DISCUSSION PAPER SERIES

No. 5895

SWEDISH LABOUR MARKET TRAINING AND THE DURATION OF UNEMPLOYMENT

Katarina Richardson and
Gerard J van den Berg

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5895.asp**

ISSN 0265-8003

SWEDISH LABOUR MARKET TRAINING AND THE DURATION OF UNEMPLOYMENT

Katarina Richardson, Institute for Labour Market Policy Evaluation (IFAU)
Gerard J van den Berg, Free University Amsterdam and CEPR

Discussion Paper No. 5895
October 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5895

October 2006

ABSTRACT

Swedish Labour Market Training and the Duration of Unemployment*

The vocational employment training program is the most ambitious and expensive training program in Sweden and a cornerstone of labor market policy. We analyze its causal effects on the individual transition rate from unemployment to employment by exploiting variation in the timing of treatment and outcome, dealing with selectivity on unobservables. We demonstrate the appropriateness of this approach in our context by studying the process leading to enrollment. We also develop a model allowing for duration dependence and unobserved heterogeneity (leading to spurious duration dependence) in the treatment effect itself, and we prove non-parametric identification. The data cover the population and include multiple unemployment spells for many individuals. The results indicate a large significantly positive effect on exit to work shortly after exiting the program. The effect at the individual level diminishes after some weeks. When taking account of the time spent in the program, the effect on the mean unemployment duration is often close to zero.

JEL Classification: C14, C41 and J64
Keywords: duration analysis, duration dependence, hazard rate, identification, program evaluation, selectivity bias, transition to work, treatment effect and vocational training

Katarina Richardson
IFAU
PO Box 513
751 20 Uppsala
SWEDEN
Email: katarina.richardson@ifau.uu.se

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=159403

Gerard J van den Berg
Department of Economics
Free University Amsterdam
De Boelelaan 1105
NL-1081 HV Amsterdam
THE NETHERLANDS
Email: gjvdberg@xs4all.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=126469

* We thank the Swedish National Labour Market Board (AMS) and Statistics Sweden (SCB) for their permission to use the data. We thank Richard Smith, Annette Bergemann, Håkan Regnér, Anders Harkman, and other participants in conferences in Stockholm, Uppsala, and Rotterdam, and seminars at Aarhus, Bristol, Mannheim, UCL London, Humboldt, Oxford, and LSE, as well as our colleagues at IFAU, for helpful comments. In addition, we thank Helge Bennmarker for help with the data.

Submitted 14 September 2006

1 Introduction

Training programs for the unemployed have been cornerstones of labor market policy for many decades. In Sweden, training programs have been used since 1918 and constitute an important part of the so-called Swedish model (or Nordic model) of labor market policy. Among Sweden's current programs, the employment training program (which we denote by its Swedish acronym AMU) is the most prestigious. AMU aims to improve the chances of unemployed job seekers to obtain a job, by way of substantive skill-enhancing courses. In 1997, on average 37,000 individuals were participating in AMU per month, which corresponds to over 10% of total unemployment.[1] AMU is the most expensive active labor market program in Sweden and as such it adds to the tax burden. Nevertheless, the number of evaluation studies is rather small, and most of these analyze the effect of AMU on the participants' annual earnings and/or use data from early eighties and/or data on special subgroups of unemployed workers, notably youths in Stockholm (see references below).

This paper provides a comprehensive empirical analysis of the effect of AMU on the individual transition rate from unemployment to employment. Note that the officially stated objective of AMU is to generate a positive effect. The results are of obvious importance for the evaluation of the AMU program and the underlying "Swedish model". In addition, they are of importance in the light of the recent policy shifts in many other countries towards an increased use of active measures of bringing the unemployed back to work, notably by way of reschooling unemployed workers with low skills or obsolete qualifications (see e.g. Fay, 1996).

We use matched longitudinal register data of the full population of individuals who were unemployed in Sweden within the period from January 1, 1993 until June 22, 2000. The data include detailed records from employment offices, records from unemployment insurance agencies, and income tax records. The employment office data report the exact types of training and the corresponding dates of entry and exit.

The empirical analysis applies a methodology in which the information in the *timing* of events (like the moment at which the individual enrolls in training and the moment at which he finds a job) is used to estimate causal treatment effects in the presence of "selectivity on unobservables". This Timing of Events approach involves estimation of models that simultaneously explain the duration until an outcome of interest and treatment status. The treatment is allowed to affect the main outcome by way of the time rate at which the latter occurs after the treatment. Abbring and Van den Berg (2003) provided a formal underpinning of the approach by proving non-parametric identification in a number of settings.[2] In addition, they provide a systematic account of the behavioral assumptions

[1]In 2000, these figures are 30,000 and 9%, respectively (see AMU, 2001).

[2]A major advantage of the approach is that it does not require exclusion restrictions on the set of

that are required for a valid use of this approach. Notably, individuals are not allowed to anticipate the moment at which the treatment occurs, although they are allowed to know the distribution of this moment over time. Many of the requirements for the use of this approach also apply to other treatment evaluation methods, including those that do not focus on dynamic treatment assignment or on a duration variable as main outcome. Nevertheless, they are often neglected in the empirical literature, including empirical studies of treatment effects on duration variables. We explain in detail that AMU fits well into the methodological framework, contrary to other labor market training programs and active labor market programs in Sweden. To substantiate our claims we use evidence from discussions with caseworkers, and we also rely on existing studies on unemployment, unemployment insurance, and active labor market programs in Sweden. These include Eriksson (1997a, 1997b), Zettermark et al. (2000), Carling and Richardson (2004), Dahlberg and Forslund (2005), Edin et al. (1998), and Carling et al. (1996). (Some of these deal with the interaction between the inflow into active labor market programs in general on the one hand, and expiration of benefits entitlement on the other; we return to this in Sections 2 and 3.) Our paper thus contributes to the evaluation literature by explicitly studying the empirical implementation of the Timing of Events approach at a very high level of detail.

A major practical advantage of the Timing of Events approach is that it does not just lead to a single estimated treatment effect, but instead it allows for estimation of how the causal training effect changes over time. In particular, we allow the effect of AMU on the exit rate to work to depend on the elapsed time in unemployment since exiting the course and on the elapsed unemployment duration at which participation took place. (Time-varying) effects on hazard rates can be more easily related to the individual economic behavior than effects on the over-all probability of finding work as a function of the time since entry into unemployment. The estimates can therefore be used to study the reasons for *why* training works or not. The paper thus illustrates the usefulness of the Timing of Events approach in understanding the reasons for the effectiveness of a policy, and this in turn facilitates the assessment of counterfactual policy changes.

Notice that unobserved heterogeneity in the treatment *effect* may be an important explanation for changes of the observed treatment effect over time. The intuition is the same as for the spurious duration dependence generated by unobserved heterogeneity in duration models (e.g. Lancaster, 1990). Treated individuals with unobserved characteristics such that their treatment effect is high are (holding every other characteristic constant)

explanatory variables that directly affect the chances of getting a job. Also, it does not require selection effects to be captured completely by observed variables (like the so-called matching approach). This is particularly useful if the set of observed variables only contains a small number of indicators of past individual labor market behavior, as is often the case. See Van den Berg (2001), for a survey, and Abbring and Van den Berg (2004) for a more detailed comparison to other evaluation methodologies.

more likely to leave unemployment quickly. This tends to decrease the average treatment effect among the treated survivors. Whether the exit rate after treatment declines because of a fading treatment effect or because of dynamic selection has major policy implications. In the former case the policy is only effective for a short while, whereas in the latter case one might want to screen individuals more closely before admission into training. We develop a model in which the treatment effect depends on the time since treatment, on covariates, and on an unobserved heterogeneity term which may be related to the unobserved heterogeneity terms affecting the treatment assignment rate and the transition rate out of the current state. This model, which could be labelled a Mixed Proportional Treatment Effect model, was not considered by Abbring and Van den Berg (2003). We demonstrate identification of this model under conditions similar to those in Abbring and Van den Berg (2003).

Duration model estimates with treatment effects are less sensitive to model assumptions if multiple spell data are available. Since our data set includes many individuals with multiple unemployment spells, we may exploit this advantage. We also estimate models that deal with participation in non-AMU programs, and we estimate models that take account of the real time spent in training. The latter mitigates any positive effect of training, in the sense that time in training by itself (the so-called lock-in effect) increases the mean unemployment duration.

To date, a few econometric studies have addressed the effect of AMU on unemployment duration. Harkman and Johansson (1999) and some replication studies examine individuals who finish a program in the final quarter of 1996. Harkman and Johansson (1999) use a subset of the data that we use and match it to data from a postal survey conducted in late 1997. They estimate a bivariate probit model on the employment probability at one year after the program, for different programs. The instrumental variable in the participation equation is the composition of programs within the employment office. The validity of the corresponding exclusion restriction is debatable. Their results indicate that persons in AMU have a higher probability to get a job. Subjective responses on the perceived importance of program participation agree to the estimation results.[3]

[3] Edin and Holmlund (1991) and Larsson (2003) examine the effect of AMU on the transition rate from unemployment to work for *young* individuals aged below 25. Edin and Holmlund (1991) use data from Stockholm from the early 1980s. They compare the unemployment spells of individuals who become unemployed and do not enter the program with the unemployment spells after exiting an AMU-program, and they attempt to deal with selective assignment by adding many variables on the individual's unemployment history. They find a positive effect. Larsson (2003) also uses a matching approach, with data from the 1990s. Her results are mixed. We do not examine these studies further because in our empirical analyses we restrict attention to individuals aged over 25 (see Subsection 3.4). See Björklund (1993) for a survey of other studies based on data from the 1970s and 1980s. Regnér (2002) studies earnings effects of AMU with register data from around the 1980s. A matching approach is used to construct a comparison group. He concludes that on average there is no effect of AMU on earnings.

The paper is organized as follows. Section 2 describes the AMU program. In Section 3 we discuss the model framework and we highlight the main assumptions. We then argue that AMU fits into the framework whereas other programs do not. Section 4 describes the data. Section 5 contains the main estimation results. We also report the sensitivity of the results with respect to a number of assumptions concerning the model, and the construction of duration variables. Section 6 concludes.

2 Labor market training in Sweden

2.1 The AMU program

The purpose of the AMU program is to improve the chances of job seekers to obtain a job, and to make it easier for employers to find workers with suitable skills. This means that it aims to increase unemployed individuals' transition rate to work. The program attempts to achieve this by way of the participation of individuals in training and education courses.[4]

The program is targeted at unemployed individuals as well as employed individuals who are at risk of becoming unemployed. The individuals have to be registered at the local job center (which we shall call the (local) employment office) and must be actively searching for a job. The lower age limit is 20, although nowadays younger individuals are entitled to participate if they are disabled or receive unemployment insurance (UI) benefits.

During the 1980s, the yearly average number of individuals in AMU per month was about 40,000. During the heavy Swedish recession of the early 1990s, this number increased up to 85,000, with seasonal peaks of about 100,000. After 1992, this number decreased again to about 30,000–40,000, which is about 1% of the total labor force (Dahlberg and Forslund, 2005; AMU, 2001). Nowadays, the annual inflow into AMU is about 80,000. The average duration of a course has fluctuated during the past decade and is now about six to seven months. In 1994, total expenditure on the AMU program amounted to about SEK 12 billion (US $ 1.2 billion), half of which was for training procurement and half for training grants. Per participant this equals about $ 10,000 for procurement and $ 10,000 for grants, on a yearly base (AMU, 1997).

There is strong evidence that in 1991 and 1992, participation in AMU was often used in order to extend benefits entitlement (Regnér, 2002, and Edin et al. 1998). This requires a brief exposition. A commonly recognized problem with Swedish labor market programs is that until 2001 they could be used to extend an individual's entitlement to unemployment benefits (which is 300 working days ($\approx$ 14 months) for those aged between 25 and

[4]See e.g. AMS (1997). The formulation of the official aims of AMU has changed somewhat over time. For example, earlier formulations sometimes even refer to the prevention of cyclical inflationary wage increases. See e.g. Harkman and Johansson (1999) and Regnér (1997).

55). By participating in a program, the unemployed individual ensured that his benefits entitlement was extended until completion of the program; in fact, if the participation exceeded a few months then the new entitlement extended further into the future. Edin et al. (1998) examine this interaction between inflow into active labor market programs in general on the one hand, and expiration of benefits entitlement on the other. They do not consider differences across programs. They find that many unemployed workers move into programs shortly before expiration. Carling et al. (1996) use data from 1991–1992 to study these issues as well, and they reach similar conclusions.[5] In January 1993, a new large program called ALU ("work experience") was introduced to end the abuse of AMU for benefits entitlement extension. ALU is specifically targeted towards individuals whose benefits entitlement expires. Participation usually amounts to performing tasks in the non-profit private that would otherwise not be carried out. Also, in 1993, the size of other non-AMU programs increased, and other new programs were designed. Again, these programs are much cheaper than AMU.

There are two types of AMU training: vocational and non-vocational. Vocational training courses are provided by education companies, universities, and municipal consultancy operations. The local employment office or the county employment board pay these organizations for the provision of courses. The contents of the courses should be directed towards the upgrading of skills or the acquisition of skills that are in short supply or that are expected to be in short supply. In recent years, most courses concerned computer skills, technical skills, manufacturing skills, and skills in services and medical health care. Vocational training is *not* supposed to involve the mastering of a wholly different occupation with a large set of new skills.

Non-vocational training (basic general training) concerns participation in courses within the regular educational system, i.e. at adult education centers and universities. Non-vocational training specifically intends to prepare the individual for other types of training (so that the aim of an increased transition rate to work is less direct here). Before 1997, a substantial part of AMU consisted of this non-vocational training. In 1997, a new program of adult education (called the Adult Education Initiative, or Knowledge Lift) has been introduced, and this program is, amongst other things, supposed to replace the non-vocational training part of AMU (see Brännäs, 2000). Nevertheless, for the period since January 1995, non-vocational training amounts to approximately 40% of all AMU courses followed. For 2000 this number is even higher (about 50%).

Concerning UI it should be mentioned that entitlement also requires registration at

[5]Note that this also suggests that workers do not enjoy training very much, since otherwise they would have entered these programs earlier. Alternatively, caseworkers may stimulate unemployed individuals to enter programs only shortly before the benefits expiration, or program participation was quantity constrained for individuals with low unemployment durations.

the employment office. In the mid-1990s, about 40% of the inflow into unemployment and about 65% of the stock of unemployed qualified for UI (Carling, Holmlund and Vejsiu, 2001). Part of the remaining 60% received "cash assistance" benefits, which are typically much lower than UI benefits. The average replacement rate for UI recipients is about 75% (Carling, Holmlund and Vejsiu, 2001).

During the training, the participants' income is called a training grant. Those who are entitled to UI receive a grant equal to their UI benefits level, with a minimum of SEK 240 per day (which is about $24). The other participants receive a grant of SEK 143 per day. These payments are made by the UI agency. In case of vocational training, the training organizations have to send in attendance reports, and the grant is withheld in case of non-attendance. In all cases, training is free of charge. In fact, additional benefits are available to cover costs of literature, technical equipment, travel, and hotel accommodation. In this sense, AMU training is far more attractive than regular education.

In Sweden there is a number of other active labor market programs (that is, apart from AMU and the above-mentioned ALU). Most of these concern subsidized employment. See AMS (1998) and Harkman and Johansson (1999) for descriptions of the programs and changes in program participation over time, respectively. In 1997, on average 191,000 individuals (4.5% of the total labor force) participated in one of the programs. The government's part of the total costs of this have amounted to over 3% of GDP (Dahlberg and Forslund, 2005, Regnér, 2002). In fact, Sweden has been the country with the highest percentage of GDP spending on active labor market policies in the world.

The benefits entitlement rules and programs for persons aged below 25 or over 55 differ from those aged between 25 and 55. Young persons must participate in a program after 100 days of unemployment, or otherwise they lose their unemployment benefits. They may use special programs that are not available for other age groups. Persons over 55 receive unemployment benefits for 450 days (instead of 300 days for those aged between 25 and 55).

Dahlberg and Forslund (2005) examine crowding out of non-participants by active labor market programs. They find no significant crowding out effects of AMU.

2.2 The training enrollment process at the individual level

In this subsection we describe the process that leads to an individual's enrollment in AMU. The information is mostly obtained from documents of the Swedish National Labour Market Board (AMS) (see e.g. AMS, 1998) and from in-depth interviews with a number of individual caseworkers.[6] In addition, we rely on Zettermark et al. (2000), who provide a

[6]We did not use a formal sampling procedure to select caseworkers to be interviewed. Rather, we contacted a number of them to get detailed information concerning the actual decision process at the work

wealth of information on the day-to-day activities of employment offices and caseworkers. Most of that information confirms the interview outcomes.

Usually the employment office advertises, at the office and in the newspapers, the availability of AMU courses. Most of the offices advertise one or two months before the scheduled starting date. In the advertisement they invite interested individuals to an information meeting. At this meeting individuals are informed about the contents of the course and about the eligibility rules. The individuals can usually talk to their personal caseworker at the meeting. Those who are interested can then apply to the course.

Enrollment requires approval from the caseworker. The eligibility rules usually include minimum requirements on the educational level upon inflow, but these are typically not binding. The caseworker also estimates the individual's "need" for AMU. In practice this means that he examines whether the individual's skills can be enhanced by the course. It is common that the applicants undergo a test in order to find out if they are able to benefit from the course. One may for example test the person's skills in mathematics or in the Swedish language. The test may also include some ability testing. Another way to address whether the individual's skills can be enhanced is by profiling the individual in terms of employment opportunities, i.e. making an educated guess about the individual's "typical" unemployment duration. This duration is regarded to be high in case of a low education or an obsolete type of education, or if the individual has an occupation in excess supply. The profiling procedure is subjective. Sometimes the applicant should write a personal letter that explains why he wishes to participate in a specific AMU-course. If the person has work experience in his occupation, the caseworker might call employer references to ask if they would consider employing the person after AMU participation. In general, caseworkers seem to be reluctant to offer AMU courses in fields that are completely different from the occupation of the individual. If an individual rejects a caseworker's offer of an AMU course then in principle the individual's unemployment benefits may be cut off completely, but such sanctions were extremely rare in practice.

The assignment may be affected by caseworkers working closely with firms that demand certain skill categories. Such firms may have an influence on who is accepted into the program. In such cases, training (of the unemployed individual) and job search effort (done by his caseworker) go hand in hand, so the effect of AMU may consist of a skill enhancing effect as well as a search effort effect.

If the number of applicants is insufficient then the course may be cancelled (i.e. may not be bought from the course provider). If there are more applicants than slots in a given course, then individuals with high elapsed durations and/or at risk of losing benefits (these are usually the same individuals) are often given priority. However, AMU is generally not offered to individuals if they are primarily concerned about the renewal of their unemploy-

floor of the employment offices.

ment benefits. It is commonly felt that such practices would not agree with the objective of AMU. Perhaps more importantly, there are in general cheaper alternative programs to deal with such cases, like workfare programs, and efforts are made to push the individual into those programs instead of AMU. Similarly, AMU is generally not offered to individuals who, in the opinion of the caseworker, need practical experience in order to be able to get a job, or who are just deemed in “need something to do” during daytime. In such cases the individual is offered another active labor market program, like a work experience program.

It takes approximately one month from the first information meeting to the first day of the course. On average, the period from application to acceptance takes 2–3 weeks, while the period from acceptance to the start of the course takes 1–2 weeks. An individual may try the AMU-course before actually starting the course. For example, if he is interested in welding then he can make a one-week visit to the school that offers welding courses. Also, individuals may drop out of the course, because they find a job or for other reasons. In fact, in the first case, they are encouraged to do so, and they can come back later and complete the course. An AMU participant may also follow a sequence of courses, starting with basic vocational training and ending in a very narrow type of vocational training. Such a sequence may take 30–40 weeks. The participants do not receive grades or test-based certificates upon finishing a course.

We now show that the above information given by caseworkers on the process that leads to an individual’s enrollment in AMU is confirmed by existing empirical studies. Eriksson (1997a, 1997b) analyzes choice and selection into different programs using register data in combination with survey data on choice and selection by the unemployed as well as the caseworkers. (The HÄNDEL register that she uses is part of the set of registers that we use in the current paper.) It is shown that the personal characteristics that are observable in HÄNDEL are not able to give a very precise prediction of actual participation in AMU versus non-participation. The predictive performance can be substantially enhanced if one takes account of self-reported (by the unemployed) measures of the amount with which AMU is expected to have certain advantages for future labor market prospects. These can be assumed to capture unobserved heterogeneity in the inflow rate into AMU and perhaps unobserved heterogeneity in the treatment effect. (Of course they may also reflect an ex-post rationalization of actual choices made in the past.) Eriksson (1997a) notes that informal interviews with caseworkers reveal that the motivation of the unemployed is a very important criterium for placing an unemployed individual into AMU.

Eriksson (1997b) exploits survey data obtained by letting caseworkers give AMU-advice on the basis of actual files of unemployed individuals that are supplied to them by the survey agency. The allocation of files to caseworkers is fully randomized. The data also allow for a comparison between the valuation of AMU as stated by the caseworkers and the actual (non-)participation of the individual. It turns out that heterogeneity of the caseworkers (which is typically unobserved but is here observed and used as an identifier) is a more important determinant of the caseworkers' stated decisions than the unobserved heterogeneity of the unemployed individuals as captured by fixed effects. So, there is a lot of variation in the caseworkers' decisions which can not be attributed to the unemployed individuals' identities but can be attributed to the caseworkers' identities. When selecting on the basis of observable personal characteristics, officials seem to use rules of thumb which are often not in accordance to the stated goals of AMU on priority groups. If the caseworkers think that an individual would benefit a lot from participation then the individual is also more likely to be an actual participant. But the actual participation also depends on the unemployed individual and on unexplained factors.

Carling and Richardson (2004) use the HÄNDEL data from 1995 onwards to study the choice of a particular type of training program conditional on going into one of these programs. They use a Multinomial Logit model for this. They find that employment agency identifiers have significant effects, and that these dominate the effects of characteristics of the unemployed individual.

According to Eriksson (1997b), caseworkers are reluctant to let current participants to non-AMU programs enter AMU. Also, work experience programs and public temporary employment are substitutes for each other but not for AMU. Caseworkers regard AMU to be a fundamentally different kind of program. So the variation in the caseworkers' behavior with respect to AMU mostly concerns the choice between AMU and no AMU, instead of the choice between AMU and another program. According to Dahlberg and Forslund (2005), nowadays, AMU is typically not used for UI entitlement extensions.

3 The model framework

3.1 A class of bivariate duration models for treatment evaluation

We normalize the point of time at which the individual enters unemployment to zero. The durations T_u and T_p measure the duration until employment and the duration until entry into the AMU training program, respectively.[7] At this stage we assume that unemployment can only end in employment, and we take the period in AMU as part of the unemployment spell. Also, for the moment we ignore other training programs during unemployment. As a result, T_u also denotes the duration of unemployment. The population that we consider concerns the inflow into unemployment, and the probability distributions that are defined

[7]Formally, different potential values t_p of T_p denote different treatments. The model framework can accordingly be developed in terms of counterfactual notation; see Abbring and Van den Berg (2003). Here we simply outline the model as a system of two equations: one for the treatment assignment mechanism and one for the actual duration outcome corresponding to the actual assigned treatment t_p.

below are distributions in the inflow into unemployment (unless stated otherwise).

The two durations are random variables. If necessary we use T_u and T_p to denote the random variables and t_u and t_p to denote their realizations, but for expositional reasons we occasionally use the latter notation for both. We assume that, for a given individual in the population, the duration variables are absolutely continuous and nonnegative random variables. We assume that all individual differences in the joint distribution of T_u, T_p can be characterized by explanatory variables X, V, where X is observed and V is unobserved to us. Of course, the joint distribution of $T_u, T_p|X, V$ can be expressed in terms of the distributions of $T_p|X, V$ and $T_u|T_p, X, V$. The latter distributions are in turn characterized by their hazard rates $\theta_p(t|x, V)$ and $\theta_u(t|t_p, x, V)$, respectively.[8]

As noted in the introduction, we are interested in the causal effect of participation in AMU on the exit out of unemployment. The treatment and the exit are characterized by the *moments* at which they occur, so we are interested in the effect of the realization of T_p on the distribution of T_u. To proceed, we assume that, conditional on X, V, the set of possible relations between T_u and T_p is characterized as follows: the realization t_p of T_p affects the shape of the hazard of T_u from t_p onwards, in a deterministic way. The assumption implies that the causal effect is captured by the effect of t_p on $\theta_u(t|t_p, x, V)$ for $t > t_p$. Note that it is ruled out that t_p affects $\theta_u(t|t_p, x, V)$ on $t \in [0, t_p]$. Obviously, it is useful to take the hazard rates as the basic building blocks of the model specification. As will become clear below, this also facilitates the discussion of the empirical relevance of some assumptions, and it enables one to interpret empirical findings in terms of an economic-theoretical framework.

Let $V := (V_u, V_p)'$ be a (2×1)-vector of unobserved covariates. As usual, we take V_p (V_u) to capture the unobserved determinants of T_p (T_u). We adopt the following model framework, in terms of the hazard rates $\theta_u(t|t_p, x, V_u)$ and $\theta_p(t|x, V_p)$ (where it should be stressed that we also estimate less restrictive model specifications),

Model 1.

$$\theta_p(t|x, V_p) \quad = \quad \lambda_p(t) \cdot \exp(x'\beta_p) \cdot V_p \tag{1}$$

$$\theta_u(t|t_p, x, V_u) \quad = \quad \lambda_u(t) \cdot \exp(x'\beta_u) \cdot \exp(\delta(t|t_p, x) \cdot \mathrm{I}(t > t_p)) \cdot V_u \tag{2}$$

where I(.) denotes the indicator function, which is 1 if its argument is true and 0 otherwise.

[8] For a nonnegative random (duration) variable T, the hazard rate is defined as $\theta(t) = \lim_{dt \downarrow 0} \Pr(T \in [t, t + dt)|T \geq t)/dt$. Somewhat loosely, this is the rate at which the spell is completed at t given that it has not been completed before, as a function of t. It provides a full characterization of the distribution of T (see e.g. Lancaster, 1990).

Apart from the term involving $\delta(t|t_p, x)$, the above hazard rates have Mixed Proportional Hazard (MPH) specifications. The term $\delta(t|t_p, x) \cdot \mathrm{I}(t > t_p)$ captures the AMU effect. Clearly, AMU has no effect if and only if $\delta(t|t_p, x) \equiv 0$. Now suppose $\delta(t|t_p, x)$ is a positive constant. If T_p is realized then the level of the individual exit rate to employment increases by a fixed amount. This will reduce the remaining unemployment duration in comparison to the case where AMU is entered at a later point of time.

More in general, we allow the effect of AMU to vary with the moment t_p of entry into AMU and with x. Moreover, the individual effect may also vary over time, as we allow it to depend on the elapsed unemployment duration t. As a result, the individual effect may also vary with the time $t - t_p$ since entry into AMU. The effect of $t - t_p$ may capture that the exit rate is low during the training course or high immediately after completion of it. Model 1 does not rule out that for each individual there is a probability that he will never get training ($\int_0^\infty \lambda_p(t)dt < \infty$) We may also allow x to be time-varying. In an extension we allow the training effect to depend on unobserved characteristics, i.e. to be heterogeneous across individuals with the same x (see Subsection 3.2).

Suppose that we have a random sample of individuals from the inflow into unemployment, containing one unemployment spell per individual (i.e. single-spell data). The data then typically provide observations on T_u and x for each individual. In addition, if T_p is completed before the realization t_u then we also observe the realization t_p, otherwise we merely observe that T_p exceeds t_u.

Consider the (sub)population of individuals with a given value of x. The individuals who are observed to enter AMU at a date t_p are a non-random subset from this population. The most important reason for this is that the distribution of V_p among them does not equal the corresponding population distribution, because most individuals with high values of V_p have already gone into AMU before. If V_p and V_u are dependent, then by implication the distribution of V_u among them does not equal the corresponding population distribution either. A second reason for why the individuals who are observed to enter AMU at a date t_p are a non-random subset is that, in order to *observe* the fact that entry into AMU occurs at t_p, the individual should not have left unemployment before t_p. Because of all this, the AMU effect cannot be inferred from a direct comparison of realized unemployment durations of these individuals to the realized unemployment durations of other individuals. If the individuals who enter AMU at t_p have relatively short unemployment durations then this can be for two reasons: *(1)* the individual causal AMU effect is positive, or *(2)* these individuals have relatively high values of V_u and would have found a job relatively fast anyway. The second relation is a spurious selection effect.

If V_u and V_p are independent (which includes the case in which unobserved heterogeneity V_u in the exit rate to work is absent) then $\mathrm{I}(t > t_p)$ is an exogenous time-varying covariate for T_u, and one may infer the AMU effect from a univariate duration analysis based on the distribution of $T_u|t_p, x, V_u$ mixed over the distribution of V_u. However, in general there is no reason to assume independence of V_u and V_p, and if this possible dependence is ignored then the estimate of the AMU effect may be inconsistent.

The joint density of $T_u, T_p|x$ at $T_u = t_u, T_p = t_p$ can be expressed as

$$\int_0^\infty \int_0^\infty (\exp(x'\beta_u)v_u\lambda_u(t_u)\exp(\delta(t_u|t_p,x)\mathrm{I}(t_u > t_p))$$
$$\exp\left(-\exp(x'\beta_u)v_u\left[\int_0^{\min\{t_u,t_p\}}\lambda_u(s)ds + \mathrm{I}(t_u > t_p)\int_{t_p}^{t_u}\lambda_u(s)\exp(\delta(s|t_p,x))ds\right]\right) \quad (3)$$
$$\exp(x'\beta_p)v_p\lambda_p(t_p)\exp(-\exp(x'\beta_p)v_p\int_0^{t_p}\lambda_p(s)ds\Big)\, dG(v_u, v_p)$$

where G is the joint distribution of V_u, V_p in the inflow into unemployment. This joint density forms the basis for the Maximum Likelihood estimation of the model.[9]

Abbring and Van den Berg (2003) show that Model 1 is identified from single-spell data, i.e., from a random sample of drawings of $\{T_u, \mathrm{I}(T_p \leq T_u), T_p \cdot \mathrm{I}(T_p \leq T_u), x\}$. This means that there is a one-to-one mapping between the data generated by the model and the set of model determinants (being the functions $\lambda_u, \lambda_p, \delta$, the unobserved heterogeneity distribution G, and the parameters β_u and β_p). This is a useful model property. It implies that the estimation results are not fully determined by parametric functional form assumptions on the functions $\lambda_u, \lambda_p, \delta$ and G.

Intuitively, what drives the identification of the training effect δ is the extent to which the moments of training and the moment of exit to work are close in time. If training is quickly followed by exit to work, no matter how long the elapsed unemployment duration before the training, then this is evidence of a causal effect of training. The spurious selection effect gives a second relation between the two duration variables, but it can be shown that that relation does not give rise to the same type of quick succession of events. So the interaction between the moment of exit and the moment of training in the conditional rate of events allows one to distinguish between the causal effect and selectivity. With specifications where δ depends on t and t_p, the identification follows from a comparison of treated and not-yet treated at points of time t and t_p, using observations of $\min\{T_u, T_p\}|x$ to correct for selectivity (see Abbring and Van den Berg, 2004).

Identification does not require exclusion restrictions on the hazard specification of either duration, so the same vector x may affect both hazards. This entails that we allow individuals to be aware of the existence of the AMU, and we allow them to influence both

[9]Note that Model 1 and (3) include a specification of the distribution of T_p for $T_p > T_u$. However, this specification is immaterial, as it does not play any role in the paper or indeed in any empirical analysis.

the rate of entry into AMU and the rate of exit into employment. This is obviously an advantage. We return to this below.

So far we have ignored time-varying covariates, although t_p can be thought of as an endogenous time-varying covariate in θ_u. It is clear that in some cases a model with time-varying covariates is not identified, for example, if $\theta_i(t|x, v_i) = \lambda_i(t) \exp(x(t)'\beta_i)$ with $x(t)$ additive in t. However, in general, variation of x over time is helpful for identification of duration models. Honoré (1991) and Heckman and Taber (1994) provide some illustrations of this. In our empirical model specifications we include exogenous x variables that vary over time.

The identification with single-spell data does require a number of assumptions that are standard in the literature on identification of MPH models. Notably, $X \perp\!\!\!\perp V$, and X includes two continuous variables with the properties that (i) their joint support contains a non-empty open set in $\mathbb{R}^2$, and (ii) the vectors of the corresponding elements of β_u and β_p form a matrix of full rank. Abbring and Van den Berg (2003) show that these assumptions can be discarded if the data provide multiple spells, i.e. if for individuals in the sample we have more than one unemployment spell with the same value of V, and if these spells are independent given the values of x and V. We assume that an individual has a given value of V_u, V_p. Since V_u and V_p are unobserved, the duration variables given x are not independent across spells. It is especially useful that identification with multi-spell data does not require independence of observed and unobserved explanatory variables, as in general such independence is hard to justify. In fact, multi-spell data also allow the relaxation of multiplicity assumptions in Model 1. Specifically, we may allow x to enter in an arbitrary nonproportional manner in the conditional hazard rates, and we do not need variation of these hazard rates with x at all. Alternatively, we may allow the dependence of the conditional hazard rates on t, x in the second spell to be different from the dependence of these rates on t, x in the first spell. The size of the AMU effect may also be different across the two spells. A causal effect of the realizations for the first spell on the outcomes for the second spell or the other way round is not allowed (although the observed outcomes are related across spells by way of their unobserved determinants). But the individual values of x may differ across spells.

3.2 Identification of models with duration dependence and unobserved heterogeneity in the treatment effect

In the model of the previous subsection, the magnitude of the causal training effect δ does not depend on unobserved characteristics, so any systematic heterogeneity of treatment effects across individuals comes from observable characteristics x. It is hard to justify this assumption. Moreover, unobserved heterogeneity in δ may be an important explanation for changes of the observed (i.e., only conditional on x) treatment effect over time. The intuition is the same as for the spurious duration dependence generated by unobserved heterogeneity in duration models (e.g. Lancaster, 1990). Treated individuals with unobserved characteristics such that their treatment effect is high are (holding every other characteristic constant) more likely to leave unemployment quickly.[10] This tends to decrease the average treatment effect among the treated survivors. Of course, if the unobserved characteristics affecting the treatment effect are inversely related to the unobserved characteristics V_u affecting the exit rate to work in general, then more subtle effects can be generated for the observed treatment effect.

As we shall see in Section 4, the decline of the observed exit rate to work among the treated is a major distinguishing feature of the raw data. It therefore makes sense to consider models that allow for both duration dependence of the individual treatment effect and spurious duration dependence due to dynamic selection as two potential explanations for the observed decline. Moreover, whether the exit rate after treatment declines because of a fading treatment effect or because of dynamic selection has major policy implications. In the former case the policy is only effective for a short while, whereas in the latter case one might want to screen individuals more closely before admission into training.

Abbring and Van den Berg (2003) demonstrate identification of a model in which δ is a sum of a term depending on t, a term depending on x, and an unobserved heterogeneity term V_δ. This function δ does not depend on t_p. For our purposes, such a model is less attractive. Instead, we consider a model in which δ is allowed to depend on $t - t_p, x$, and V_δ. Specifically, in Model 1 we replace δ by

$$\delta(t - t_p, x, V_\delta) \;=\; \lambda_\delta(t - t_p) + x'\beta_\delta + V_\delta \tag{4}$$

where V_δ is allowed to be stochastically related to V_u and/or V_p. Note that the exit rate to work (or, more generally, the transition rate out of the state of interest) is proportional to $\exp(\delta)$, so that by analogy to the Mixed Proportional Hazard model we may call our model the Mixed Proportional Treatment Effect model.

In the Appendix we present the model assumptions in detail and we prove identification of this model under conditions similar to those in Abbring and Van den Berg (2003) and in the previous subsection. To be short,

Proposition 1. *The Mixed Proportional Treatment Effect model is identified.*

[10]The heterogeneity may also be due to heterogeneity of characteristics of the training course. The individuals who follow a good course find a job quickly, and those who follow a bad course remain unemployed longer.

3.3 Implicit assumptions in the model specifications

The model specifications reflect a number of implicit assumptions. First of all, the future realization of the moment t_p of entry into training does not affect the individual's exit rate θ_u prior to that moment t_p. So the individual's exit rate at t is the same irrespective of whether training will occur at $t+1$ or whether it will occur at $t+100$. This rules out anticipation of the future individual realization of the moment of training. If an individual would foresee participation in AMU at a particular future date t_p then he may use this as an input of his current behavior, for example he may want to wait for the treatment by reducing his search intensity for jobs, and this may decrease the probability that T_u is quickly realized. If this is ignored in the empirical analysis then the training effect may be over-estimated. However, if the time span between the earliest moment at which anticipation can occur and the moment of the actual training is short relative to typical values of the durations T_p and $T_u - T_p$, and if the anticipatory effect is not very large, then estimation results may be relatively insensitive to the assumption of no anticipation.

It is important to distinguish anticipation of the realization of T_p from ex ante knowledge of the existence of the program and ex ante knowledge of the individual distribution of T_p. With well-established programs like AMU, it is plausible that *determinants* of the stochastic process of training assignment affect the individual's exit rate out of unemployment before the actual entry into training. For example, if the individual knows that he has a relatively high training enrollment rate and if he enjoys training then he will reduce his job search effort. In such cases the program is said to have an *ex ante* effect on exit out of unemployment before training. The "ex ante" effect contrasts to the *ex post* effect of training, which is the effect of actual training on the individual exit rate. The ex ante effect is an example of the macro effects that are present in a world in which a particular program is implemented. There may also be ex ante or macro effects on the magnitude and composition on the inflow into unemployment and on the behavior of employers.

The model framework is compatible with ex ante effects. However, we do not aim to disentangle such effects from other determinants of the hazard rates. Identification of the ex ante effect on the exit rate to work before training requires additional information, such as strong functional-form assumptions, instruments for a comparison of a world with AMU to a world without it, or the imposition of an economic-theoretic structure on the model (see Abbring and Van den Berg, 2005). The first option is undesirable, whereas the others are beyond the scope of this paper. This means that the treatment effect δ is defined relative to the exit rate to work in absence of a treatment but within a world in which treatments are present.

We now turn to a different type of anticipation. The model framework rules out that the future realization of the variable of interest T_u has an effect on the current level of θ_p. In reality, an individual may have private knowledge on a future job opportunity that is independent of whether the training will occur, and the individual may use this knowledge to avoid training. If something like does occur in reality then a positive effect of training on exit to employment is under-estimated. However, if the training course takes a long time, then this bias may be empirically unimportant, as employers may be unwilling to wait for a new employee for many months. Also, if the time span between the moment at which the anticipation occurs and the moment of the actual exit to work is relatively short, and if the anticipatory effect is not very large, then estimation results may be rather insensitive to this. Again, absence of anticipation does not rule out that individuals know the determinants of the process leading to employment and use these as inputs in their decision problem. For example, the individuals may know that $\lambda_u(t)$ increases in the near future, and modify their strategy accordingly, which may affect their θ_p. The latter can be captured in the model through $\lambda_p(t)$.

Finally, the fact that we specify the assignment of training by way of specifying the hazard rate of a duration distribution implies that there is a random component in the assignment that is independent of all other variables (see e.g. Ridder, 1990, and Abbring and Van den Berg, 2003). The model framework thus postulates that there is variation in T_p at the individual level. (This variation affects T_u only by way of the treatment.) To see the importance of this, consider the extreme case where individuals can only enter AMU at, say, exactly one year after flowing into unemployment. Then it is impossible to distinguish the effect of AMU from the duration dependence in the exit rate to work after one year. (In such a case it is of course also hard to justify that entry into AMU is not anticipated.)

3.4 Applicability of the model framework to AMU

In this subsection we argue that the model framework (covering the different specifications we consider) is particularly well suited for our study of the AMU program. We focus on the following issues: dependent unobserved heterogeneity, randomness in the moment of treatment assignment, absence of anticipatory effects, and absence of substitution with other programs.

From the information in Subsection 2.2 and from the studies by Eriksson (1997a, 1997b), it is obvious that unobserved (to us) heterogeneity of the unemployed individuals plays an important role in the assignment to AMU. The corresponding variables taken into account by the caseworker (like motivation, subjectively assessed expected unemployment duration, and subjective assessments of other aspects of the future career) are also indicative of unobserved determinants of the individual exit rate to work. The empirical analysis should therefore take account of potentially related unobserved heterogeneity terms in θ_u and θ_p.

If the individual knows that a variable is an important determinant of the treatment assignment process (like the amount and type of discretionary behavior of his caseworker), and the individual knows that he may be subject to treatment, then he has a strong incentive to inquire the actual value of the variable. Subsequently, he will take his value of the variable into account to determine his optimal strategy, and this strategy in turn affects the rate at which he moves to employment. We should note that the variables that are observed by us and that may have an effect on assignment to AMU are also observable to the individuals under consideration, so that we cannot impose exclusion restrictions on β_u, and we take the same vector x to affect both θ_u and θ_p.

Now let us consider the presence of randomness in the moment of entry into AMU. To some extent this may be generated by changes in the behavior of the caseworker or the employment agency that are beyond observation of the unemployed individual. More importantly, it is generated by the variation in the moment at which AMU courses start. In addition, admission to a course may depend on the extent to which other individuals apply to the course, which is random from the individual's point of view. Recall that Eriksson (1997b) finds residual variation in the AMU assignment process that can not be attributed to the individual or the caseworker.

We now turn to anticipation of the moment of entry into AMU. From Subsection 2.2, the time period between the moment at which the individual is informed about the possibility of enrolling into an AMU course and the moment at which the course starts is very short. There are however two reasons for why some individuals may anticipate the moment of entry, and both of these lead us to restrict the focus of the empirical analysis somewhat.

First, as discussed in Section 2, in 1991 and 1992 AMU was often used to extend benefits entitlement. In that case, the date of inflow into AMU is mostly determined by the date of expiration of benefits entitlement. The latter date is known in advance by the unemployed individual and his caseworker (this date does not vary much across the unemployed; see the references). This allows for anticipation of the inflow into AMU, which violates a key assumption of our evaluation approach. Moreover, such self-selection into AMU is governed by different motives than self-selection in other years, so we may expect the unobserved heterogeneity distribution to be different across time. From January 1993 onwards, other programs took over its role as means to extend benefits entitlement. We therefore restrict attention to data from 1993 onwards.

Secondly, recall from Section 2 that part of AMU concerns non-vocational training, in particular before 1997. Non-vocational training is often given within the regular school system. This implies that the starting date of the non-vocational training is often determined by institutional features of the school system, like the starting dates of the school seasons. As a result, it is straightforward for unemployed individuals to anticipate the date of inflow into such a program. We therefore restrict ourselves to vocational training. There are two additional reasons to do so. First, vocational training is relatively expensive, so the participation costs are higher. Secondly, vocational training is difficult to obtain in alternative labor market programs, whereas non-vocational training is easier to obtain elsewhere, implying that in the latter case there are substitution possibilities.

Concerning substitution possibilities in general, recall from Subsection 2.2 that case-workers regard vocational AMU training as a very different type of program than the other active labor market programs. The latter are regarded to be substitutable to a high degree. For persons under 25, there are programs that are more similar to AMU vocational training. Also, for these individuals, the similarity with vocational courses and tracks in the regular school system may be important. For this reason we restrict attention to individuals aged over 25. Also, young individuals must enter a training course after 100 days of unemployment, which may generate anticipatory effects. We omit individuals over 55 because they face a different unemployment benefits system and because for them vocational AMU training seems to have relatively small advantages.

It follows from the above that our model framework may be less suited for the analysis of the effects of the other active labor market programs on unemployment duration. With other programs, individuals may anticipate their enrollment a long time in advance, because of their link to benefits entitlement expiration and/or because of their connection to the regular school system. Moreover, it is difficult to analyze them in isolation from each other because of the high degree of substitutability.

4 The data

4.1 Data registers and unemployment spells

The data are taken from a combination of two Swedish register data sets called HÄNDEL (from the official employment offices) and AKSTAT (from the unemployment insurance fund). HÄNDEL covers all registered unemployed persons since August 1991 (approximately 2 million observations). According to Carling, Holmlund and Vejsiu (2001), more than 90are ILO-unemployed according to labor force surveys also register at the employment offices. HÄNDEL includes detailed information on the individuals' training activities and work experience activities, including the starting and ending dates of program participation. HÄNDEL is also informative on whether an individual in AMU receives vocational training or non-vocational training. AKSTAT is available from 1994 onwards and provides information on the wage level and working hours in the job prior to the spell of unemployment, for individuals who are eligible for UI.

Our observation window runs from January 1, 1993 until June 22, 2000. The unit of observation is an individual. For each individual who is in HÄNDEL at least once during the observation window, we can construct an event history from HÄNDEL. For any spell of unemployment (to be defined below), HÄNDEL and AKSTAT provide characteristics at the beginning of the spell, and a list of dates within the spell at which changes occur, including the nature of the change. We also include the information on participation in non-AMU programs, since such participation may temporarily rule out a transition to AMU, or may at least reduce the transition rate to AMU and/or work.

We only use information on individuals who become unemployed at least once within the observation window. An individual becomes unemployed at the first date at which he registers at the employment office as being "openly" unemployed. This eliminates registration spells that start because the individual wants to change employer and also eliminates spells that start because the individual knows that he is going to be unemployed in the future (short term contract or notification of lay-off), at least until the individual does actually become unemployed. We also ignore unemployment spells that are already in progress at the beginning of the observation window, because using them would force us to make assumptions about the period before the beginning of the window. We thus obtain a so-called inflow sample of unemployment spells, and we follow the individuals over time after this moment of inflow. (Note that we also use information available on the period prior to such spells, notably on wages.) We exclude individuals who have experienced unemployment between August 1991 and January 1, 1993. The years 1990–1992 witness an unusually severe recession in Sweden, and the individuals who became unemployed in that period may be different from those who did not become unemployed then but who became unemployed later. As we have seen, the former individuals were certainly exposed to a different active labor market policy regime before 1993, and this might affect their outcomes after 1993 as well.

For convenience, we use the term "unemployment spell" to include possible spells in AMU, relief work, ALU, etc. The spell ends if the individual leaves the employment office register or if he moves from the unemployment categories in the employment office register to a non-unemployment category in the register. If the exit destination is employment then we observe a realization of the duration variable of interest. If the exit destination is different (e.g. "regular education", or "other reason") then this duration variable is right-censored (independence of right-censoring may be checked in a sensitivity analysis). The duration is independent right-censored if the spell is continuing at the end of the observation window.

Occasionally, we observe coding errors in data at points of time at which individuals move between different categories in the register. Obvious typing errors are corrected, whereas otherwise we right-censor the duration variables at the moment at which such an error occurs. If exit occurs into "wage subsidy" or "(public) sheltered employment" then we remove the individual from the sample, since these programs are for handicapped people (who are typically are not in open unemployment anyway). As mentioned in Subsection 3.4, we restrict attention to individuals who were at least 25 and below 55 at the moment they enter unemployment. As a result, our data set contains 500,960 individuals. Note that by following the individuals over time we may observe multiple unemployment spells per individual. For each individual we use at most 3 unemployment spells. The analyses are based on a random subsample of the full data set at our disposal, containing 16467 individuals, with in total 28451 unemployment spells.

Even though vocational AMU and other programs are fundamentally different and are not used as substitutes, we are forced to consider the participation in other programs during unemployment, as such participation spells are likely to affect the transition rates into AMU and into work for a certain amount of time. Since participation in those other programs takes place at points in time that are dispersed across individuals and that may to some extent be random, the common deterministic duration dependence functions in Model 1 cannot capture this. Also, we have seen that expanding the Timing of Events model framework to include multiple types of treatment is hard to justify. If we treat participation in other programs before participation in AMU as regular unemployment, then the transition rate from unemployment into AMU is extremely low during the participation in the other programs. Participation in non-AMU programs most likely also reduces the transition rate into employment. So, during such a period of program participation, it may be preferable to halt the time clock of the duration until regular employment. As a starting point, the time spent in training (in non-AMU programs as well as in AMU) is therefore assumed not to contribute to the unemployment duration, and the time spent in other training programs is assumed not to contribute to the duration until AMU. Note that this also means that time spent in non-AMU programs after AMU does not contribute to the unemployment duration. We subsequently relax these assumptions in additional analyses.

4.2 Descriptive statistics

Table 1 provides summary statistics of the unemployment duration, the participation in labor market programs, and their interrelation. Of all 28451 spells, 2185 (i.e. 7.7%) are observed to include a period of participation in a vocational AMU course. Some of the other spells are right-censored due to the finiteness of the observation window, and in reality some of those may include AMU participation afterwards. The median value of the duration until training across the 2185 spells that are observed to include training is 153 days. Except where stated otherwise, the duration outcomes in Table 1 are measured while ignoring time in training and other programs.

In a setting where one duration outcome of interest (T_p) is right-censored by the other (T_u) and both durations are subject to end-of-follow-up right-censoring, the information in summary statistics of outcomes is limited. Spells with observed AMU participation are longer than spells without simply because it takes time before T_p is realized. To complicate matters further, note that right-censoring due to finiteness of the observation window takes place at the duration value equal to the difference between June 22, 2000, and the moment of inflow into unemployment, and the latter is dispersed across spells. One notable aspect is that a sizeable fraction of spells with AMU participation ends with a transition to work within a few days after leaving training.

Of the 2185 spells that are observed to include AMU participation, 47% are also observed to include participation in another type of active labor market program. Not surprisingly, this happens predominantly in long spells. Of the spells with t_p smaller than 160 days, only 12% are also observed to include participation in another type of active labor market program before AMU participation. Of the spells observed to be shorter than 160 days that are not observed to include participation in AMU, 10% are observed to include participation in another type of active labor market program. This suggests that participation in other programs is not related to AMU participation. The fact that spells with AMU participation relatively often also include participation in other programs is because of the fact that by conditioning on AMU participation we condition on high realized durations.

Table 2 provides summary statistics of explanatory variables in the empirical analysis, across all spells and across all first spells. The latter reflects the composition across individuals better than the former, as we allow the x variables to differ across the spells of a given individual.

Concerning education we distinguish between five levels: junior high school or lower, short senior high school, long senior high school, short tertiary education, and full university degree or higher. These are roughly equivalent to ≤ 9, 10–11, 12–13, 14, and ≥ 15 years of education, respectively. Concerning nationality we distinguish between three categories: Eastern Europe, Africa / Asia, and otherwise (including Sweden). Concerning the type of unemployment benefits received during unemployment we distinguish between three categories: UI, cash allowance, and neither. For UI recipients in 1994 and beyond, the AKSTAT data include the hourly wage earned in the job that was held just before the onset of the spell of unemployment. This is almost linearly related to their UI level (see e.g. Carling, Holmlund and Vejsiu, 2001). For non-UI-recipients the wage variable is set to zero. The latter also applies to UI recipients who become unemployed and subsequently employed within 1993. However, if they move back to unemployment in 1994 we use the corresponding pre-unemployment wage to quantify the pre-unemployment wage for the unemployment spell in 1993. The "large city" dummy equals 1 iff the individual lives in one of the counties covering Stockholm, Göteborg, and Malmö. Notice that some variables concern subjective assessments by the caseworker (e.g. whether the individual needs guidance) or subjective statements by the individual concerning the span of jobs that he searches for.

The sample means across spells are virtually equal to those across individuals in their first spell. This suggests that the observation of multiple spells is not strongly driven by

Table 1: *Summary statistics for the treatment and the outcome.*

	All spells	**First spell**
regardless of treatment		
# spells	28 451	16 467
# individuals	16 467	16 467
% with exactly one spell		53
% with exactly two spells		21
% with ≥ 3 spells		26
% spells with t_p observed	7.7	8.1
% spells with t_u observed	58.0	57.1
average observed t_u	149 (181)	162 (199)
median observed t_u	89	95
% spells with time in other programs	20.7	20.0
average time spent in other programs	43 (117)	45 (125)
id. for spells with observed $t_u <$ its median	14 (53)	12 (49)
concerning spells with observed t_p		
# spells	2185	1339
% spells with t_u observed	56.5	53.1
average observed t_p	211 (205)	240 (219)
median observed t_p	153	187
average observed t_u	328 (285)	379 (310)
median observed t_u	246	294
average observed $t_u - t_p$	132 (188)	152 (210)
id. incl. censored t_u	162 (229)	185 (254)
average observed $(t_u - t_p)/t_p$	1.9 (8.1)	1.7 (7.1)
id. incl. censored t_u	2.3 (15.1)	2.3 (17.9)
average observed $(t_u - t_p)/t_u$	0.3 (0.3)	0.3 (0.3)
id. incl. censored t_u	0.4 (0.3)	0.4 (0.3)
% spells with $t_u \approx t_p$	24.8	21.9
id. incl. censored t_u	19.4	16.7
average time in training	119 (114)	120 (114)
% spells with time in other programs	47.4	49.3
average time spent in other programs	114 (186)	128 (200)
id. for spells with observed $t_u <$ its median	46 (104)	58 (108)

Explanatory note: Standard deviations in parentheses. The time unit is one day. The condition $t_u \approx t_p$ is shorthand for $t_p \leq t_u \leq t_p + 5$.

Table 2: *Averages of explanatory variables.*

	Across all spells	**Across first spells**
age	35.9 (8.3)	35.4 (8.7)
level of education:		
junior high school or lower	0.31	0.33
short senior high school	0.26	0.25
senior high school	0.20	0.21
short tertiary education	0.06	0.05
university	0.17	0.16
female	0.51	0.50
unemployment benefits:		
UI recipient	0.66	0.63
cash allowance recipient	0.07	0.08
nationality:		
from Eastern Europe	0.05	0.05
from Africa/Asia/S.America	0.05	0.05
hourly wage if observed	87.6 (33.0)	86.8 (32.9)
experience in occupation (dummy)	0.63	0.62
education in occupation (dummy)	0.63	0.63
occupation:		
manufacturing	0.22	0.20
professional, technical, agric.	0.23	0.23
health, nursing and social care	0.14	0.14
adm., managerial, sales, clerical,service	0.41	0.43
large city (dummy)	0.52	0.53
needs guidance (dummy)	0.08	0.09
willing to move (dummy)	0.16	0.16
accepts part-time work (dummy)	0.06	0.06
local unemployment rate	0.09 (0.03)	0.09 (0.03)

Explanatory note: Standard deviations are in parentheses.

selectivity. Age is on average slightly higher across spells than across individuals in their first spell, but this is a consequence of the fact that an individual's age necessarily increases over consecutive spells.

5 The empirical analysis

5.1 Parameters

For the duration dependence functions and the bivariate unobserved heterogeneity distribution we take flexible specifications. We take both $\lambda_u(t)$ and $\lambda_p(t)$ to have a piecewise constant specification,

$$\lambda_i(t) = \exp\left(\sum_{j=1,2,\ldots} \lambda_{ij} I_j(t)\right) \qquad i = u, p$$

where j denotes time intervals and $I_j(t)$ are time-varying dummy variables that are one in consecutive time intervals. Note that with a sufficiently large number of time intervals any duration dependence pattern can be approximated closely.

In most of the empirical analyses we take 8 intervals for λ_u and 6 for λ_p. In both cases the length of an interval is 56 days, except for the last intervals which are unbounded from the right.

We take the joint distribution of the unobserved heterogeneity terms V_u and V_p to be bivariate discrete with two unrestricted mass point locations for each term. This specification is popular, flexible, and computationally feasible (see Van den Berg, 2001, for an overview). Let v_1, v_2, v_3 and v_4 denote the points of support of V_u and V_p, respectively (note that V_u and V_p are random variables whereas $v_1, .., v_4$ are realizations). The associated probabilities are denoted as $p_{ij} := \Pr(V_u = v_i, V_p = v_j)$ with $i = 1, 2$ and $j = 3, 4$, and with $p_{24} = 1 - p_{13} - p_{14} - p_{23}$. Note that unobserved heterogeneity adds 7 parameters to the model, but two of these need to be normalized as V_i enters θ_i multiplicatively.

The covariance of V_u and V_p equals

$$\text{cov}(V_u, V_p) = (p_{13}p_{24} - p_{14}p_{23}) \cdot (v_1 - v_2) \cdot (v_3 - v_4)$$

It is easy to show that V_u and V_p are independent if and only if $\text{cov}(V_u, V_p) = 0$.

In the estimation procedure we actually estimate the transformed probabilities q_{ij} which are implicitly defined by

$$p_{ij} = \frac{\exp(q_{ij})}{\sum_{i^*=1}^{2}\sum_{j^*=3}^{4}\exp(q_{i^*j^*})} \qquad i = 1,2;\ j = 3,4.$$

Because the p_{ij} sum to one, we normalize by taking $q_{24} = 0$. There is a one-to-one mapping between admissible values of p_{13}, p_{14} and p_{23} on the one hand, and q_{13}, q_{14} and q_{23} on $(-\infty, \infty)$ on the other. So, estimating the q_{ij} instead of the p_{ij} has the advantage that no boundary restrictions have to be imposed on the parameter space. Moreover, conditional on $v_1 \neq v_2$ and $v_3 \neq v_4$, there holds that $\text{corr}(V_u, V_p) = 0$ if and only if $q_{23} = q_{13} - q_{14}$.

5.2 Estimation results for the basic model

We estimate the models using the method of Maximum Likelihood. We take the unit of time to be one calender time day. For the categorical variables in x we have the following baseline categories: education = less than short senior high school; gender = male; unemployment benefits type = none; nationality = not in Eastern Europe, Africa, Asia, or South America; occupation type = manufacturing. Log age and log hourly wage in the previous job are measured in deviation from their mean across all spells. The "constant terms" in θ_u and θ_p are represented by the means of V_u and V_p, respectively, which is why we normalize $\lambda_{u1} = \lambda_{p1} = 0$ and why x does not include a constant.

The parameter estimates in Table 3 concern the basic model specification, i.e. Model 1 with the following restrictions: δ is a constant, the lengths of the time intervals spent within AMU and within other programs are set to zero, and within a spell any subsequent participation in AMU after the first course is ignored. We do include data on multiple unemployment spells per individual. To keep the computational burden manageable, we do not disaggregate the 4 occupational categories further. Also, we capture local labor market conditions by the local unemployment rate instead of using yearly or monthly dummy variables.[11] As a result, the number n_x of elements in the vector x equals 21, and the model has $20 + 2n_x = 62$ unrestricted parameters: $v_1, v_2, v_3, v_4, q_{13}, q_{14}, q_{23}, \delta, \beta_u, \beta_p, \lambda_{uj}(j = 2, ..8)$, and $\lambda_{pj}(j = 2, ..6)$.

The reported value of $-\infty$ for $\log v_3$ requires some explanation. The iterative estimation routines always converge to large negative values for $\log v_3$, but the value varies with the starting values of the estimation routine, and the corresponding standard error is always very large. The likelihood value and the estimates and standard errors of the other estimates are always the same and indeed are the same as when $\log v_3 = -\infty$ is imposed. Clearly, the convergence values are driven by numerical limitations of the computer program. Taken literally, the results imply that there is a fraction of workers who have a zero inflow rate

[11]Specifically, we include the mean-centered log municipal unemployment rate in the inflow year.

Table 3: *Estimation results for the basic model specification (first part).*

	To work θ_u		**To AMU training** θ_p	
Training effect				
δ	0.68	(0.053)*		
Individual characteristics				
log (age)	−0.43	(0.061)*	0.15	(0.15)
level of education:				
short senior high school	0.063	(0.036)	0.23	(0.086)*
senior high school	−0.000	(0.039)	0.15	(0.091)
short tertiary education	0.063	(0.062)	0.20	(0.15)
university	0.23	(0.044)*	0.094	(0.11)
female	0.026	(0.030)	−0.037	(0.072)
unemployment benefits:				
UI recipient	0.26	(0.033)*	0.22	(0.081)*
cash allowance recipient	0.14	(0.057)*	0.34	(0.13)*
nationality:				
from Eastern Europe	−0.49	(0.071)*	0.19	(0.13)
from Africa/Asia/S.America	−0.75	(0.078)*	−0.10	(0.15)
log (hourly wage)	0.094	(0.049)	−0.062	(0.093)
experience in occupation	0.089	(0.030)*	0.13	(0.074)
education in occupation	0.19	(0.029)*	0.12	(0.071)
occupation:				
professional, technical, agric.	−0.026	(0.040)	0.003	(0.093)
health, nursing and social care	0.18	(0.047)*	−0.41	(0.13)*
adm., managerial, sales, clerical,service	−0.22	(0.037)*	0.019	(0.086)
large city	−0.19	(0.026)*	−0.35	(0.064)*
needs guidance	−0.54	(0.058)*	0.27	(0.11)*
willing to move	0.059	(0.034)	0.20	(0.081)*
accepts part-time work	−0.042	(0.056)	−0.45	(0.16)*
relative unemployment rate	−0.68	(0.041)*	0.19	(0.11)

Explanatory note: Standard errors in parentheses. The superindex $*$ denotes significance at the 5% level (only for elements in β_i and λ_i (with $i = u, p$) and δ).

Table 3 *(continued)*.

	To work θ_u		**To AMU training** θ_p	
Duration dependence				
λ_{i2}	0.089	(0.028)*	−0.50	(0.089)*
λ_{i3}	0.10	(0.035)*	−0.51	(0.10)*
λ_{i4}	0.048	(0.043)	−0.40	(0.11)*
λ_{i5}	0.028	(0.051)	−0.36	(0.12)*
λ_{i6}	−0.068	(0.063)	−0.26	(0.10)*
λ_{i7}	−0.088	(0.072)		
λ_{i8}	−0.19	(0.057)*		
Unobserved heterogeneity				
$\log v_1$	−5.30	(0.058)		
$\log v_2$	−6.67	(0.075)		
$\log v_3$			$-\infty$	
$\log v_4$			−7.17	(0.21)
q_{13}	-0.23 (0.78)			
q_{14}	-0.069 (0.33)			
q_{23}	-1.78 (1.39)			
log likelihood	-126453.2			
number of individuals	16467			

Explanatory note: Standard errors in parentheses. The superindex $*$ denotes significance at the 5% level (only for elements in β_i and λ_i (with $i = u, p$) and δ).

into training. This may be true, or it may be that the actual inflow rate is a small positive number.

The main parameter of interest is the causal effect δ of training on the transition rate to work. The estimated value of δ is 0.68 and is significantly different from 0. Training thus raises this transition rate with about 100%, which means that it doubles. The effect on the mean or median unemployment duration depends on the moment at which training occurs. If the training is given within the first month then the mean duration is more or less reduced by half. Similarly, training at a relatively early stage in an unemployment spell has a large effect on the probability of long-term unemployment. (Of course, such a policy can be costly if implemented on a wide scale.) Recall that (part of) the effect may be due to increased search effort on the part of the caseworker, both before and during the participation period.

Now let us turn to the covariate effects β_u and duration dependence λ_u of the transition rate to work. To the extent that they are also estimated in other studies on recent Swedish unemployment durations, like Carling, Holmlund and Vejsiu (2001), the results are similar to those reported in those studies. The signs of the significant covariate effects are as expected. The exit rate to work is significantly lower for older and non-Swedish individuals and higher for university graduates. It is also higher for unemployment benefits recipients, reflecting the stronger labor market attachment of these individuals. There are no significant disincentive effect of high benefits as represented by the previous wage. Note however that this variable presumably also captures the mean wage offer. Individuals in large cities and in areas and years with high unemployment have a lower exit rate to work, whereas individuals with experience in their occupation or with an education that fits in with their occupation have a higher rate. Finally, individuals deemed to be in need of guidance by the caseworker at the moment of entry into unemployment have a much lower exit rate than others. This captures characteristics of the individual that are not fully described by the observed explanatory variables. The estimated duration dependence of θ_u is such that the individual transition rate to work decreases as the duration increases. Apparently, stigmatization and discouraged worker effects play a significant role here. Also, some individuals may enter a loop of successive periods of unemployment and workfare.

To some extent, the effects β_p of individual characteristics on θ_p can be interpreted as resulting from cost-benefits considerations by the caseworkers. For example, for individuals with the lowest education, AMU courses are presumably too difficult so they should not enter training. Also, individuals with occupations in the health, nursing and social care sectors do not need AMU because their job finding rates are relatively high anyway. Individuals who are willing to accept part-time jobs may benefit less from human capital accumulation in terms of earnings capacity, so they should have lower priority. Note that such considerations call for an analysis of heterogeneous treatment effects (see the next subsection). Individuals who are entitled to unemployment benefits should have a higher priority because of their opportunity costs. However, entitlement also signals a prolonged commitment to labor market institutions, and this may enhance their chances of being admitted to AMU training. If the individual is in need of guidance then the rate of entering training is much higher than otherwise. Finally, the estimated rate of entering training is highest during the first 56 days of unemployment.

Concerning the estimated unobserved heterogeneity distribution we find that $v_1 > v_2$ and $v_3 < v_4$. It is not difficult to see that the estimated correlation between V_u and V_p is negative, implying that individuals with unobserved factors that increase the exit rate to work have a lower rate into training.

As a first informal check on the robustness of the covariate effects, we compare them to those obtained from simpler specifications in which it is imposed that there is no unobserved heterogeneity or that the heterogeneity is independent across the hazard rates. In both cases, the treatment is by assumption exogenous. Also, in both cases, the parameters of θ_u can be estimated in isolation from those in θ_p. In the no-unobserved heterogeneity case, the estimated treatment effect is equal to 0.41. It turns out that the other estimates are very close to those reported in Table 3.[12] The covariate effects are a few percent smaller. This is not surprising. Typically, when unobserved heterogeneity is ignored in duration analysis, the estimated duration dependence is more negative (i.e., θ_u decreases more over time), and the estimated covariate effects on the hazard rate are smaller (see e.g. Van den Berg, 2001, for an overview). If we allow for unobserved heterogeneity but erroneously assume that $V_u \perp\!\!\!\perp V_p$ then the estimated treatment effect equals 0.55. So, ignoring selectivity leads to a slight under-estimation of this effect. The other estimates are almost indistinguishable from those reported in Table 3.[13]

5.3 Duration dependence and unobserved heterogeneity of the treatment effect on the individual transition rate to work

The previous subsection assumed homogeneity of the treatment effect δ on the exit rate to work, over individuals and over time. (Of course, the treatment effect on other outcomes of interest, like the mean duration or the fraction employed within a year is heterogeneous, due to the nonlinear way in which they depend on δ and x, v_u, v_p.) We now allow for heterogeneous treatment effects.

[12] For brevity, we do not report these. All unreported estimates are available upon request.

[13] We also attempted to estimate the model with one spell per individual. The estimates of δ and the distribution of V converge to limiting values such that part of the population has a zero inflow rate into training while the other part has a zero inflow rate into work before training and a moderate inflow rate into work after training. With multiple-spell data this is impossible because the data contain individuals who sometimes move into training and sometimes do not.

First, we only allow δ to depend on the time $t - t_p$ that has elapsed since AMU participation. The data show that many individuals move to employment closely after they leave training, and this shows that δ is not constant over time. The treatment effect may be smaller if the elapsed time is large, because of the termination of the caseworker's increased search assistance during participation, and because the acquired human capital may depreciate after some time. Also, as we have seen, heterogeneity of δ across individuals generates spurious duration dependence of δ as a function of $t - t_p$.

We take δ to be a piecewise constant function of $t - t_p$, by analogy to the duration dependence parameterization of the hazard rates in Subsection 5.1, so $\delta(t-t_p) = \sum_j \lambda_{\delta,j} I_j(t-t_p)$ where j denotes time intervals and $I_j(t)$ are time-varying dummy variables that are one in consecutive time intervals. We report estimates for the case in which δ is constant within the first three two-week intervals after training and is constant after the sixth week (so j attains 4 values). The results are not very sensitive to the choice of intervals, and most of the action occurs within the first weeks after training.

Table 4 gives the estimates for δ, or, more precisely, the estimates of $\lambda_{\delta,j}$. Clearly, the training effect is very large right after the training participation period. It is 6.5 times as likely to move to employment within two weeks after AMU training, in comparison to when the individual would not have participated in the training. After the first two weeks, the effect is still positive but it is smaller in magnitude. Between 2 and 6 weeks the transition rate to work is about 1.6 times larger, whereas after 6 weeks it is about 1.3 times larger. The likelihood ratio test of constancy of δ results in rejection of the null hypothesis at all conventional levels of significance.

Table 4: *Estimation results for the model in which the training effect on the transition rate to work is allowed to depend on the elapsed time since training.*

Training effect on θ_u		
time since training:		
≤ 2 weeks	1.87	(0.071)*
between 2 and 4 weeks	0.49	(0.15)*
between 4 and 6 weeks	0.49	(0.16)*
> 6 weeks	0.28	(0.070)*
log likelihood	-126213	
number of individuals	16467	

Explanatory note: Standard errors in parentheses. The superindex $*$ denotes significance at the 5% level.

For sake of brevity we do not report the other parameter estimates for this model. The estimates of the covariate effects β_u and β_p and their standard errors are virtually the same as in Table 3. This is also true for the estimates of the duration dependence λ_p. The estimated duration dependence λ_u is slightly less negative, which is not surprising given that now $\delta(t-t_p)$ has become a source of negative duration dependence as well. The estimates of the unobserved heterogeneity distribution also change slightly. Notably, $\log v_3$ is estimated to equal -8.04, so now the estimate of v_3 strictly exceeds 0.

As noted, one explanation for the observed negative duration dependence of the training effect is that the individual effect is heterogeneous. To proceed, we estimate models that allow for such heterogeneity. We start by incorporating individual characteristics x in δ, by specifying that δ is the sum of the above-used piecewise constant duration dependence term and a term $x'\beta_\delta$. For computational reasons we restrict the vector x in δ to six elements. We also add t_p as an explanatory variable.

Table 5: *Estimation results for the model in which the training effect on the transition rate to work is allowed to depend on a number of individual characteristics, the moment of training, and the elapsed time since training.*

Training effect on θ_u		
time since training:		
$\leq$ 2 weeks	2.15	(0.13)*
between 2 and 4 weeks	0.77	(0.16)*
between 4 and 6 weeks	0.79	(0.17)*
$>$ 6 weeks	0.61	(0.086)*
individual characteristics:		
log(age)	−0.51	(0.19)*
education $>$ senior high school	−0.34	(0.11)*
female	−0.13	(0.087)
unemployment benefits	−0.20	(0.12)
immigrant (E.Eu.,Af.,As.,S.Am.)	0.30	(0.15)
relative unemployment rate	−0.034	(0.15)
log (t_p)	−0.096	(0.037)*
log likelihood	-126188	
number of individuals	16467	

Explanatory note: Standard errors in parentheses. The superindex $*$ denotes significance at the 5% level.

Table 5 gives the estimates for δ, or, more precisely, the estimates of $\lambda_{\delta,j}$ and $\beta_{\delta,j}$. Again, we do not report the other parameter estimates for this model because these are virtually the same as before (even the β_u coefficients corresponding to the covariates included in δ). We first examine the covariate effects $\beta_{\delta,j}$. The training effect is significantly smaller for elderly individuals, for those with a high level of education, and for those who are trained when they are long-term unemployed. The educational effect can be explained by noting that not many courses are available at an academic level so that individuals with a high level of education may not be able to benefit as much from training as other individuals. The effect of t_p on the treatment effect can be due to selectivity of those who are treated later. We return to this below.[14] Also, the effect of t_p may reflect an effect of t, because of the "age-period-cohort" problem that only two of the three effects of $t, t_p, t - t_p$ are identified. The estimated effects for women and immigrants are not significantly different from zero. The likelihood ratio test of the hypothesis that δ does not depend on observed individual characteristics leads to rejection at all conventional levels of significance.

To compare the duration dependence coefficients in Table 5 to those in Table 4, notice that the explanatory variables are not measured in deviation from their mean, except for log(age), the relative unemployment rate, $\log t_p$. In addition, the average of x among those who are exposed to training is different from the average of x in the inflow into unemployment, because those with favorable characteristics will have found a job before T_p is realized. For an individual with characteristics equalling the average in the inflow, the training effect δ has estimated values 1.89, 0.51, 0.53, and 0.35 respectively as the time since training proceeds over the four different intervals that we distinguish. The number of 1.89 for when the time since training $t - t_p$ is less than or equal to two weeks is slightly larger than the corresponding value in Table 4. After the initial two weeks, the treatment effect heterogeneity in x gives rise to a dynamic selection, leading to negative duration dependence in the effect averaged over x among the survivors. So, part of the estimated negative duration dependence in δ in Table 4 is now captured by the heterogeneity in x. As a result, the estimated duration dependence in Table 5 is less negative than in Table 4. In other words, not taking heterogeneity of the treatment effect into account leads to an over-estimate of the speed at which the treatment effect vanishes after the treatment. Nevertheless, the shape of the training effect as a function of the elapsed time since training is qualitatively the same as before.

This line of reasoning naturally leads to the question whether the negative duration dependence in the training effect may be due to dynamic selection because of unobserved heterogeneity. We therefore estimate a model in which δ not only depends on $t - t_p$ and x but also on unobserved heterogeneity. This relies on the novel identification result that we proved in Subsection 3.2. Note that we should not include t_p as a covariate in δ as it is not clear whether such a model is identified without parametric functional form restrictions.

To keep the estimation manageable, we assume that V_δ (and therefore $\log \delta$) is a linear function of $\log V_u$ which is the unobserved heterogeneity term in the exit rate to

[14] Recall that we allow for selectivity of being treated at t_p, but not yet for systematically different treatment effects of those with certain unobserved characteristics.

work. This is equivalent to a one-factor loading specification for $V_\delta, \log V_u$. Specifically, $\delta(t - t_p, x, V) = \sum \lambda_{\delta,j} I_j(t - t_p) + x'\beta_\delta + \alpha \log V_u$, where $I_j(t)$ are again time-varying dummy variables that are one in consecutive time intervals, and β_δ is a vector.

Table 6: *Estimation results for the model in which the training effect on the transition rate to work is allowed to depend on a number of individual characteristics, the elapsed time since training, and on unobserved heterogeneity.*

Training effect on θ_u		
time since training:		
coefficients on		
≤ 2 weeks	1.39	(0.18)*
between 2 and 4 weeks	0.74	(0.19)*
between 4 and 6 weeks	0.75	(0.20)*
> 6 weeks	0.47	(0.14)*
individual characteristics:		
coefficients on		
log(age)	−0.43	(0.19)*
education > senior high school	−0.30	(0.12)*
female	−0.025	(0.090)
unemployment benefits	−0.10	(0.13)
immigrant (E.Eu.,Af.,As.,S.Am.)	0.51	(0.16)*
relative unemployment rate	0.11	(0.16)
unobserved heterogeneity V_u:		
coefficient on $\log v_1$	−0.68	(0.043)*
Unobserved heterogeneity distribution		
$\log v_1$	−5.21	(0.059)
$\log v_2$	−6.34	(0.062)
$\log v_3$	−8.70	(0.30)
$\log v_4$	−6.97	(0.16)
q_{13}	−0.041	(0.17)
q_{14}	−4.24	(3.51)
q_{23}	−1.12	(0.57)
log likelihood	-126086	
number of individuals	16467	

Explanatory note: Standard errors in parentheses. The superindex $*$ denotes significance at the 5% level (only for coefficients of δ).

The model with $\delta(t, t_p, x)$ and the model with $\delta(t - t_p, x, V)$ are not nested. However, they only differ in whether t_p or v_1 is included as a regressor in δ, whereas the other 71 parameters are the same. The model with v_1 in δ produces a much higher likelihood value than the other model. From this one may conclude that unobserved heterogeneity in the treatment effect is an important feature, and, indeed, is more important than the way in which the treatment effect depends on the moment of treatment.

The covariate effects in Table 6 are not very different from those in Table 5, except that now immigrants have a significantly higher treatment effect than natives.[15] The level of the duration dependence coefficients in Table 6 is again difficult to compare to those in previous tables. The unobserved heterogeneity term V_u is not included in deviation from its mean, and, more importantly, the mean of V_u among those who are exposed to training is lower than the mean of V_u in the inflow into unemployment. The latter is due to the dynamic selection driven by the effect of V_u on the exit rate to work before the treatment is realized.

The observed and unobserved heterogeneity in the determinants of the treatment effect give rise to a large amount of heterogeneity of the treatment effect itself across individuals. For individuals with $V_u = v_1$, the treatment effect on the exit rate to work is so large that virtually all of them will leave unemployment within some weeks after training. Of course, with V_u having a high value, the exit rate to work is relatively high, so that many would move to work before T_p is realized. Also, among those for whom V_u is large, the probability that V_p is large is very small (in the inflow it is 1.4%, which follows from $\Pr(V_u = v_1, V_p = v_3) = 0.417$ and $\Pr(V_u = v_1, V_p = v_4) = 0.006$), so their treatment rate is very small. In sum, very few v_1-individuals will be exposed to the treatment. Note that in reality, for given x, there are most likely more than two types of individuals, and it is not clear for how many of them the estimated value of v_1 is appropriate.

Let us return to the duration dependence shape of the treatment effect. The treatment effect heterogeneity in V gives rise to an additional dynamic selection, leading to negative duration dependence in the treatment effect averaged over V among the survivors. As follows from the previous paragraph, in our estimated model this issue is primarily relevant in the very first weeks after training. It implies that part of the estimated negative duration dependence in δ in Table 5 is now explained by the heterogeneity in V. As a result, the estimated duration dependence in Table 6 is less negative than in Table 5. In other words, not taking unobserved heterogeneity of the treatment effect into account leads to an overestimate of the speed at which the treatment effect vanishes after the treatment.

Nevertheless, for many individuals the shape of the training effect as a function of the elapsed time since training is qualitatively similar to before. For many individuals, the exit rate to work in the first two weeks after training is more than 3 times larger than the counterfactual exit rate in the absence of training. Also, the individual exit rate to work

[15] In Table 5 the t-value is 1.95.

in the seventh week after training is only 40% (=exp(0.47 − 1.39)) of the exit rate during the first two weeks after training.

Consider for example a native male individual, aged 36, with a high-school education level, who receives unemployment insurance, lives in a region with average labor market conditions, and has $V_u = v_2$. His exit rate to work in the first two weeks after training is 3.6 times larger than in the absence of training, and his exit rate to work in the seventh week after training is only 1.4 times larger than in the absence of training. For an otherwise equal woman, these numbers are virtually the same (3.5 and 1.4). If they would have age 55 instead of 36 then these numbers are 3.0 and 1.2. If the individual is an immigrant and/or aged in their twenties then the effects are larger, while if he/she is highly educated then they are smaller. Notice that for identification reasons we do not allow for interaction effects between covariates and the time since training.

5.4 Interpretation of the treatment-effect estimates

Recall that the official purpose of AMU is to increase the transition rate to work, and that this is supposed to be achieved by way of skill enhancements, i.e. by way of productivity improvements. Human capital accumulation by itself cannot explain the peak in the individual exit rate to work right after training. After all, skills do not depreciate at the rate at which the training effect decreases. The same applies to signaling-and-screening effects of having completed training. It is also difficult to explain the peak by the accumulation of vacancies or job offers during the period of training: it is hard to imagine that an employer is willing to reserve a vacancy for an individual who may or may not be available 3 or 6 months later (and who may by then have accumulated the appropriate skills).

It is therefore likely that the peak is driven by the increased job-search assistance efforts by the caseworkers towards the end of the training period. These efforts make it easier for employers to find suitable workers. After the training is completed, the job-search assistance efforts return to their normal level. This explains why the treatment effect does not stay at a high level as the elapsed time since training increases.

In our estimates, the effect on the exit rate to work does not vanish as time proceeds. Six weeks after the training, the exit rate is often still 40% higher than in the absence of training, which is substantial. Of course it remains to be seen whether this magnitude is persistent in case of model general model specifications. Presumably, the resulting estimates for the long-run effect will be smaller than 40% if we allow for a larger number of time intervals in the piecewise-constant duration-dependence of the effect, and/or more possible realizations of unobserved heterogeneity in the effect, and/or a non-deterministic relation between V_δ and V_u, and/or observed course heterogeneity, and/or interactions between covariates and duration dependence in the effect. (The computational demands would however increase dramatically.) The fact that other studies have consistently failed to find AMU effects on income supports the hypothesis that the productivity-enhancement effect is small. This is in line with a treatment effect on the exit rate to work that converges in the long run to a small number.

If the success of the job search assistance efforts right after training is due to a productivity improvement of the worker, then the total training effect can be increased by extending the job search assistance efforts in time. If the success of the job search assistance efforts is simply due to bridging the search-frictions information gap, then the training courses are not needed in the first place. In either case, it seems that the over-all effect can be improved if resources are reallocated away from the training courses towards job search assistance. Also, in the light of the amount of heterogeneity in the training effect, it seems that AMU can be made more effective by a stronger pre-screening and selection of potential candidates on their characteristics.

5.5 Time in training

We now incorporate the time spent in AMU training into the duration analysis, that is, we drop the rule that the length of the time interval spent within AMU training is set to zero. There are two major issues involved. First, the time in training adds to the time out of work. This is the so-called lock-in effect of training. Treatment effects on the total time out of work should incorporate this. This can be carried out in a straightforward way with the results of the previous subsections, by carefully adding the time in training to the duration out of work for someone who has been in training, when calculating effects on the mean duration out of work.

The second issue is that individuals may influence the time they are in training. In the data, the time in training is not constant across spells with training. To a large extent this reflects course heterogeneity. It is not known whether course heterogeneity gives rise to selectivity in course enrollment, and also not whether the moment at which training ends is driven by selective drop-out behavior. It is difficult deal with such selection effects. Relatively short times in training are rarely observed, so perhaps selective drop-out is not a major concern.

Incorporating the time in training into the duration model may lead to different estimated effects on the exit rate to work. Consider for example the training effect directly after leaving training. Let t_0 be the duration at entry into training. So far, the effect has been obtained by comparing (a) the exit rate of individuals who complete training, evaluated at the moment they complete training, to (b) the exit rate of individuals who have not entered training at t_0, evaluated at t_0, while appropriately taking selection effects into account. With a positive time in training, one may argue that (b) should be replaced by a measure that reflects the exit rate of non-trained individuals evaluated at t_0 plus the training time. After all, the duration dependence in the exit rate to work may be driven by the time out of work rather than the time in open unemployment.

From the previous two paragraphs it follows that models incorporating the time in training would be complex, and we feel that the empirical analysis of such models is beyond the scope of the paper. Instead we provide estimates of treatment effects on the time out of work based on results from the previous subsections, and we provide estimates of relatively simple duration models to shed some further light on the size of the lock-in effect. It should be borne in mind that all these results do not take account of course-length selection effects or drop-out selection effects.

Let us now discuss the duration analyses. Consider the basic model estimated in Subsection 5.2, where we now let δ depend on whether the total time in training was short (smaller than 90 days) or long. If, as before, we stop the time clock during training, so that the duration variable does not take account of the time in training, then we obtain that δ is estimated by 0.67 for short courses (standard error 0.066) and by 0.76 for long courses (standard error 0.064). This suggests that long courses are slightly beneficial in their effect on the exit rate to work. Now suppose that we let the time clock run during training, and we let the treatment effect δ affect the exit rate to work from the *onset* of training, where δ only depends on whether the course is short or long. Clearly, this ignores that in reality the exit rate to work is very small during training and is large after training. The parameter estimate will capture some average of the low exit rate to work during training and the training effect after leaving training. Consequently, we expect the estimate of δ to be smaller than in Table 3. Indeed, we obtain that δ is estimated by 0.24 for short courses (standard error 0.069) and by –0.35 for long courses (standard error 0.084). What this shows is that the over-all training effect that comprises the lock-in effect and the post-training effect together is very small. As seen from the moment at which the individual enters training, the lock-in effect is of a similar order of magnitude, but with an opposite sign, as the post-training effect on the exit rate to work. This is due to the fact that individuals do not often move from training to work during the first five months of AMU training. Note that this explanation is consistent with the fact that the over-all effect is more negative for long courses than for short courses, After all, the longer courses have a substantially larger lock-in effect but only a marginally larger post-training effect.

The fact that the other parameter estimates are similar to before means that they are insensitive to whether we include time spent in other programs or not.

5.6 Participation in other programs

The present paper does not aim to analyze the effects of participation in other active labor market programs, neither in isolation from AMU training, nor in possible interaction with AMU training and its effects. We have already seen that AMU training is regarded as being fundamentally different from other programs, and substitution seems to be absent. The data suggest that AMU training and other programs are unrelated activities. Nevertheless, the way we handle the actual time spent in other programs may affect our estimate of the training effect on the exit rate to work. To shed some more light on this we estimate a model version in which time spent in other programs not ignored but is added to time in unemployment. For T_u this is appropriate if individuals move to employment at the same rate within other programs as they do when they are openly unemployed. For T_p this is appropriate if individuals move into AMU at the same rate when they are in other programs as they do when they are openly unemployed. The estimate of δ in the basic model now equals 0.54 (standard error 0.052), which is slightly lower than in Table 3. It was to be expected that the estimate would be somewhat lower, because now the treated are effectively compared to not-yet treated who are at an earlier stage of the unemployment spell than the comparison group used earlier.

The fact that the other parameter estimates are similar to before means that they are insensitive to whether we include time spent in other programs or not. The duration dependence of the inflow rate into AMU is more negative than before. This reflects the fact that individuals cannot be in AMU training and in another program at the same time.

We can combine our current approach to that of the previous subsection, meaning that we let the time clock run during training and other program participation, and we let the training effect δ affect the exit rate to work from the onset of training, where δ only depends on whether the training course is short or long. This gives an estimate for δ of 0.18 for short courses (standard error 0.067) and of –0.26 for long courses (standard error 0.081). So, the over-all training effect on the total time out of work is even closer to zero if we take time in other programs into account.

Finally, we return to the framework of Subsection 5.3 where the clock is halted during program participation and the treatment effect is heterogeneous. We estimate a model in which the effect is also allowed to depend on a dummy variable indicating whether the individual has participated in another program during the current spell of unemployment before entry into training. The corresponding parameter estimate equals 0.17 (standard error 0.068). This suggests that the effect of AMU training benefits from an earlier participation in another program. Recall however that we do not control for selectivity of such program participation.

6 Conclusions

The individual transition rate from unemployment to employment is significantly and substantially raised as a result of the individual's participation in an AMU vocational training course. Individual re-employment rates are more or less tripled upon completion of the AMU course. However, this large effect only holds during the first few weeks after completion of the course. This may reflect the fact that caseworkers provide extra effort to find a job for AMU participants towards the end of the course. The observed decline of the effect can only to a limited degree be explained by dynamic weeding out due to heterogeneity of individuals' skills and other characteristics. When we take the time spent within the program into account as well, then the net effect on the individual's unemployment duration is about zero. Thus, the program does not appear to be cost-effective.

It seems that the over-all effect can be improved if resources are reallocated away from the training course, in the sense that their durations are shortened, towards a more prolonged job search assistance effort by the caseworker after completion of the course. Also, in the light of the amount of heterogeneity in the training effect, it seems that AMU can be made more effective by a stronger pre-screening and selection of potential candidates on their characteristics.

There are some topics for further research. We have argued that the empirical analysis has benefited from the availability of multi-spell data. In fact, any stochastic dependence across spells of the same individual can only be due to the presence of heterogeneity. It is ruled out by assumption that realizations of the unemployment duration or the training program in one spell affect the distributions of these in another spell. This may be a strong assumption. A topic for future work would be to examine to what extent AMU courses have effects that carry over to future unemployment spells (although the finding that the effect is strongest right after completion of the course suggests that such long-term effects may be negligible). Another topic for further research is to add hourly wages as an outcome variable. Comprehensive Swedish register information on hourly wages is expected to become widely available in the near future.

Appendix. Proof of Proposition 1

Suppose first that the distribution of the observed explanatory variables X is degenerate. As in Abbring and Van den Berg (2003), the information in a large data set can then be summarized by

$$Q_p(t,t_p) := \Pr(T_u > t, T_p > t_p, T_u > T_p) \quad \text{and} \quad Q_u(t) := \Pr(T_u > t, T_u < T_p) \tag{5}$$

for all $(t,t_p) \in \mathbb{R}_+^2$. These are the so-called sub-survival-functions of (T_u, T_p) and T_u for the sub-populations with $T_u > T_p$ and $T_u < T_p$, respectively.

Note that the distribution of the *identified minimum* of (T_u, T_p), *i.e.* the smallest of T_u and T_p, together with the identity of this smallest duration, is fully characterized by (Q_p^0, Q_u), with $Q_p^0(t_p) := Q_p(-\infty, t_p)$ for all $t_p \in \mathbb{R}_+$.

Now let X be allowed to be heterogeneous. In obvious notation, the data then come in the form of a collection $\{Q_p, Q_u\} := \{(Q_p(\cdot|x), Q_u(\cdot|x)); x \in \mathcal{X}\}$ of conditional sub-survival functions. Here, $\mathcal{X} \subset \mathbb{R}^k$, $1 \leq k < \infty$, is the support of X.

We adopt the model of Subsection 3.2. In fact, we consider a slightly more general model where $\exp(x'\beta_i)$ is replaced by $\phi_i(x)$,

$$\begin{aligned}
\theta_p(t|x, V_p) &= \lambda_p(t) \cdot \phi_p(x) \cdot V_p \\
\theta_u(t|t_p, x, V_u, V_\delta) &= \lambda_u(t) \cdot \phi_u(x) \cdot V_u \cdot \exp(\delta(t - t_p, x, V_\delta) \cdot \mathrm{I}(t > t_p)) \\
\delta(t - t_p, x, V_\delta) &= \lambda_\delta(t - t_p) + \phi_\delta(x) + V_\delta
\end{aligned}$$

Notice that $\lambda_u(.), \phi_u(.)$ and V_u act multiplicatively on $\theta_u(t|t_p, x, V_u, V_\delta)$ while at the same time $\exp(\lambda_\delta(.)), \exp(\phi_\delta(.))$ and $\exp(V_\delta)$ act multiplicatively on $\theta_u(t|t_p, x, V_u, V_\delta)$. This asymmetry between notation indexed by u and p on the one hand and notation indexed by δ on the other hand complicates some expressions below.

We make the following regularity conditions, assumptions, and normalizations.

Assumption 1. Variation over observed covariates.
ϕ_u, ϕ_p, and ϕ_δ are continuous functions $\phi_p : \mathcal{X} \to (0,\infty)$ and $\phi_u : \mathcal{X} \to (0,\infty)$ and $\phi_\delta : \mathcal{X} \to (-\infty,\infty)$. Further, $\{(\phi_u(x), \phi_p(x)); x \in \mathcal{X}\}$ contains a non-empty open set in $\mathbb{R}^2$ and $\{\phi_u(x) \cdot \exp(\phi_\delta(x)); x \in \mathcal{X}\}$ contains a non-empty open interval in $\mathbb{R}$.

Assumption 2. Baseline hazards.
The functions $\lambda_u : \mathbb{R}_+ \to (0,\infty)$ and $\lambda_p : \mathbb{R}_+ \to (0,\infty)$ are continuous except at at most a finite number of known points. They have integrals

$$\Lambda_u(t) := \int_0^t \lambda_u(\tau) d\tau < \infty \quad \text{and} \quad \Lambda_p(t) := \int_0^t \lambda_p(\tau) d\tau < \infty$$

for all $t \in \mathbb{R}_+$. The function $\lambda_\delta : \mathbb{R}_+ \to (-\infty, \infty)$ is continuous except at at most a finite number of known points. Moreover, its integral

$$\Lambda_\Delta(t) := \int_0^t \exp(\lambda_\delta(\tau))d\tau$$

exists and is finite for all $t \in \mathbb{R}_+$. The functions λ_δ and λ_u are such that

$$K(t, t_p) := \int_{t_p}^t \lambda_u(\tau)\exp(\lambda_\delta(\tau - t_p))d\tau$$

exists and is finite for all $\{(t, t_p) \in \mathbb{R}^2 : t > t_p \geq 0\}$.

Assumption 3. Independence of observed and unobserved heterogeneity.
(V_u, V_p, V_δ) is independent of X.

Assumption 4. Positive unobserved heterogeneity.
$\Pr((V_u, V_p, \exp(V_\delta)) \in (0, \infty)^3) = 1$.

Assumption 5. Finite means of unobserved heterogeneity.
$\mathbb{E}[V_u] < \infty$ and $\mathbb{E}[V_p] < \infty$ and $\mathbb{E}[V_u V_p \exp(V_\delta)] < \infty$.

Assumption 6. Normalizations.
For some a priori chosen $x^* \in \mathcal{X}$, there holds that $\phi_u(x^*) = \phi_p(x^*) = 1$ and $\phi_\delta(x^*) = 0$. For some a priori chosen $t^* \in (0, \infty)$, $\Lambda_u(t^*) = \Lambda_p(t^*) = K(t^*, 0) = 1$.

We now introduce some new notation. Let $\lambda_\Delta, \phi_\Delta$, and V_Δ be defined such that for every t, x, V_u, V_δ,

$$V_\Delta = V_u \exp(V_\delta),$$

$$\lambda_\Delta(t) = \exp(\lambda_\delta(t))$$

$$\phi_\Delta(x) = \phi_u(x)\exp(\phi_\delta(x))$$

Consequently, the exit rate to work at $t > t_p$ can be expressed as

$$\theta(t|t_p, x, V_\Delta) = \lambda_u(t)\phi_\Delta(x)\lambda_\Delta(t - t_p)V_\Delta$$

By Proposition 2 in Abbring and Van den Berg (2003), the functions $\Lambda_u, \Lambda_p, \phi_u$ and ϕ_p and the joint distribution of V_u, V_p are identified from $\{Q_p^0, Q_u\}$, given the assumptions listed above. We proceed by subsequently identifying the functions Λ_Δ and ϕ_Δ and the joint distribution of V_u, V_Δ, V_p from Q_p. Once we established this, the identification of the functions Λ_δ and ϕ_δ and the joint distribution of V_u, V_p, V_δ follows immediately. The structure of the remainder of the proof is similar to the structure of the proof in Abbring and Van den Berg (2003) of identification of a model where δ depends on x, t, and unobserved heterogeneity V (their Proposition 4).

Let $\mathcal{L}$ denote the trivariate Laplace transform of the distribution G_Δ of (V_u, V_Δ, V_p), and $\mathcal{L}^{(\Delta p)}(z_1, z_2, z_3) := \partial^2 \mathcal{L}(z_1, z_2, z_3)/\partial z_2 \partial z_3$ for all $(z_1, z_2, z_3) \in (0, \infty)^3$. Note that Assumption 3 implies that $(V_u, V_p, V_\Delta) \perp\!\!\!\perp X$.

For fixed $x \in \mathcal{X}$, $\partial^2 Q_p(t, t_p|x)/\partial t \partial t_p$ and $\partial^2 Q_p(t, t_p|x^*)/\partial t \partial t_p$ exist for almost all $t, t_p \in \mathbb{R}_+$ such that $t_p < t$. For these (t, t_p),

$$\frac{\partial^2 Q_p(t, t_p|x)/\partial t \partial t_p}{\partial^2 Q_p(t, t_p|x^*) \partial t \partial t_p} = \phi_p(x) \phi_\Delta(x) \frac{\mathcal{L}^{(\Delta p)}(\phi_u(x)\Lambda_u(t_p), \phi_\Delta(x) K(t, t_p), \phi_p(x)\Lambda_p(t_p))}{\mathcal{L}^{(\Delta p)}(\Lambda_u(t_p), K(t, t_p), \Lambda_p(t_p))} \quad (6)$$

If $t \downarrow 0$ and $t_p \downarrow 0$, both sides of (6) above reduce to $\phi_p(x)\phi_\Delta(x)$ because, by assumption, $\mathbb{E}[V_\Delta V_p] = \lim_{z \downarrow (0,0,0)} \mathcal{L}^{(\Delta S)}(z) < \infty$. We have already identified ϕ_p, and the left-hand side is data, so this identifies ϕ_Δ.

For arbitrary $x \in \mathcal{X}$ and $t, t_p \in \mathbb{R}_+$ such that $t_p < t$ there also holds that

$$\frac{\partial Q_p(t, t_p|x)/\partial t_p}{\lambda_p(t_p)\phi_p(x)} = \mathcal{L}^{(p)}(\phi_u(x)\Lambda_u(t_p), \phi_\Delta(x) K(t, t_p), \phi_p(x)\Lambda_p(t_p)), \quad (7)$$

with $\mathcal{L}^{(p)}(z_1, z_2, z_3) := \partial \mathcal{L}(z_1, z_2, z_3)/\partial z_3$. Note that the left-hand side of this equation is already identified for all t_p arbitrarily close to zero. As $t_p \downarrow 0$, the right-hand side reduces to

$$\mathcal{L}^{(p)}(0, \phi_\Delta(x) K(t, 0), 0). \quad (8)$$

After imposing $t = t^*$, we can identify the completely monotone function $-\mathcal{L}^{(p)}(0, \cdot, 0)$ on a nonempty open set in $\mathbb{R}$ by appropriately varying x in equation (8). This identifies $-\mathcal{L}^{(p)}(0, z, 0)$ for all $z \in (0, \infty)$ because of the real analyticity of $-\mathcal{L}^{(p)}(0, \cdot, 0)$ (see below). Subsequently, the right-hand side of (8) is strictly monotone in K, so $K(t, 0)$ can be traced out by varying t, and as a result the function $K(t, 0)$ as a function of t is identified. Recall that $K(t, 0) = \int_0^t \lambda_u(\tau)\lambda_\Delta(\tau) d\tau$. Given Assumption 2, this suffices to identify $\Lambda_\Delta(t)$ for all $t > 0$.

Finally, by appropriately varying x and t in equation (7), we can trace $\mathcal{L}^{(p)}$ on an nonempty open subset of $\mathbb{R}^3$. This identifies $\mathcal{L}^{(p)}$ on $(0, \infty)^3$ because $-\mathcal{L}^{(p)}$ is real analytic. This in turn identifies $\mathcal{L}$ (see Abbring and Van den Berg, 2006). $\square$

References

Abbring, J.H. and G.J. van den Berg (2003), "The non-parametric identification of treatment effects in duration models", *Econometrica*, 71, 1491–1517.

Abbring, J.H. and G.J. van den Berg (2004), "Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models", *Empirical Economics*, 29, 5–20.

Abbring, J.H. and G.J. van den Berg (2005), "Social experiments and instrumental variables with duration outcomes", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H. and G.J. van den Berg (2006), "Corrigendum on the non-parametric identification of treatment effects in duration models", Working paper, Free University Amsterdam, Amsterdam.

AMS (1997), "Employment training", Working paper, AMS, Solna.

AMS (1998), "Labour market policy programmes", Working paper, AMS, Solna.

AMS (2001), "Yearly report on labour market policy programmes", Working paper, AMS, Solna (*in Swedish*).

Björklund, A. (1993), "The Swedish experience", in *Measuring labour market measures*, Ministry of Labour, Denmark, Copenhagen.

Brännäs, K. (2000), "Estimation in a duration model for evaluating educational programs", Working paper, Umeå University, Umeå.

Carling, K., P.A. Edin, A. Harkman, and B. Holmlund (1996), "Unemployment duration, unemployment benefits and labor market programs in Sweden", *Journal of Public Economics*, 59, 313–334.

Carling, K., B. Holmlund, and A. Vejsiu (2001), "Do benefit cuts boost job findings? Swedish evidence from the 1990s", *Economic Journal*, 111, 766–790.

Carling, K. and K. Richardson (2004), "The relative efficiency of labor market programs: Swedish experience from the 1990's", *Labour Economics*, 11, 335–354.

Dahlberg, M. and A. Forslund (2005), "Direct displacement effects of labour market programmes", *Scandinavian Journal of Economics*, 107, 475–494.

Edin, P.A., A. Harkman, B. Holmlund, and H. Söderberg (1998), "Escaping long-term unemployment in Sweden", Working paper, Uppsala University, Uppsala.

Edin, P.A. and B. Holmlund (1991), "Unemployment, vacancies and labour market programmes: Swedish evidence", in F. Padoa Schioppa, editor, *Mismatch and Labour Mobility*, Cambridge University Press, Cambridge.

Eriksson, M. (1997a), “Comparison of compensatory and non-compensatory models for selection into labour market training”, Working paper, Umeå University, Umeå.

Eriksson, M. (1997b), “Placement of unemployed into labour market programs: a quasi-experimental study”, Working paper, Umeå University, Umeå.

Fay, R.G. (1996), “Enhancing the effectiveness of active labour market policies: Evidence from programme evaluations in OECD countries”, Working paper, OECD, Paris.

Harkman, A. and A. Johansson (1999), “Training or subsidized jobs – what works?”, Working paper, AMS, Solna.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999), “The economics and econometrics of active labor market programs”, in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Heckman, J.J. and C.R. Taber (1994), “Econometric mixture models and more general models for unobservables in duration analysis”, *Statistical Methods in Medical Research*, 3, 279–302.

Honoré, B.E. (1991), “Identification results for duration models with multiple spells or time-varying covariates”, Working paper, Northwestern University, Evanston.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Larsson, L. (2003), “Evaluation of Swedish youth labour market programmes”, *Journal of Human Resources*, 38, 891–927.

Regnér, H. (1997), “Training at the job and training for a new job: two Swedish studies”, Working paper, SOFI, Stockholm.

Regnér, H. (2002), “A nonexperimental evaluation of training programs for the unemployed in Sweden”, *Labour Economics*, 9, 187–206.

Ridder, G. (1990), “The non-parametric identification of generalized accelerated failure-time models”, *Review of Economic Studies*, 57, 167–182.

Van den Berg, G.J. (2001), “Duration models: Specification, identification, and multiple durations”, in J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.