studies as well as for work. Table 1 shows that average insider times are much longer for LMT participants. These to a lesser extent pursue further studies. The average participant in either program has a history of unemployment reflected by the large average number of days of unemployment during 1995-1996 as well as the number of states occupied. The differences between the two participant groups are not significant. Figure 3 indicates a median insider time of 238 days for AEI participants while the median exceeds 350 days for LMT participants. The sudden drop at only a few days as well as at about 150 days for AEI participants appears to arise from

<sup>5</sup>No connection to the insider-outsider theory is intended.



Table 2: Parameter estimates for the duration model.

Variable	Powell		Buckley-James	
	Est.	s.e.	Est.	s.e.
Gender	0.115	0.100	-0.006	0.017
Age/10	0.021	0.051	0.007	0.005
Children at home	-0.015	0.032	-0.003	0.007
Disability	-0.022	0.129	-0.052	0.031
Nationality (Swedish)	0.084	0.203	0.074	0.053
Nationality (Nordic country)	0.039	0.294	0.067	0.072
Nationality (Europe)	-0.053	0.300	-0.075	0.073
Education (Primary school)	0.016	0.148	-0.036	0.020
Education (Secondary school)	0.170	0.175	-0.037	0.035
Education (Post secondary school)	-0.316	0.192	0.089	0.037
Unemployment benefit	-0.715	0.273	-0.607	0.027
Unemployment, cash assistance	-0.343	0.232	-0.379	0.054
Unemployment rate/10	0.081	0.278	0.070	0.056
Days in event database/1000	-0.316	0.192	-0.185	0.043
Number of states in event database/10	-0.174	0.179	-0.170	0.043
Participant in AEI	-1.640	0.538	-1.547	0.090
Ending late 1998	1.313	0.083	1.377	0.028
Constant	5.305	0.377	5.260	0.091

Note: Buckley-James  $\hat{\sigma}^2 = 1.018$ .

those that end or have a break from AEI report themselves as unemployed from the early days of the summer.

For the creation of an instrumental variable for  $d$  we use a first stage Probit model estimated by ML. The results and the specification are given in Table C1 in Appendix C. The correlation between  $d_i$  and  $\hat{d}_i = \Pr(d_i = 1|\hat{\gamma})$  is 0.59. Running a regression between the  $\hat{d}_i$  obtained by the selection model and the explanatory variables of the model gives an  $R^2$  of 0.30, which we regard as quite innocent.

The estimation results for the duration model are summarized in Table 2. The given estimates (Powell and Buckley-James) correspond to the estimators that account for both censoring and selection. The key parameter of interest is that of the AEI dummy variable and for both estimators we get a negative sign corresponding to shorter insider times for AEI participants. The effect is in fact substantial as it corresponds to 80.6 percent shorter insider times for the Powell estimator and 78.7 percent for the Buckley-James estimator. The effect is significant for both estimators. The only other variables that have significant effects using the Powell estimator are those of the unemployment benefit and the indicator for ending education in the early part of the summer 1998. The former variables has the effect of shortening insider times which is the expected result. Using the Buckley-James estimator indicates some differences, but the main conclusions remain even stronger. As the

Buckley-James estimator has substantially smaller standard errors several variables now have significant effects. Both forms of unemployment support have significant effects of the expected negative signs. The history variables, days and number of states, both have significant effects of expected signs as has the ending late 1998 indicator variable.

In Table C2 in Appendix C we give estimates obtained by inappropriately used OLS, Powell and Buckley-James estimators. These do not account for censoring, selection or both. These estimators are utilized to indicate the impact of neglecting censoring and selection and not for their estimates per se.<sup>6</sup> We note, however, that signs in most cases remain unaltered between these estimators and those of Table 2. The sizes of parameter estimates are frequently smaller. The effect of the AEI program participation is substantially smaller when the selection effect is neglected. For these estimates the AEI insider times are 71.3, 73.5 and 69.8 percent shorter, respectively. Notable is also the much larger standard errors of the Powell than of the OLS and Buckley-James estimators. In most cases the OLS and Buckley-James estimates are relatively close to each other.

Next we turn to the implied density and hazard functions of the estimated models. As the basis for this exercise we choose to use the Buckley and James (1979) estimator, as the distribution function,  $F_{\Omega}(\omega)$ , of  $\omega$  is already estimated in this case. The model estimated by the Powell estimator would require that we first estimate  $F_{\Omega}(\cdot)$ . In Figure 4 we give the estimated density function  $\hat{f}_{\Omega}(\omega)$ . The used kernel estimator is based on a standard normal density function as the kernel and the bandwidth is set equal to 1. As is obvious the estimated density is negatively skewed with a fat left tail. The tails are fatter than those of a standard extreme-value distribution (corresponding to Weibull distributed duration  $t$ ). The estimated asymmetric density function is not very different from the density employed in the Monte Carlo study as the second experiment. Hence, though the Powell estimator presumes symmetry it was not found to be excessively sensitive to asymmetry, and therefore the Powell estimates can not be disregarded for this reason. Obviously, the skewness speaks against the use of normality based ML estimation or the least squares procedure of Appendix B.

To obtain the hazard function of  $t$  we use the result that the distribution function of  $t$  for given  $d$  can be written in terms of the  $F_{\Omega}$  distribution:

$$F(t) = F_{\Omega}[(\ln t - \mathbf{x}\boldsymbol{\beta} - \alpha d)/\sigma].$$

The hazard function conditional on  $d$  can with substituted estimates be written on the form

$$\hat{\lambda}(t) = \frac{\hat{\sigma}^{-1}t^{-1}\hat{f}_{\Omega}[(\ln t - \mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\alpha}d)/\hat{\sigma}]}{1 - \hat{F}_{\Omega}[(\ln t - \mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\alpha}d)/\hat{\sigma}]}.$$

---

<sup>6</sup>To test formally against a selection effect using a Durbin-Wu-Hausman test is not straightforward in this case. The test rests on one estimator being efficient (and consistent) under the null of no selection effect. Obviously, we have no such estimator.

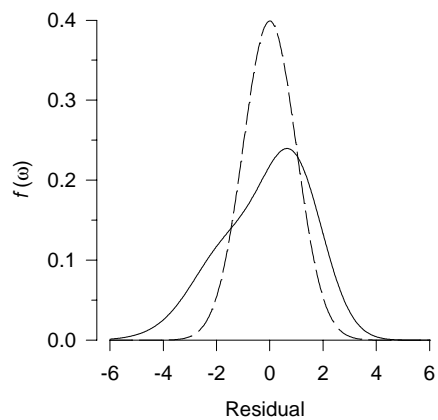


Figure 4: Estimated density function  $\hat{f}(\omega)$  (solid line) and standard normal density function (dashed line).

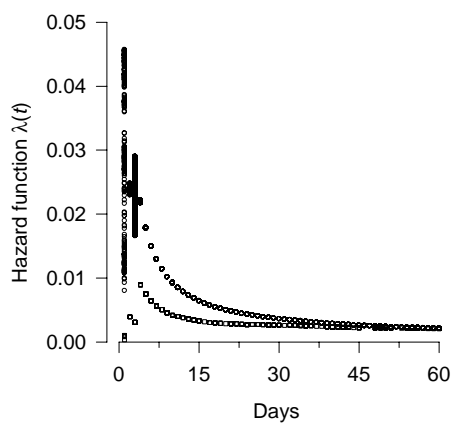


Figure 5: Estimated hazard function based on Buckley-James estimates of the model (uncensored observations; AEI, circle marker, and LMT square marker).

Here,  $(\ln t - \mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\alpha}d)/\hat{\sigma}$  corresponds to the standardized residual. By evaluating the function at noncensored values we get the hazard functions of Figure 5. The graph is based on observed values and the markers indicate individual values on the calculated hazard function values. One conclusion is that proportionality does not appear to receive support for low values on  $t$ , while for higher values the functions coincide. Further support for this claim is that there appears to be an interior maximum which rules out a Weibull distributed  $t$  – the only proportional hazard that could arise from the present model setup. The second conclusion is that there appears to be negative duration dependence, so that returning to the unemployment state gets less likely with time.

As a final point we return to the second measure of the participation effect, cf. (3). Given residuals  $\hat{\omega}_i$  we estimate the conditional expectation for given  $d_i = 0$  and  $d_i = 1$ , respectively, by their sample analogs. The measure is on the log scale  $-1.672$  for the Buckley-James estimator, i.e. larger than  $\hat{\alpha} = -1.547$ . The corresponding measures for the day scale  $t$  are 0.053 vs.  $\exp(-1.547) = 0.213$ . The former corresponds to -94.7 percent.

## 6. Discussion

The paper has suggested instrumental variable adaptations to two well-known estimators for censored data. One advantage with the suggested procedures is that only weak distributional assumptions need to be invoked. The results of the small Monte Carlo study suggests that the two estimators are quite well-behaved in sample size regions that can be expected to arise in empirical work. The estimators by and large give analogous results in the empirical study, albeit the adapted Powell estimator has larger standard errors. This may be due to using less information for this estimator but also due to heteroskedasticity. Correcting the Buckley and James covariance matrix estimator for heteroskedasticity does not alter standard errors by much.

Among other estimators that are based on weak assumptions it appeared difficult to adapt the Honoré and Powell (1994) estimator for our purposes. The Robinson (1988) estimator cannot be applied as the  $\mathbf{x}$  variables will reappear in the correction factor for censoring. In some approximate version with  $\mathbf{x}$  variables that do not enter directly but rather in some mixed form the Robinson estimator could easily be applied for the present purposes.

For models based on specified hazard functions no particular problem arises, except that conventional programs for duration data can not be applied. For correct instrumental variable estimation, the model in the residual part of the moment condition should be based on  $d$ , while  $\hat{d}$  is to be used in the instrumental variable part. Obviously, the proportional hazard function is at present the only reasonable point of departure if some variable(s) are time dependent.

Recent evaluation research (e.g., Ham and LaLonde, 1996) is frequently based on multiple-spell models. With the present model type we may consider, say, the

following model version

$$\begin{aligned}
y_{ai} &= \mathbf{x}_i\boldsymbol{\beta} + \alpha d_i + \omega_{ai} + \varepsilon_i && \text{(after)} \\
y_{pi} &= \mathbf{v}_i\boldsymbol{\lambda} + \omega_{pi} + \psi_p\varepsilon_i && \text{(program)} \\
y_{bi} &= \mathbf{w}_i\boldsymbol{\theta} + \omega_{bi} + \psi_b\varepsilon_i && \text{(before)} \\
d_i &= \mathbf{1}(\mathbf{z}_i\boldsymbol{\gamma} + \xi_i + \psi_d\varepsilon_i \geq 0) && \text{(selection rule).}
\end{aligned}$$

These equations are correlated through the common random effect  $\varepsilon_i$  for individual  $i$  and also due to correlation between, at least, the pair  $\omega_{ai}$  and  $\xi_i$ . It still remains legitimate to focus only on the  $y_{ai}$  equation and to use an instrumental variables estimator for only this equation. Obviously, there will be an additional loss in efficiency from not estimating all the equations jointly. The  $y_{ai}$ ,  $y_{pi}$  and  $y_{bi}$  equations could be estimated jointly by IV/GMM techniques, and this requires some extensions to the considered estimator. Estimation of these types of models with a left censored  $y_{ai}$  variable was first considered by Nelson and Olson (1978) and Amemiya (1979). Brännäs (1992b) discussed adaptations to these estimators for right censored data by using the Powell (1986) estimator. In this framework it is obviously important to reflect on simultaneity vs. conditional interpretations, that reflect the time ordering of spells. The current interest in discrete mixing distributions for  $\varepsilon_i$  is here facing an identifiability barrier as features of, e.g.,  $\omega_{ai}$  and  $\varepsilon_i$  can not be separately estimated unless additional information or assumptions are introduced. If that could be done, the  $y_{ai}$  is the dependent variable of a switching regression model with unknown sample separation.

When it comes to the empirical exercise, the results are very much the expected ones. The dependent variable, which we for brevity reasons chose to call insider time, mainly represents the times to employment and further studies. During studies an individual is outside the labor force. For those leaving AEI, a substantial fraction (about 73 percent of the exits are to studies) of the individuals advance to higher levels of education within AEI or elsewhere. As the AEI attempts to enhance the educational level this may be as positive an effect as it is that some end up in regular employment. The newly developed estimators suggest that AEI participants have significantly shorter times before re-entering unemployment. On comparison of estimates it appears that the selection effect on estimates is stronger than that of censoring.

## Appendix A: Present Value

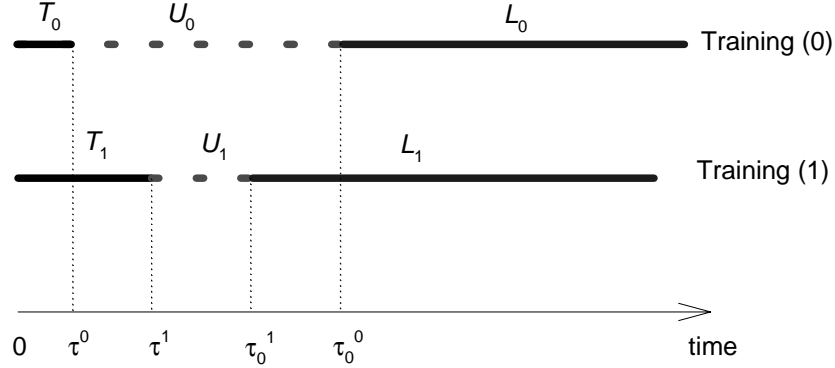


Figure A: The spells used for present value calculations [ $T$  indicates training with benefit  $B_U$ ,  $U$  is unemployment with benefit  $B_U$ , and  $L$  is work with wages  $w^0$  (Training state, 0) and  $w^1$  (Training state, 1)].

We give a simple and stylized example, cf. Figure A, of the net present value calculation. Under the Training (0) and Training (1) regimes we get the present values

$$\begin{aligned}
 P^0 &= \int_0^{\tau_0^0} B_U e^{-\theta t} dt + \int_{\tau_0^0}^{\tau_0^1} B_U e^{-\theta t} dt + \int_{\tau_0^1}^{\infty} w^0 e^{-\theta t} dt \\
 P^1 &= \int_0^{\tau^1} B_U e^{-\theta t} dt + \int_{\tau^1}^{\tau_0^1} B_U e^{-\theta t} dt + \int_{\tau_0^1}^{\infty} w^1 e^{-\theta t} dt,
 \end{aligned}$$

where  $\theta$  is the discount rate. Hence,  $B_U$  is the compensation under unemployment as well as under both types of training. It follows that the difference  $\delta$  is a random variable since the  $\tau_0^0, \tau_0^1, w^0$  and  $w^1$  are assumed to be random variables:

$$\delta = P^1 - P^0 = \frac{B_U}{\theta} \left( e^{-\theta \tau_0^0} - e^{-\theta \tau_0^1} \right) + \frac{1}{\theta} \left( e^{-\theta \tau_0^1} w^1 - e^{-\theta \tau_0^0} w^0 \right)$$

Therefore,  $\delta$  and the duration variables are obviously correlated and will produce a selection effect in the duration model. If the unemployment spell for trainees (0) follows an exponential distribution with parameter  $\mu_0$  and for trainees (1) with parameter  $\mu_1$ , both spells are independent of the random wages with  $E(w^1) = \eta + \alpha = E(w^0) + \alpha$ , and in addition times in training are fixed, then the expected difference is

$$E(\delta) = E(P^1 - P^0) = \frac{1}{\theta} [c_0 (B_U - \eta) + c_1 (\eta + \alpha - B_U)]$$

where  $c_0 = \mu_0 \exp(\mu_0 \tau_0^0) / (\mu_0 + \theta)$  and  $c_1 = \mu_1 \exp(\mu_1 \tau_0^1) / (\mu_1 + \theta)$ .

## Appendix B: Full Information Estimation

Consider the case when the training impact is measured by (3). The following estimator is based on joint normality in  $(\omega, \xi)'$ . Obviously, the ML estimator may be applied in this case, but we prefer to expand on a procedure that is closer to least squares. Under normality and using (4-5) we may write the model for full observations as

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \alpha d_i + \sigma\sigma_{\omega\xi}\lambda_i + \varepsilon_i, \quad (\text{B1})$$

where  $\sigma_{\omega\xi}$  is the covariance (and the correlation in this case) between  $\omega$  and  $\xi$  and  $\lambda_i = \phi(\mathbf{z}_i\boldsymbol{\gamma})/\Phi(\mathbf{z}_i\boldsymbol{\gamma})$  with  $\phi(\cdot)$  and  $\Phi(\cdot)$  the standard normal density and distribution functions, respectively. Here,  $E(\varepsilon_i) = 0$  and the conventional normalization  $V(\xi_i) = 1$  was already employed. For censored observations we need the conditional expectation expression

$$E(y_i|y_i > c_i) = \mathbf{x}_i\boldsymbol{\beta} + \alpha d_i + \sigma\sigma_{\omega\xi}\lambda_i + E(\varepsilon_i|\varepsilon_i > c_i - \mathbf{x}_i\boldsymbol{\beta} - \alpha d_i - \sigma\sigma_{\omega\xi}\lambda_i), \quad (\text{B2})$$

where for evaluating the final term a standardization by the variance of  $\varepsilon_i$ ,  $\sigma_{\varepsilon_i}$ , is required. Under the normality assumption we get

$$V(\varepsilon_i) = \sigma_{\varepsilon_i}^2 = \sigma^2\sigma_{\omega\xi}^2 \{1 - (\mathbf{z}_i\boldsymbol{\gamma})\lambda_i\varphi_i - \lambda_i^2\varphi_i^2\},$$

where we have used the notation  $\varphi_i = [d_i - \Phi(\mathbf{z}_i\boldsymbol{\gamma})]/[1 - \Phi(\mathbf{z}_i\boldsymbol{\gamma})]$ .

Corresponding to (B2) we then get:

$$E(y_i|y_i > c_i) = \mathbf{x}_i\boldsymbol{\beta} + \alpha d_i + \sigma\sigma_{\omega\xi}\lambda_i + \sigma_{\varepsilon_i}\lambda_i^*, \quad (\text{B3})$$

where  $\lambda_i^*$  is defined as  $\lambda_i$  but evaluated at  $(c_i - \mathbf{x}_i\boldsymbol{\beta} - \alpha d_i - \sigma\sigma_{\omega\xi}\lambda_i)/\sigma_{\varepsilon_i}$ .

For estimation we will lay out the following estimation procedure that can be seen as an extension of the Heckman 2-step estimator and that is also related to the Fair as well as the EM-algorithm. Probit ML can be used to estimate  $\hat{\boldsymbol{\gamma}}$  and then to calculate  $\hat{\lambda}_i$ , while to get initial values on  $\boldsymbol{\theta}$  and on  $\sigma\sigma_{\omega\xi}$ , we may adopt, say, a nonideal OLS estimator based on (4-5) using observed values on  $y$  and  $c$  to regress on  $\mathbf{x}$ ,  $d$  and  $\hat{\lambda}$ . With such estimates in hand the Fair algorithm, that resembles an EM-algorithm, alternates between;

- [1 ] predicting observations for  $y > c$ ; for this  $\hat{y} = \hat{E}(y|y > c)$  in (B3) is used, with  $\sigma_{\varepsilon_i}$ ,  $\lambda_i$  and  $\lambda_i^*$  calculated from Probit ML estimates and the initial guesses on  $\alpha$ ,  $\boldsymbol{\beta}$  and  $\sigma\sigma_{\omega\xi}$  [the E-step], and
- [2 ] the observations  $y$  that correspond to fully observed durations are then combined with predicted observations  $\hat{y}$  and employed for LS estimation of  $\alpha$ ,  $\boldsymbol{\beta}$  and  $\sigma\sigma_{\omega\xi}$  in (B1) [the M-step].

The algorithm stops the iterations when changes between successive values on  $\boldsymbol{\theta}^* = (\alpha, \boldsymbol{\beta}', \sigma\sigma_{\omega\xi})'$  are small enough. To estimate the covariance matrix we will employ the White estimator

$$Cov(\hat{\boldsymbol{\theta}}^*) = (\mathbf{X}'_*\mathbf{X}_*)^{-1}\mathbf{X}'_*\mathbf{S}\mathbf{X}_*(\mathbf{X}'_*\mathbf{X}_*)^{-1},$$

where  $\mathbf{X}_*$  has columns consisting of  $\mathbf{x}_i$ ,  $d_i$  and  $\hat{\lambda}_i$  and where  $\mathbf{S}$  is a diagonal matrix with squared residuals as elements on the diagonal.



## Appendix C: Probit Model Result and Duration Model Estimates

Table C1: Probit model maximum likelihood estimates.

Variable	Est.	s.e.
Gender	0.647	0.082
Age	0.342	0.020
Age <sup>2</sup>	-0.004	0.0003
Children at home	-0.054	0.018
Disability	-1.140	0.142
Nationality (Swedish)	0.360	0.091
Nationality (Nordic country)	0.471	0.152
Nationality (Europe)	-0.190	0.120
Education (Primary school)	-0.088	0.047
Education (Secondary school)	-0.725	0.049
Education (Post secondary school)	-0.794	0.060
Unemployment benefit	0.782	0.070
× Gender	-0.017	0.090
Unemployment, cash assistance	0.769	0.145
× Gender	0.143	0.228
Unemployment rate	-0.081	0.012
Days in event database	-0.0001	0.0001
× Disability	0.001	0.0003
Number of states in event database	-0.043	0.009
× Disability	0.058	0.026
Constant	-7.008	0.365

Table C2: Parameter estimates for duration models neglecting censoring and/or selection.

Variable	OLS		Powell		Buckley-James	
	Est.	s.e.	Est.	s.e.	Est.	s.e.
Gender	0.059	0.019	0.089	0.074	0.011	0.017
Age/10	0.017	0.012	0.010	0.043	0.018	0.011
Children at home	-0.013	0.008	-0.017	0.032	-0.004	0.007
Disability	-0.036	0.032	-0.010	0.121	-0.039	0.031
Nationality (Swedish)	0.052	0.054	0.080	0.200	0.060	0.053
Nationality (Nordic country)	0.036	0.074	0.036	0.293	0.044	0.072
Nationality (Europe)	-0.053	0.075	-0.109	0.296	-0.053	0.073
Education (Primary school)	-0.031	0.021	-0.049	0.086	-0.031	0.020
Education (Secondary school)	0.019	0.028	0.063	0.110	0.016	0.027
Education (Post secondary school)	0.120	0.035	0.219	0.159	0.122	0.034
Unemployment benefit	-0.623	0.026	-0.817	0.148	-0.658	0.025
Unemployment, cash assistance	-0.397	0.055	-0.410	0.191	-0.427	0.053
Unemployment rate/10	0.091	0.057	0.127	0.242	0.096	0.055
Days in event database/1000	-0.183	0.044	-0.322	0.190	-0.197	0.043
Number of states in event database/10	-0.127	0.042	-0.163	0.176	-0.136	0.041
Participant in AEI	-1.248	0.033	-1.346	0.100	-1.198	0.032
Ending late 1998	1.039	0.021	1.256	0.079	1.320	0.021
Constant	4.941	0.080	5.195	0.312	4.961	0.078

Notes: The OLS estimator does not account for censoring nor selection. The Powell and Buckley-James estimators do not account for selection.

## References

- Amemiya, T. (1979). The Estimation of a Simultaneous-Equations Tobit Model. *International Economic Review* **20**, 169-181.
- Axelsson, R. and Westerlund, O. (1999). Deltagare i kunskapslyftet höstterminen 1997 och jämförelsegrupper: Arbetsmarknads- och inkomstförhållanden före utbildning. Umeå Economic Studies 498.
- Brännäs, K. (1992a). Residuals for Limited Dependent Variable Models. *Statistica* **LII**, 197-205.
- Brännäs, K. (1992b). *Econometrics of the Accelerated Duration Model*. Umeå Economic Studies 269.
- Brännäs, K. and Laitila, T. (1989). Heteroskedasticity in the Tobit Model. *Statistical Papers* **30**, 185-196.
- Buckley, J. and James, I. (1979). Linear Regression with Censored Data. *Biometrika* **66**, 429-436.
- Ham, J.C. and LaLonde, R.J. (1996). The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training. *Econometrica* **64**, 175-205.
- Heckman, J. (1990). Selectivity Bias: New Developments. *American Economic Review: Papers and Proceedings*, 313-318.
- Honoré, B.E. and Powell, J.L. (1994). Pairwise Difference Estimators of Censored and Truncated Regression Models. *Journal of Econometrics* **64**, 241-278.
- Horowitz, J. (1986). A Distribution-Free Least Squares Estimator for Censored Linear Regression Models. *Journal of Econometrics* **32**, 59-84.
- Hujer, R., Mauerer, K-O. and Wellner, M. (1999). Estimating the Effect of Vocational Training on Unemployment Duration in West Germany. *Jahrbücher für Nationalökonomi und Statistik* **218/5+6**, 619-646.
- James, I.R. and Smith, P.J. (1984). Consistency Results for Linear Regression with Censored Data. *Annals of Statistics* **12**, 590-600.
- Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.
- Melkersson, M. (1999). Unemployment Duration and Heterogenous Search Behavior among Swedish Disabled Workers. IFAU Working Paper 1999:5, Uppsala.
- Moon, C.-G. (1989). A Monte Carlo Comparison of Semiparametric Tobit Estimators. *Journal of Applied Econometrics* **4**, 361-382.
- Nelson, F.D. and Olson, L. (1978). Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables. *International Economic Review* **19**, 695-720.
- Powell, J.L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica* **54**, 1435-1460.
- Robinson, P.M. (1988). Root- $n$ -Consistent Semiparametric Regression. *Econometrica* **56**, 931-954.
- Santos Silva, J.M.C. (1998). Powell's SCLS: A Newton Type Algorithm, Influence Diagnostics and Alternative Covariance Matrix Estimators. Working Paper,

ISEG, Universidade Técnica de Lisboa.

Wu, C-S.P. and Zubovic, Y. (1995). A Large Scale Monte Carlo Study of the Buckley-James Estimator with Censored Data. *Journal of Statistical Computation and Simulation* **51**, 97-119.