Evaluation of Swedish Youth Labor Market Programs

Laura Larsson

ABSTRACT

A nonparametric matching approach is applied to estimate the average effects of two active labor market programs for youth in Sweden: youth practice and labor market training. The results of the evaluation indicate either zero or negative effects of both programs on earnings, employment probability, and the probability of entering education in the short run, whereas the long-run effects are mainly zero or slightly positive. The results also suggest that youth practice was more effective—or "less harmful"—than labor market training. However, there is considerable heterogeneity in the estimated treatment effects among individuals.

I. Introduction

It is a well-known fact in many European countries that youth unemployment is more sensitive to fluctuations in the business cycle than adult unemployment. Traditionally, this also has been the case in Sweden. The unemployment rates of the youth labor force have also been higher. Thus, the explosive rise in youth unemployment during the crisis of the 1990s is hardly surprising: From a level of around 3 percent in 1990, the unemployment rate for individuals aged 20–24 rose to above 18 percent in 1993, as shown by Figure 1. For the youngest age group, the

Laura Larsson is a researcher at the Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Sweden. *She is grateful to Per-Anders Edin, Denis Fougère, Bertil Holmlund, Per Johansson, Jochen Kluve, Winfried Koeniger, Michael Lechner, Gerard van den Berg, and two anonymous referees for their helpful comments. Previous versions of this paper were presented at the IZA Summer School in Munich, EALE 1999, the IWH workshop in Halle, and seminars at IFAU and the Department of Economics in Uppsala. The author claims responsibility for all remaining errors. The data used in this article can be obtained beginning April 2004 through March 2007 from Laura Larsson, IFAU, P.O. Box 513, S-751 20 Uppsala, Sweden. E-mail: laura.larsson@ifau.uu.se.*

[Submitted October 2000; accepted May 2002]

ISSN 022-166X

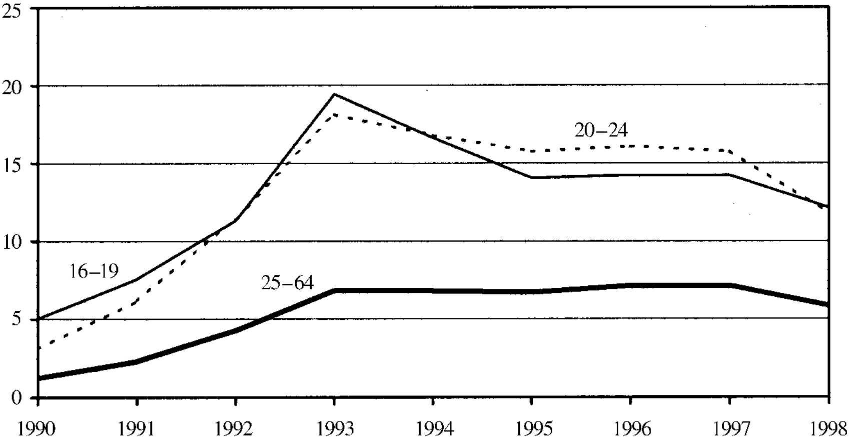

Figure 1
Unemployment Rate in Sweden by Age 1990–98.
Source: Statistics Sweden, Labor Force Surveys.

level of unemployment was even higher until 1994. Adult (aged 25–64) unemployment rose from slightly more than 1 percent to 7 percent. After the peak in 1993, the situation has improved for the young cohorts, whereas adult unemployment remained on the same level until 1997.

In response to rising unemployment figures, the Swedish government increased its spending on active labor market policy in order to improve the chances of the unemployed to return to regular employment. In 1992, a new large-scale program called youth practice, targeted at unemployed youth, was introduced. Since participants in active labor market programs are defined either as employed or as being outside the labor force, the immediate effect of such programs is that unemployment falls.[1] But this is solely a matter of accounting, whereas the longer-term effects remain largely uncertain. Thus, the evaluation of active labor market programs has become an increasingly important issue.

This paper evaluates the two most comprehensive active labor market programs in Sweden for youth, aged 20–24 years, in the first half of the 1990s, namely *youth practice* and *labor market training*. The objective is to determine the effects of the programs as compared to the outcome if the individual had continued to search for a job as openly unemployed.[2] The effects are measured in terms of earnings, employment probability, and the probability of entering studies provided by the regular educational system. The focus is on the direct effects of the programs; no attempt is made to assess the general equilibrium implications.[3]

1. In principle, participants in training programs (including youth practice) are excluded from the work force, whereas subsidized work programs are defined as employment.
2. ''Openly unemployed'' refers to the unemployed not participating in any active labor market program.
3. For a theoretical macroeconomic framework for studying both the direct and indirect effects, see Layard, Nickell, and Jackman (1991). Dahlberg and Forslund (1999) estimate the displacement effects of various active labor market programs, and find that programs providing subsidized labor displace on average 65

Identification of the average treatment effects is based on the conditional independence assumption (CIA), according to which participation in the various programs is independent of the post-program outcome, conditional on observable factors influencing both the decision to participate and the outcome. Given the CIA, matching on the propensity score using the multiple treatment approach introduced by Imbens (2000) and Lechner (2001) can be applied to obtain unbiased estimates of the average treatment effects on both the treated and the population. Here, a part of the paper is devoted to discussing the plausibility of the CIA in this context. Indirect tests of the CIA, as suggested by Heckman and Hotz (1989), are discussed, and the matching method is compared to some alternative, well-known methods for estimating average treatment effects based on different identifying assumptions.

Previous microeconomic studies of active labor market programs for Swedish youth report varying results. Edin and Holmlund (1991) and Korpi (1994) find negative effects on post-program employment, but positive or insignificant effects on the re-employment probability in subsequent unemployment spells. Ackum (1991) and Regnér (1997) mainly estimate negative program effects on earnings. However, except for Regnér (1997), these studies use the same small data set from the 1980s, and apply methods that rely on restrictive parametric assumptions. None of the previous studies evaluates the effects of youth practice.

Consequently, this study contributes to the Swedish and the international literature in several ways. First, it provides a number of new results on the effects of youth programs in Sweden. Second, it applies recently developed methodology to program evaluation. Third, it offers an example of how to make use of data based on comprehensive Employment Service records.

The paper is organized as follows. The evaluation problem, as well as the identification and estimation of average treatment effects under the conditional independence assumption is addressed in Section II. The labor market programs and the data are described in Section III. Section IV outlines the econometric analysis based on the propensity score matching approach, while Section V considers the sensitivity of the results. Section VI contains a discussion of alternative identification strategies and ways of (indirectly) testing conditional independence, and, finally, Section VII concludes.

II. Econometric Evaluation Strategies

A. The Evaluation Problem

This study attempts to determine and compare the outcomes of three alternative strategies available to a young unemployed individual: to participate in either youth practice or labor market training, or to continue searching for a job as openly unemployed. In other words, the aim is to determine the causal effect of a program compared to (1) the no-program state, and (2) the other program. Following Lechner (2001), among others, this multiple evaluation problem may be introduced as follows.

percent of the corresponding regular employment. Youth practice is regarded as such a program. Labor market training is not found to have any significant displacement effect, however.

Consider participation in $(M + 1)$ mutually exclusive treatments, denoted by an assignment indicator $T \in \{0, 1, \ldots, M\}$. Let the zero category indicate the *no-treatment* alternative. Moreover, denote variables unaffected by treatments, often called *attributes* (Holland 1986) or *covariates*, by X. The outcomes of the treatments are denoted by $\{Y^0, Y^1, \ldots, Y^M\}$ and, for any participant, only one of the components can be observed in the data. The remaining M outcomes are called counterfactuals. The number of observations in the population is N, such that $N = \sum_{m=0}^{M} N^m$, where N^m is the number of participants in treatment m.

The evaluation problem is to define the effect of treatment m compared to treatment l, for all combinations of $m, l \in \{0, 1, \ldots, M\}$, $m \neq l$. More formally, the outcomes of interest in this study are shown in the following equations:

$$(1)\quad \theta_0^{ml} = E(Y^m - Y^l | T = m) = E(Y^m | T = m) - E(Y^l | T = m),$$

$$(2)\quad \gamma_0^{ml} = E(Y^m - Y^l) = EY^m - EY^l.$$

θ_0^{ml} in Equation 1 denotes the expected average treatment effect of treatment m, relative to treatment l, for participants in treatment m (sample size N^m). In the binary case, where $m = 1$ and $l = 0$, this is usually called the 'treatment-on-the-treated' effect. γ_0^{ml} in Equation 2 is the corresponding expected effect for an individual drawn randomly from the whole population (N).[4]

The evaluation problem is characterized by missing data: the counterfactual $E(Y^l | T = m)$ for $m \neq l$ cannot be observed, since it is impossible to observe the same individual in several states at the same time. Thus, the true causal effect of treatment m relative to treatment l can never be identified. However, the *average* causal effects defined by Equations 1 and 2 can be identified under the conditional independence assumption; see subsection 2.3.[5]

B. Matching as an Evaluation Estimator

In experimental studies, participants are randomly assigned to treatment(s) from a large group of eligible applicants. In a binary case, a comparison between the treated and the control group, which consists of the individuals not assigned to the treatment, yields an unbiased estimate of the average treatment effect. Similarly, in a multiple case, an unbiased estimate of the average effect of one treatment compared to another is obtained by comparing the two randomly assigned treatment groups. This is not the case in nonexperimental studies, because the various treatment groups are likely to differ from each other in a nonrandom way. Hence, the objective of a nonexperi-

4. Note that the latter expected effect is symmetric in the sense that $\gamma_0^{ml} = -\gamma_0^{lm}$, whereas the same is not valid for the treatment effect on the treated, that is $\theta_0^{ml} \neq -\theta_0^{lm}$, as long as participants in treatments m and l differ in a nonrandom way.

5. Moreover, to make causal analysis possible, the stable-unit-treatment-value assumption (SUTVA) must be satisfied for all individuals in the population. The SUTVA has several consequences, the most important of which in our context is that the potential outcomes for an individual are independent of the treatment status of other individuals in the population. Thus, cross effects and general equilibrium effects are excluded. The term ''stable-unit-treatment-value'' refers to another implication of the assumption, namely that the treatment status of an individual (or ''unit'') is unrelated to the treatment status of other individuals. For a more detailed description and discussion of the SUTVA, see, for example, Angrist, Imbens, and Rubin (1996).

mental evaluation study is to construct a comparison group that is as close as possible to the experimental control group. One method suggested for solving this problem is matching.

Matching methods have been developed and widely used in the statistics and medical literature (Rubin 1977; Rosenbaum and Rubin 1983, 1984, 1985; Rubin and Thomas 1992), but are relatively new to economics and labor market policy evaluation. In short, matching involves pairing individuals from various treatment groups who are similar in terms of their observable characteristics. When selection into treatments and the outcome are based exclusively on these observable characteristics, matching yields unbiased estimates of the average treatment effects.

C. Conditional Independence Assumption

The crucial assumption behind matching is that all differences affecting the selection and the outcome between the groups of participants in treatment m and treatment l are captured by (to the evaluator) observable characteristics, X. In the evaluation literature, this assumption is called conditional independence, or unconfoundness. In the multiple case considered in this paper, the conditional independence assumption (CIA) is formalized as[6]

(3) $\{Y^0, Y^1, \ldots, Y^M\} \amalg T|X = x, \forall\, x \in \chi,$

where $\amalg$ is a symbol for independence and χ denotes the set of covariates for which the average treatment effect is defined. In words, the CIA requires treatment T to be independent of the entire set of outcomes, given X. That is, given all the relevant observable characteristics (X), when choosing among the available treatments (including the no-treatment alternative), an individual does not base her decision on the *actual* outcomes of the various treatments.[7] Individuals can, however, base their decisions on expected outcomes, as long as these are determined by X only. This implies that individuals expect their outcomes to equal the mean outcomes for people with similar (observed) characteristics. Moreover, in order for the average treatment effect to be identified, the probability of treatment m must be strictly between zero and one:

(4) $0 < P^m(x) < 1, \quad \text{where} \quad P^m(x) = E[P(T = m|X = x)], \quad \forall\, m = 0, 1, \ldots, M.$

In the binary case of two treatments, Rosenbaum and Rubin (1983) show that if the CIA is valid for X, it is also valid for a function of X called the *balancing score* $b(X)$, such that $X \amalg T|b(X)$. The balancing score property holds even for the multiple case:

6. The significance and consequences of the CIA in the binary case of one treated and one nontreated state have been explored and formalized by Rubin (1977) and Rosenbaum & Rubin (1983). The analysis of the multiple case presented here closely follows the analyses in Lechner (2001) and Imbens (2000).

7. Naturally, for identification of a single treatment with fixed m and l, it is sufficient to assume pair-wise independence $Y^l \amalg T = m,l|X = x, \forall x \in \chi$. Moreover, instead of conditional independence as in Equation 3, it is sufficient to assume conditional *mean* independence, which is a somewhat weaker assumption. However, in practical applications, it is difficult to find a situation where the latter, but not the former, is fulfilled. For a thorough discussion on identifying assumptions, see Heckman, Ichimura, and Todd (1998).

$$\text{(5)}\quad \{Y^0, Y^1, \ldots, Y^M\} \amalg T|X = x, \forall\, x \in \chi \rightarrow \{Y^0, Y^1, \ldots, Y^M\} \amalg T|b(X) = b(x),$$
$$\forall\, x \in \chi, \text{ if } E[P(T = m|X = x)|b(X) = b(x)] = P[T = m|X = x] = P^m(x),$$
$$0 < P^m(x) < 1, \forall\, m = 0, 1, \ldots, M.$$

The main advantage of the balancing score property is the decrease in dimensionality: Instead of conditioning on all the observable covariates, it is sufficient to condition on some function of the covariates. In the binary case of two treatments, the balancing score with the lowest dimension is the propensity score $P^1(x) = E[P(T = 1|X = x)]$. In the case of multiple treatments, a potential and quite intuitive balancing score is the M-dimensional vector of propensity scores $[P^1(x), P^2(x), \ldots, P^M(x)]$. Lechner (2001) shows, however, that the dimension can be further reduced to two, or even one. This is illustrated in the following section, which addresses identification of the average treatment effects.

D. Identification

Let us begin by considering the identification and estimation of the average treatment effect on the treated, θ_0^{ml}. The mean outcome of treatment m for participants in m, $E(Y^m|T = m)$, is identified and estimated by, for example, the sample mean. Lechner (2001) and Imbens (2000) show that the latter part of Equation 1, the mean outcome of treatment l for participants in m, $E(Y^l|T = m)$, can also be identified in sufficiently large samples, given conditional independence. To estimate it, they show that instead of the M-dimensional balancing score, the dimension of the condition set can be reduced to $[P^m(x), P^l(x)]$. Thus,

$$\text{(6)}\quad E(Y^l|T = m) = E[E(Y^l|P^m(X), P^l(X), T = l)|T = m].$$

Lechner (2001) shows that the dimension can be further reduced:

$$\text{(7)}\quad E(Y^l|T = m) = E[E(Y^l|P^{l|ml}(X), T = l)|T = m],$$

where $P^{l|ml}$ is the conditional choice probability of treatment l, given either treatment m or l. Both Equations 6 and 7 are suggested for estimating the average treatment effect on the treated.[8]

The identification and estimation of the average treatment effect for the whole population, γ_0^{ml}, may be carried out in several ways. Lechner (2001) suggests the following:

$$\text{(8)}\quad \gamma_0^{ml} = E(Y^m|T = m)P(T = m) + \underset{P^m(X)}{E}[E(Y^m|P^m(X), T = m)|T \neq m]P(T \neq m)$$
$$- E(Y^l|T = l)P(T = l) + \underset{P^l(X)}{E}[E(Y^l|P^l(X), T = l)|T \neq l]P(T \neq l).$$

8. $P^{l|ml}$ is identified as

$$E[P^{l|ml}(X)|P^l(X), P^m(X)] = E\left[\frac{P^l(X)}{P^l(X) + P^m(X)}\,\middle|\, P^l(X), P^m(X)\right] = P^{l|ml}(X).$$

In words, Equation 8 implies that the average treatment effect on the population is identified by a weighted sum of the treatment effects on all subsamples. For a more detailed description of the identification of θ_0^{ml} and γ_0^{ml}, see Imbens (2000) and Lechner (2001).

III. The Programs and the Data

Conditional independence cannot be regarded as a plausible assumption unless one is acquainted with the institutional settings—what was the purpose and content of the program? who participated and why?—and has reliable data on all these factors.

A. Description of the Programs

Youth practice (ungdomspraktik) was launched in July 1992, during the most severe period of rising unemployment in Swedish postwar history. By January 1993, the stock of participants aged 20–24 in youth practice reached its peak at 60,000, which corresponds to approximately 10 percent of the population in this age group.[9] Simultaneously, labor market training, the second largest program for that cohort, decreased from about 25,000 to 15,000 participants. During the period of July 1992–July 1993, participants in these two programs on average accounted for 85 percent of all people in this age group taking part in any program; in the following year, the share was 75 percent.[10] In October 1995, youth practice was replaced by new programs.

Youth practice consisted of a subsidized work program aimed at providing working experience for the young unemployed with a high school diploma.[11] Participants were placed in both the private and the public sector, and the program period was generally six months. For individuals aged 20–24, the allowance for participation was SEK 338[12] per day, of which the employers paid only a very small fraction. In the relatively rare cases where the participant was entitled to unemployment benefits, she received an allowance equal to the benefit.

According to the program regulations, participation should be preceded by at least four months' active job search as openly unemployed. In addition, participants should be a supplementary resource for the employer and not displace regular employment, and they should allocate 4–8 hours a week to job-seeking activities at the local employment office. In practice, however, participants often worked with tasks that would otherwise have required hiring a regular employee, and allocated very

9. Unemployed individuals aged 18–19 were eligible for youth practice but not for training. Thus, they are excluded from the study in order to fulfill the balancing score property, $X \amalg T|b(X)$.

10. Thus, it seems plausible to focus on the evaluation of these two programs only.

11. Formally, the program was supposed to be a ''mixture of subsidized work and training'' in the sense that it would improve the participants' human capital. However, implementation studies show that the tasks were often very simple, so that the share of training was more or less negligible (see, for example, Hallström 1994 and Schröder 1995).

12. Approximately USD 36.5, June 2002.

Table 1
Differences between Youth Practice and Labor Market Training

	Youth Practice	Labor Market Training
Content of the program	Subsidized work	Training courses
Duration of the program	Generally six months (some variation)	No general rule, up to 12 months (much variation)
Formal target group:		
Age	18–24 years	20–65 years
Education	High school diploma (some variation)	Low/wrong type of education for labor demand
Work experience	Little work experience	Low/wrong type of experience for labor demand
Labor market status before assignment to the program	Unemployed for at least four months	Unemployed or at risk for unemployment

little time to job seeking.[13] Moreover, the length of preprogram unemployment varied noticeably from two or three days to several months.

Labor market training, which has existed in various forms for decades and is still in effect, is aimed at improving the skills of the unemployed job seeker in order to match her to labor demand. Thus, it has traditionally been directed at individuals with low education and skills. However, the Swedish high school system seldom prepares fully trained workers, so that individuals with a high school diploma are part of the target group. The program consists of courses of various length and content, both vocational and nonvocational.[14] The age limit and the size of the allowance have changed over time, but during the period under study, the minimum age limit for participating in the program was 20 years. Moreover, the size of the allowance was the same in labor market training as in youth practice, and, according to the program regulations, participants should continue their job-seeking activities during the program. Table 1 summarizes the differences between the two programs.

Typically, an unemployed individual, in consultation with a placement officer at the local employment office, decided whether to participate in any of the programs and which program to choose. The reason for wanting to participate varied. Except for individuals who were eligible for unemployment benefits (and who thus received the same amount as participants in the programs), participation in either of the programs implied a financial benefit. Moreover, surveys among job seekers and place-

13. For example, participants might assist with some simple administrative tasks in a firm, or take care of children at a daycare center.

14. Although the heterogeneity of the program is ignored in the main analysis, results from an analysis where vocational and nonvocational courses are treated separately are reported in Section V.

ment officers indicate that many job seekers believed that participation in a program would improve their chances of finding a job, and many regarded youth practice as a ''real job'' (see, for example, Hallström 1994; Schröder 1995; Eriksson 1997).

An individual interested in youth practice was usually encouraged by the placement officer to find an employer willing to offer placement. This was intended to increase the individual's power of initiative. Consequently, individuals who managed to find an employer on their own might have had a better chance of participating than those who needed assistance from the local employment office. Sometimes, employers took the initiative and offered placement in youth practice if the local employment office arranged the financing.

Rejecting an offer to participate could, in principle, lead to suspension from unemployment benefits, if the unemployed person was entitled to any. However, in a situation where local employment offices were deluged with job searchers, those who needed help the most, comprising the least educated and experienced—and not entitled to benefits—were most likely to receive an offer, with perhaps one exception. In Sweden, unemployment benefits expire after 300 unemployment days unless the individual has qualified for a new 300–day period by working or participating in a labor market program for at least six months.[15] Therefore, unemployed people close to the benefit expiration date may have been more likely to be assigned into a program; see Section IIIA.

To conclude, it is reasonable to assume that the more experienced and better educated the unemployed individual and the shorter her unemployment period, the lower the probability of being offered and assigned to a program. Moreover, having a high school diploma should increase the propensity for youth practice relative to labor market training.

B. Description of the Data

The data used in this study, a random sample of approximately 200,000 individuals, were collected from the databases maintained by the Swedish National Labor Market Board and Statistics Sweden. The former database includes records of all individuals who have been registered with the Employment Service, whereas the latter records the annual earnings of all individuals residing in Sweden. For each individual in this study, registration dates, labor market status, and individual characteristics between August 1991 and March 1997 were combined with information on annual earnings for the years 1985–95. A more exact description of the variables used in the empirical analysis is given in Tables A1–A3 in the Appendix. Details regarding the outcome variables are given in Section IIIE.

In the Employment Service records, each job seeker is registered under some ''job-seeker category'' defining her labor market status. Examples of such categories are full-time openly unemployed, part-time openly unemployed, or participant in a labor market program. When signing up with the Employment Service, the unemployed persons are asked to fill out a ''search form'' that contains questions about individual characteristics, such as year of birth, citizenship, formal education, previous labor market experience, and type of job they are looking for. If an individual

15. The exact rules for qualifying for unemployment benefits are somewhat more complicated.

wishes to apply for several jobs, she is asked to give each application either a high or a low priority. The job seeker's county of residence and the code of the local employment office she visited are also recorded.

During a period in the Employment Service register, an individual may—and probably will—change categories prior to de-registration. In other words, an individual may have entered the register as openly unemployed, then participate in some labor market program, and again be openly unemployed before de-registration due to, for example, the transition to a regular job. All the relevant dates are provided in the data. The reason for de-registration is also recorded.

The database at Statistics Sweden covers all individuals residing in Sweden at the end of December each year. Information on earnings is based on firms' reports to the tax authorities. Earnings are measured on a yearly basis, and there is no information about the number of working hours. As a dependent variable in the empirical analysis of earnings, I used the annual sum of work-related income including the allowance for maternity or sickness leave and other work-related allowances from the social insurance system. Unemployment benefits are, of course, not included in this variable. The variation in the dependent variable can thus reflect changes in both wage rates and working hours.

C. Is it Plausible to Assume Conditional Independence?

The description of the programs indicates that the level of education, previous work experience, and preprogram unemployment history are important factors in determining whether an individual will participate in any program, as well as in which of the programs. These factors are also likely to influence the future labor market outcome, and thus, in order for conditional independence to be plausible, they should be included in the estimation of the propensities.

The importance of labor market history prior to a program is emphasized in various evaluation studies, starting with Aschenfelter (1978). Examples of more recent studies that all point to pretraining earnings as one of the most essential factors to be controlled for in a labor market program evaluation are Hotz, Imbens, and Mortimer (1998); Dehejia and Wahba (1999); Heckman, Ichimura, Todd, and Smith (1998).

Annual earnings for the preceding year, pretraining unemployment periods, level of education, and work experience are all included in the data available for this study.[16] Moreover, the data provide detailed information on other personal characteristics (see Table A2). Information is missing on whether a job searcher is entitled to unemployment benefits that, as discussed above, may provide an incentive to participate in a program. However, there are two arguments that may alleviate this potential shortcoming.

First, entitlement requires work experience which, in turn, implies labor earnings. Thus, by controlling for the latter two, we indirectly control for entitlement. Second, the mean preprogram unemployment periods in the participant samples are far from

16. The search form includes a question as to whether the job seeker thinks she has the relevant work experience for the type of work she wants. In the remainder of the paper, this is referred to as ''specific work experience.''

300 days, which is the benefit exhaustion limit.[17] Consequently, it is reasonable to assume that the participation decision of these individuals, even if entitled to benefits, is not significantly influenced by qualification for a new benefit period.

A factor often suggested as causing selection bias is ''motivation'' or some other unobservable personal quality of the job searcher that makes her more or less successful on the job market, and that also plays a role in the program assignment process.[18] It may be that the most highly motivated job seekers show the most interest in a program, and are thus most likely to be assigned to it. The opposite is also plausible: Caseworkers may be more eager to help the unemployed persons who are the least motivated. Either way, the estimated program effect will turn out to be biased.

In the Employment Service data, each openly unemployed job seeker is assigned a grade indicative of her readiness to take a job if employment is found. Examples of grades are ''can take a job directly'' or ''needs guidance.'' This grading is based on the employment officer's assessment of the job seeker and thus provides a measure of the job seeker's expected success on the job market.

Finally, the willingness to assign people into programs in general, and into the two programs under study in particular, varied among the local employment offices. It may be that the willingness to assign into programs is correlated with the ability to match the unemployed people with employers. Thus, variables based on records from the local employment offices also are included in the estimation of the propensities.

The bottom line is that the available data include *much*, but *not necessarily all*, information on factors which affect the selection and the outcome. The crucial question—that is left to the reader to decide—is whether there is *sufficient* information to justify the conditional independence assumption.

Later on, in Section VI, I discuss different ways of indirectly testing the plausibility of the CIA, either through preprogram outcome tests suggested by Heckman and Hotz (1989) or by applying various methods to the same problem and comparing the results. In short, I find that different methods produce somewhat different estimates for the program effects, but the sign of the effects is essentially the same across methods. Moreover, the preprogram outcome tests—as far as it is possible to apply and draw conclusions from them—provide support for the conditional independence assumption.

D. Sample Construction

From the database, I collected all individuals aged 20 to 24 who registered with the Employment Service during 1992 and 1993 as openly unemployed for the first time and with the grade ''can take a job directly.'' This procedure yielded 10,579 individuals. From this group, I then collected all individuals who, after having been openly unemployed, directly enrolled in youth practice or labor market training. The final

17. Table A1 in the Appendix shows that only 7–8 percent of the program participants had been unemployed more than 270 days before the start of the program.

18. In fact, very few evaluations based on the CIA explicitly discuss this motivational factor. One nice exception is the study by Gerfin and Lechner (2002) that applies rich Swiss data that actually does include such a variable.

group consisted of 1,657 youth practice participants and 606 labor market training participants.[19]

A potential comparison group consisted of individuals who entered the register as openly unemployed during the same period, and never participated in any of the programs. There were slightly more than 5,000 such individuals. All of them could, in principle, have been used as the group of nonparticipants in the empirical analysis. However, as already pointed out, the length of the unemployment period immediately before starting a program is an important factor in determining whether an individual will participate in any program and to which program she will be assigned. Hence, in order to be able to use this information when estimating of the propensities, I created a hypothetical starting date for nonparticipants. The following procedure is similar to the *random* procedure suggested by Lechner (1999).

First, the group of participants (here, participants in practice and training are regarded as a single group) and the group of nonparticipants were divided into subgroups by the month of registration with the Employment Service. Then, each of the nonparticipants in a subgroup was randomly assigned an observation of ''length of preprogram unemployment'' from the distribution of the contemporaneous group of participants. In cases where the nonparticipant's actual unemployment period was shorter than the assigned preprogram unemployment period, the individual was removed from the sample. This procedure deleted approximately 60 percent of the sample and left me with slightly more than 2,000 nonparticipants.[20]

It should be noted, however, that this group of nonparticipants does not necessarily represent a world without programs; such a construction is possible only in a case where the individuals know that choosing not to participate implies that they will never take part in that particular program. This is not a realistic assumption for Sweden, however, where most programs continue to exist, and the unemployed who have not succeeded in finding a job (or are deregistered from the Employment Service for some other reason) are offered new possibilities to participate.[21] Thus, in a strict sense, the 2,000 individuals in the comparison group represent the alternative not to participate but to *wait*, when the first chance is offered to them. But, they are referred to as nonparticipants because they never participated in any program.

Tables A1–A2 in the Appendix report descriptive statistics of some selected variables for the three groups. There are clear differences in both the program character-

19. By restricting the program period to represent a second ''job seeker category,'' heterogeneity in the participants' unemployment history could be reduced. Moreover, choosing the *first* program for every unemployed also appears to be the easiest way of handling multiple program participation. For a discussion of this dynamic program evaluation problem, see among others Gerfin and Lechner 2002; and Lechner and Miguel 2001. I also removed observations with negative program periods or other curious dates from the complete sample of 10,579 individuals.

20. This group of nonparticipants consists of individuals with, on average, longer unemployment periods than the original group of 5,000 individuals, because the risk of being excluded from the sample due to a ''too late'' assigned start of the program is higher, the shorter is the individual's unemployment period. This is desirable, however, because the aim is to match participants with nonparticipants who were unemployed long enough to be *potential* program participants.

21. The time limit for unemployment benefits, along with the possibility of renewing benefit entitlement by participating in programs, presumably strengthen the incentives to participate when approaching the 300-day limit. Thus, at least among the entitled, the probability of participating within 300 days, given that the individual is still unemployed, is very close to unity.

istics as well as the individual characteristics among participants in various states. As shown in Table A1, the duration of both preprogram unemployment and the program itself are shorter among participants in labor market training as compared to youth practice participants.

Moreover, the sample of labor market training participants consists of individuals who were registered with the Employment Service quite early and thus also started the program earlier than the practice participants. There are also differences in age, citizenship, education, and experience among the three groups.

Table A3 lists some selected statistics from the local employment offices. I assume the probability of being assigned into one of the three states to depend on, among other things, the proportion of all unemployed assigned to any program at the specific local employment office, the month before the actual assignment. That proportion may be considered a measure of how readily the office assigns individuals to programs. Furthermore, the decision between the two programs is assumed to be dependent on the ratio between participants in these programs. Given that these figures also reflect local labor market conditions and/or the effectiveness of the offices in finding jobs, they should be included in the propensity estimations.

As expected, the share of youth practice assignments is above the country average in the sample of youth practice participants. A corresponding pattern holds for training. Somewhat surprisingly, though, the total number of program assignments is below the country average in all three samples. As expected, however, the ratio is lowest in the sample of nonparticipants.

E. What is the Outcome of Interest?

An explicit aim of active labor market policy is to improve the employability of the unemployed people. Hence, a higher probability of future employment and higher earnings are obvious measures of a program's success. However, especially in the case of youth, a possible track to stable future employment might be regular education. Thus, in addition to employment probability and earnings, I used the probability of transition from unemployment to studies as a third measure of success.

Earnings are measured by a continuous variable, whereas dummy variables were constructed for the other outcome measures. Figure 2 illustrates the way the various outcomes are defined for a hypothetical individual in the sample. This individual signs up with the Employment Service in March 1992. In November 1992, she enrolls in youth practice for a period of six months. She is defined as "employed within one year (two years) after program start" if she is deregistered from the Employment Service due to regular employment by November 1993 (1994). Analogous definitions are used for regular education.[22]

Earnings one and two years after the start of a program are measured in a slightly less precise manner because I only had access to an annual sum of earnings with no information on working hours. For an individual who enrolled in a program during the first half of a calendar year, "earnings one year after program start" comprise

22. The data provide solely one kind of information; an individual who is both employed and a student is classified according to her "main activity."

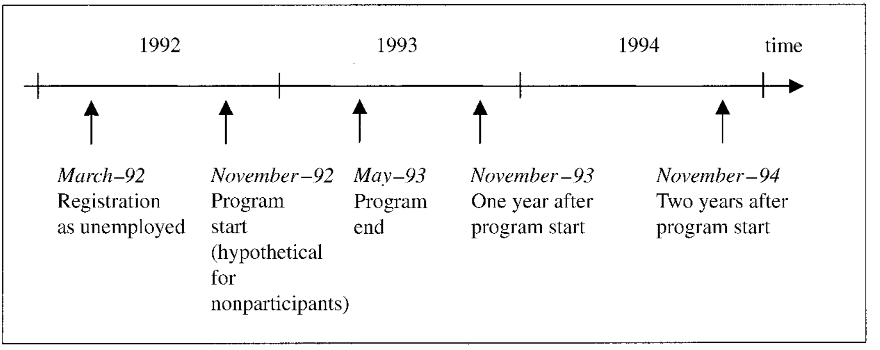

Figure 2
A Registration Period and Its Outcome Measures

the annual sum of earnings for the following calendar year. For individuals who started their program in July–December, I instead use the average of the following two calendar years to avoid counting (zero) earnings during or directly after the start of the program. Thus, for the hypothetical individual in the above example, ''earnings one year (two years) after program start'' are the average of her earnings in 1993 and 1994 (1994 and 1995). Nonparticipants' earnings are similarly defined

Table 2
Sample Means of the Outcome Measures

	Non (1)	*YP* (2)	*LMT* (3)
Earnings one year after program start (SEK)	73,750	52,110	44,120
Earnings two years after program start (SEK)	89,300	74,770	66,700
Employed within 12 months after program start (percent)	37	29	24
Employed within 24 months after program start (percent)	42	41	39
Started regular studies within 12 months after program start (percent)	11	10	5
Started regular studies within 24 months after program start (percent)	12	13	9
Number of observations	2,024	1,657	606

Notes: SEK 100 ≈ USD 10.8 (June 2002). The low value of mean annual earnings is due to a large share of zero earnings. *Non* refers to nonparticipants; *YP* to youth practice; and *LMT* to labor market training.

using the hypothetical start of a program as described in Section IIID.[23] Table 2 shows that, in the raw data, all average outcome measures except the long-term study effect are highest for nonparticipants and lowest for participants in labor market training.

IV. Empirical Application

A. Estimation of the Propensity

The matching algorithm applied in this study is, in many respects, similar to that in Lechner (2001), and it is outlined in detail in Appendix 2.[24] The discrete choice model for estimating the propensities is a multinomial logit model with three alternatives:

$$(10) \quad Pr(T_i = l) = \frac{\exp(X_i\eta_l)}{\sum_{m=0}^{M} \exp(X_i\eta_m)},$$

where m indexes the choice, and i the individual. X is a vector of covariates. The choice alternatives are no treatment ($T = 0$), youth practice ($T = 1$), and labor market training ($T = 2$) and thus, $M = 2$.[25] To test the assumption of the independence of irrelevant alternatives (IIA) underlying the multinomial logit model, I estimate binomial logit models for all three comparisons: (0,1); (0,2); and (1,2). The estimated coefficients of the binomial and multinomial models are similar and, thus, the IIA assumption is considered to be sufficiently valid.[26]

The results in Table A4 in the Appendix show that the statistical significance of various explanatory variables differs across the two programs. However, the variables for preprogram unemployment history, as well as those from the local employment offices seem to be highly significant in general. It is shown in Section V that they play an important role for the results: Excluding them from the propensity estimation would significantly alter the results.

23. I also estimated the treatment effects using earnings for the subsequent calendar year for all participants independent of the starting date of the program. As expected, the results for ''earnings one year after the program start'' are clearly more negative. For example, the effect of youth practice on participants is 10 percentage points lower than the effect reported in Table 3. The relative effectiveness of the programs, however, is not affected to any large extent by the different definitions of the outcome variable, nor by the effect after two years, which indicates that the earnings effect is stabilized quite rapidly after a program ends. The results can be obtained from the author on request.
24. Heckman, Ichimura, and Todd (1998) suggest other possible estimators. Estimators based on nonparametric kernel regressions have somewhat better asymptotic properties, whereas the main advantage of the estimators suggested by Lechner (2001) is their computational simplicity.
25. The specification of the multinomial logit is based on likelihood-ratio tests for omitted variables in a binary framework.
26. To be more exact, the Hausman test (see, for example, Chapter 9 in Greene 1993) could be applied to check whether the estimated coefficients differ significantly from each other. Nevertheless, matching based on the predicted probabilities from the binomial logit framework produces results similar to those in Tables 3–4. These results can be obtained from the author on request.

The predictive power of the model is reported in Table A5, and I consider it satisfactory: Approximately 60 percent of the observations are predicted correctly when the highest of the propensities determines the prediction. At least 70 percent of the observations in the subsamples of nonparticipants and youth practice participants are correctly predicted. Outcomes in the smallest subsample of labor market training, though, are merely predicted correctly in 7 percent of the cases.[27] However, the crucial outcome of interest is the match quality produced by the model, discussed in the next subsection.

A correct estimation of the average treatment effects, θ_0^{ml} and γ_0^{ml}, requires common support for the treatment and the comparison group, or $0 < P^m(x) < 1$ for all $m = 0, 1, \ldots, M$. In practice, this implies that some of the observations are excluded from the sample, if the propensity distributions do not cover exactly the same interval. In other words, an observation in the subsample m with an (estimated) propensity vector equal to $\{p_1^*, p_2^*, \ldots, p_M^*\}$ was excluded from the sample if any of these propensities was outside the distribution of that specific propensity in any of the other subsamples l.[28] Due to this common support requirement, approximately 200 observations were deleted, leaving a sample size of 4,084.

B. Matching

In the binary case of two treatments, the subsample of nonparticipants generally consists of a large number of observations, and it is thus plausible to use each comparison unit only once. This is not meaningful in the multiple case, since pair-wise comparisons were made across all subsamples, and for some comparisons, the potential comparison group is much smaller than the treatment group. Thus, matching was done with replacement, whereby each comparison unit was allowed to be used more than once, given that it was the nearest match for several treated units. The covariance matrix for the estimates of average effects, proposed in Lechner (2001), considers the risk of ''over-using'' some of the comparison units: The more times each comparison is used, the larger is the standard error of the estimated average effect.

A detailed description of the matching algorithm is outlined in Appendix 2. The pair-wise matching procedure was carried through six times altogether. Each individual in the treated subsample m was matched with a comparison in subsample l. The criteria for finding the nearest possible match was to minimize the Mahalanobis distance of $[P^m(X), P^l(X)]$ between the two units.

Furthermore, covariates in the matched samples ought to be balanced according to the condition $X \amalg T|b(X)$, referred to as the balance of the covariates. Following Lechner (2001), the match quality is judged by the mean absolute standardized biases of the covariates. The results show that the covariates are sufficiently balanced by the reported model specification.

27. The distributions of the predicted propensities also should be considered. In a broad outline, a good model produces large differences in the mean of predicted propensities across the various groups. This is the case for propensities to participate in youth practice and to not participate in any program, whereas the distributions of propensities to participate in labor market training look very similar. Once again, this may be a result of the small size of this subsample compared to the other subsamples.

28. This procedure assumes that there are no gaps in the empirical distributions, which is the case here.

C. Results

Aggregating the pair-wise differences over the common support yields an estimate of the average treatment effects on the treated, θ_0^{ml}. Average treatment effects on the population, γ_0^{ml}, are obtained by taking weighted sums of the treatment effects on the treated.[29] The exact expressions for θ_0^{ml} and γ_0^{ml} are found in Lechner (2001).

1. Average Treatment Effect on the Treated

Table 3 reports the effects of the six different treatments on the treated effects. Each estimated effect is reported in both absolute and relative terms. By presenting the absolute size of the effects, it is possible to compare the magnitude of the effects between the treated and the nontreated. The relative effects indicate the extent of the magnitude of the effect and help to explain how the results are changed due to the sensitivity analysis in Section V.

First, let us compare the programs to the state of no participation shown in the first four columns. Columns 1 and 3 report the program effects on program participants, as compared to nonparticipation, whereas the *potential* effects on those who did not participate in any program are shown in Columns 2 and 4. The last two columns report the effects of youth practice as compared to training, first on participants in practice and then in training.

In general, there is little heterogeneity between the groups; for example, the effects of youth practice compared to nonparticipation are roughly the same for participants and nonparticipants. The short-term effects on both earnings and employment are significantly negative for both programs and all groups throughout.[30] However, after two years from the start of a program, they are more positive and the only significantly negative results are found for the effects of labor market training on earnings. Youth practice does not seem to have any effect on the probability of entering education, whereas participation in labor market training would have significantly decreased the study probability of nonparticipants, as shown in Column 4.

A comparison of the two programs indicates that practice was better than training for those actually participating in it in terms of all outcome measures. All effects reported in Column 5 are statistically significant and positive except for the long-term employment effect. For the group of participants in labor market training, the difference between the programs seems to be less significant, although in the same direction, as for youth practice participants.

29. The weights for calculating the average population effect of treatment m compared to treatment l are based on the number of times each unit is used in all comparisons, that is, not only the comparisons between treatments m and l. Consequently, the average population effect may differ quite considerably from the average of the treatment effects on the treated, $(\theta_0^{ml} + (-\theta_0^{lm}))/2$.

30. Recall that, in practice, short term also refers to the time *after* the end of a program. The immediate earnings effect during participation may well be positive since compensation is received while participating. Individuals entitled to UI benefits receive compensation equal to the UI and can therefore not gain from participation, but individuals not entitled to UI receive either nothing or some supplementary benefit as openly unemployed. Thus, for them, the compensation of SEK 338 per working day when participating does presumably exceed income as openly unemployed.

Table 3
Results for the Mean Treatment Effect on the Treated: $\theta_0^{ml} = E(Y^m|T = m) - E(Y^l|T = m)$, *Expressed in Absolute Terms*

	YP–Non (1)	*Non–YP* (2)	*LMT–Non* (3)	*Non–LMT* (4)	*YP–LMT* (5)	*LMT–YP* (6)
Earnings one year after program start (SEK)	**−14,565**	**16,380**	**−23,440**	**27,100**	**15,560**	−6,690
	(−3.82)	**(4.03)**	**(−5.22)**	**(7.16)**	**(3.92)**	(−1.50)
	−22%	***29%***	***−35%***	***59%***	***42%***	*−13%*
Earnings two years after program start (SEK)	−3,330	5,060	**−14,080**	**14,450**	11,480	−2,170
	(−0.50)	(0.76)	**(−2.33)**	**(2.22)**	(1.45)	(−0.34)
	−4%	*6%*	***−17%***	***20%***	*18%*	*−3%*
Employment within 12 months after program start (percentage points)	**−0.07**	**0.10**	**−0.10**	**0.11**	**0.06**	−0.01
	(−2.46)	**(3.25)**	**(−3.30)**	**(3.77)**	**(2.03)**	(−0.16)
	−18%	***37%***	***−30%***	***41%***	***27%***	*−2%*
Employment within 24 months after program start (percentage points)	0.02	0.03	−0.01	−0.02	0.03	0.00
	(0.82)	(1.00)	(−0.31)	(−0.64)	(0.71)	(−0.05)
	6%	*8%*	*−3%*	*−5%*	*6%*	*0%*
Studies within 12 months after program start (percentage points)	−0.01	0.00	−0.03	**0.06**	**0.06**	**−0.04**
	(−0.42)	(0.05)	(−1.69)	**(3.77)**	**(3.20)**	**(−2.00)**
	−7%	*1%*	*−33%*	***102%***	***127%***	***−42%***
Studies within 24 months after program start (percentage points)	0.01	−0.00	0	**0.05**	**0.04**	−0.02
	(0.64)	(0.21)	(0)	**(2.40)**	**(2.06)**	(0.96)
	10%	*−4%*	*0%*	***62%***	***51%***	*−20%*
Number of observations*	1,592–711	1,912–722	580–439	1,852–459	1,592–425	580–388

Notes: **Bold type** indicates statistical significance at the 5 percent level. t-values in parentheses, *relative effects in italics*. For a description of the dependent variables, see Figure 2. *The number of observed earnings two years after program start is somewhat lower than for the other outcome variables.

Table 4
Results for the Average Treatment Effect on the Population: $\gamma_0^{ml} = E(Y^m - Y^l) = EY^m - EY^l$, Expressed in Absolute Terms

	YP–Non (1)	*LMT–Non* (2)	*YP–LMT* (3)
Earnings one year after program start (SEK)	**−15,740**	**−27,760**	**12,020**
	(−4.12)	**(−7.46)**	**(3.73)**
	−23%	***−39%***	***30%***
Earnings two years after program start (SEK)	−2,320	2,900	−5,220
	(−0.49)	(0.64)	(−1.36)
	−3%	*5%*	*−7%*
Employment within 12 months after program start (percentage points)	**−0.09**	**−0.12**	0.03
	(−3.00)	**(−4.34)**	(1.26)
	−23%	***−33%***	*11%*
Employment within 24 months after program start (percentage points)	−0.01	0.01	−0.02
	(−0.26)	(0.24)	(−0.62)
	−2%	*3%*	*−5%*
Studies within 12 months after program start (percentage points)	0.00	**−0.06**	**0.06**
	(0.10)	**(−3.75)**	**(3.56)**
	0%	***−50%***	***150%***
Studies within 24 months after program start (percentage points)	0.01	−0.03	**0.04**
	(0.51)	(−1.93)	**(2.64)**
	8%	*−25%*	***44%***

See notes to Table 3.

2. Average Treatment Effect on the Population

Table 4 reports the estimated average treatment effects on the population. These results confirm the impression given by Table 3: In the short run, both programs result in lower earnings, as well as a lower probability of employment, compared to the outcome without any program. Similar to the treatment-on-the-treated results, the negative effects more or less disappear in the course of time. Youth practice has no effect on the probability of studies, while the effect of labor market training is significantly negative.

All in all, youth practice seems to have been ''less harmful'' than labor market training, except for the effect on employment probability, where the difference is statistically insignificant.[31]

V. Heterogeneity and Sensitivity Analysis

Let us now examine the robustness of the results reported in Section IV. First, the sensitivity of the results to the availability of the covariates is explored.

31. The result that subsidized employment is relatively more effective than training is supported by other Swedish studies, see among others Carling & Richardson (2001).

Second, I examine heterogeneity among various types of individuals, and between various types of labor market training. Third, the definition of the outcome variables is changed in order to examine whether the negative program effects could be a result of declining search activity during participation in a program.

A. Availability of the Covariates

Preprogram earnings and unemployment, local employment office variables, and education and experience were excluded one by one from the propensity estimation in order to check the sensitivity of the results to the availability of these suggested key covariates. As an example, Table 5 shows the changes in the short-term effects of youth practice on participants.

The results are indeed sensitive to a reduction in information. The initially strong negative earnings and employment effects of youth practice as compared to nonparticipation become less negative when any of the covariates are excluded. Note, however, that the unadjusted differences are more negative than the initial estimates obtained by matching on all covariates. Given that our main model is correctly specified, this suggests that excluding some of the key covariates may sometimes be worse than excluding all of them.

The results further indicate that the importance of a covariate depends on the comparison group. Preprogram unemployment is an example: If it is excluded, the employment and earnings effects of youth practice become less negative when compared to nonparticipation, but less positive when compared to training. The covariates also play a different role for different outcome variables. For instance, controlling for education and experience seems to be important when examining the employment effect, but less so for the earnings effect.

Information on the relative program magnitude at the local employment office is always essential when measuring the effect of youth practice. Excluding preprogram unemployment also has an impact on most of the estimates. Moreover, these two variables are important for the estimated effect of labor market training on participants.[32]

B. Heterogeneity among Individuals

I have examined the variation in the estimated effects (i) between sexes, (ii) among the cohorts of program participants, and (iii) among individuals with various propensities to participate in the programs. In short, there is some heterogeneity in all respects.

The programs generally seem to have been slightly better for women than for men. The earnings effects are more or less the same for both sexes, whereas the effects on both study and employment probability differ significantly. This holds for the effects on the treated as well as for population effects. Both programs, but labor market training in particular, have more negative short-term effects on employment for men than for women. Youth practice seems to be superior to, or at least as good as, labor market training for both sexes in all respects.

32. Comprehensive results for all effects can be obtained from the author on request.

Table 5

Sensitivity to the Availability of Covariates; Average Effects of Youth Practice on its Participants

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
			Youth practice—nonparticipation			
Earnings one year after program start (SEK)	**−14,565**	**−10,900**	**−8,780**	**−9,000**	**−13,390**	**−21,640**
	(−3.82)	**(−3.04)**	**(−3.81)**	**(−3.29)**	**(−4.99)**	**(10.0)**
Employment within 12 months after program start (percentage points)	**−0.07**	−0.04	**−0.06**	−0.03	−0.02	**−0.08**
	(−2.46)	(−1.45)	**(−2.74)**	(−1.02)	(−0.57)	**(−5.16)**
Studies within 12 months after program start (percentage points)	−0.01	−0.02	−0.02	**−0.04**	−0.03	−0.01
	(−0.42)	(−1.30)	(−1.53)	**(−2.29)**	(−1.59)	(−0.97)
			Youth practice—labor market training			
Earnings one year after program start (SEK)	**15,560**	**13,950**	**8,530**	**11,980**	**15,270**	**8,000**
	(3.92)	**(3.37)**	**(2.72)**	**(3.68)**	**(4.39)**	**(2.85)**
Employment within 12 months after program start (percentage points)	**0.06**	**0.07**	0.04	0.04	0.06	**0,05**
	(2.03)	**(2.30)**	(1.48)	(1.41)	(1.85)	**(2.41)**
Studies within 12 months after program start (percentage points)	**0.06**	**0.05**	**0.04**	0.02	**0.06**	**0.05**
	(3.20)	**(2.74)**	**(2.08)**	(1.32)	**(3.66)**	**(4.20)**

Notes: **Bold type** indicates statistical significance at the 5 percent level. *t*-values in parentheses. (i) All variables included (main model); (ii) preprogram earnings excluded; (iii) preprogram unemployment excluded; (iv) local employment office variables excluded; (v) education and experience excluded; (vi) unadjusted differences.

The state of the business cycle also has an impact. As shown in Table A1 in the Appendix, the dates for the start of a program (or the hypothetical start of a program) vary considerably among the individuals in the three subsamples. In the analysis in Section IV, I did not consider time variation—that is, the fact that ‘‘one year after the start of a program’’ may imply early 1993 for one individual and early 1995 for another. If labor demand or study opportunities vary over the period, the results may be influenced by the systematic difference in registration dates among the samples.

Besides correcting for a potential bias, an analysis in which participant groups are divided into subgroups by the year of the start of a program may also reveal heterogeneity in the treatment effects among various cohorts of participants. In fact, there turns out to be a considerable amount of variation between the subgroups. The earnings and employment effects of the programs are the least favorable for those who enrolled in a program in 1992, then gradually improve for the latter cohorts of 1993 and 1994. Regarding the study effects, labor market training seems to have a clearly negative effect on the cohorts of 1992 and 1993. For the latest cohort, the employment and study effects of both programs compared to nonparticipation are estimated to be positive, though mainly statistically insignificant.

Heterogeneity was also examined with respect to the propensity of a treatment. A positive correlation between the propensity of a treatment and the treatment effect would indicate that the criteria for assignment are correct. Consequently, a negative correlation, or no correlation at all, implies that the selection rules are not optimal. Plotting the differences in earnings and study probability for each matched pair against the propensity of the treatment indeed reveals a great deal of variation, but no correlation. The effect of labor market training on employment as compared to nonparticipation seems to be slightly more positive, the higher is the propensity of labor market training. In general, however, the average effects of the programs compared to nonparticipation and to each other appear to be approximately the same for individuals likely and not likely to be selected into a program, respectively. This, in turn, may imply a nonoptimal selection criteria.

C. Heterogeneity between Various Types of Labor Market Training

Labor market training is a relatively heterogeneous program consisting of courses of various content and length. In a broad outline, the courses are divided into vocational and nonvocational categories that are often preparatory in the sense that participants already have ex ante plans to participate in further programs. An example of such courses is Swedish for immigrants. Consequently, participants in these courses are not expected to de-register from the Employment Service as quickly as participants in vocational courses or other programs. The effects of nonvocational courses may thus be less advantageous than those of vocational courses.

Approximately 34 percent of the labor market training participants took a nonvocational course.[33] To examine whether the effects differ between the two types of training, I applied an analysis where vocational and nonvocational courses were

33. In the sample of 606 labor market training participants used in this study, 518 observations include information on the type of course.

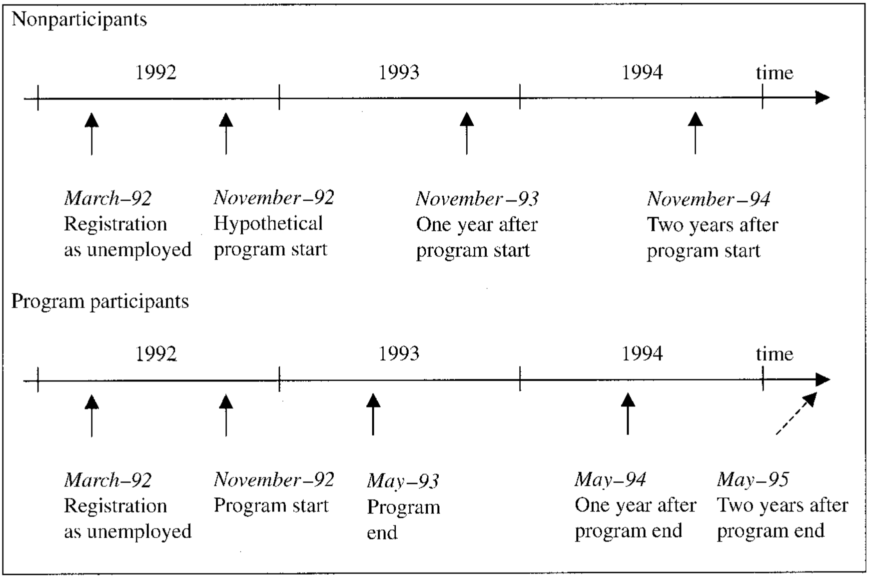

Figure 3
Outcome Measures When a Program Period is Excluded

treated as separate programs. In short, the results show that the type of training has a relatively small effect. The estimated earnings and employment effects of vocational training are only marginally higher (less negative) than those of nonvocational training. Hence, the strongly negative average effects of labor market training remain robust, even when the various types of training are considered.

D. Definition of the Outcome Variables

The analysis in Section IV is based on the assumption that individuals who participated in the programs continued their job search during the program, as required by the program regulations. Thus, the program period is included in the outcome measures of the participants. However, in practice, search activity may diminish considerably during participation in a program. (For evidence, see Ackum Agell 1996; or Edin and Holmlund 1991.) Thus, it might be argued that the program period should be excluded when defining the outcome variables, as illustrated in Figure 3.[34]

As before, the time span "within one or two years after" begins with the hypothetical start of a program for nonparticipants, while for program participants, it instead begins *at the end of a program*. In this analysis, earnings after the start of a program are defined as follows. For all program participants with a program *end*, and for all

34. Counting the time from the end of a program instead of the start may, however, imply an endogeneity bias, since the length of participation is not necessarily exogenous.

nonparticipants with a hypothetical program *start* during, say, 1992, earnings one year after the *end/start* of the program are the annual sum of earnings for the calendar year 1993. Consequently, more positive effects of the programs as compared to the state of no participation would be expected. Moreover, since the average participation period in labor market training is shorter than the average period in youth practice, more positive average effects of practice as compared to training would also be expected.

The results are more or less as anticipated: The earnings and employment effects of both programs are diminished, whereas the study effects are more or less unchanged compared to the effects in Section IV. The effects of youth practice on participants are, in fact, estimated to be slightly positive, though statistically insignificant. Labor market training seems to have negative effects even when the program period is excluded, however. When estimated for the whole population, all three short-run effects are statistically significant and negative. All in all, it seems that the negative effects of youth practice presented in Section IV may be explained by declining search activity among participants during the program, whereas further explanations are needed to account for the deleterious effects of labor market training.

VI. Discussion on Identification

The fundamental problem of an evaluator is to choose the right estimator. The decision should be based on available data and the design of the program(s), but, in the end, it will always be subjective. As Heckman, LaLonde, and Smith (1999) formulated it, ''there is no magic bullet.''

In this study, I have based the analysis on the conditional independence assumption, according to which the data provide information on all factors affecting selection as well as the outcome. This is a strong assumption which, as is always the case with identifying assumptions, cannot be tested directly. However, an indirect way of testing its plausibility, suggested by among others Heckman and Hotz (1989), is to apply the matching estimator to the outcome variables prior to the program period. According to such a test, an insignificant difference in the preprogram outcomes between two groups provides support for conditional independence.

However, as pointed out earlier, all available information on preprogram labor market history should be included in estimation of the propensities. Once this is done, application of a preprogram outcome test is not meaningful, since the matching procedure (applied correctly) implies that the preprogram outcome variables are balanced across the samples. That is also the case in this study. Preprogram earnings are included in the estimation of the propensities, and the results for balance of the covariates show that the differences in mean preprogram earnings are negligible among the three samples.[35]

35. Exact information on employment and study spells, in turn, are not included in the data. However, the samples are constructed so that all individuals were unemployed immediately prior to the start of a program, which implies that the share of employed or students at this point in time is the same—zero—in all three samples. Moreover, the length of preprogram unemployment is used for matching and, once again, the test for balance of the covariates shows insignificant differences in this variable between groups.

Another idea for testing the plausibility of the identifying assumption, or at least the robustness of the results, could be to apply various estimators to the same problem to see whether the results differ. It is not obvious how the results of such comparisons should be interpreted, however. Let us say that all methods produce similar estimates for the program effect. What, then, does this say about the validity of the identifying assumptions underlying the various estimators? It might imply that there is no selection at all, a viewpoint supported by, for example, Heckman, LaLonde, and Smith (1999). If so, it would nevertheless convince the evaluator that the estimated program effect is the true one.

Another, perhaps more pessimistic way of interpreting such results is to argue that they only imply that some or all identifying assumptions are invalid without pointing out which ones. According to this view, different estimators *should* produce various results since they are based on various identifying assumptions, and something is wrong—but we do not know what—if they produce the same results.

Bearing these alternative interpretations in mind, I compared the results obtained by matching with some alternative, well-established estimators.[36] The first approach involves the standard OLS regression for the continuous dependent variable, and a probit model for the discrete dependent variables. As in the matching approach, identification of the average treatment effects in these models requires conditional independence. Moreover, the estimators are based on further parametric restrictions.

The results are reported in Table A6 in the Appendix. The set of covariates included in the OLS and the probit estimations are the same as those used to estimate the propensity scores. A comparison of Table A6 with Table 3 shows that, in this specific case, OLS and probit on the one hand, and matching on the other produce fairly similar estimates of the average treatment effects on the population. But this is not very surprising since identification is based on the same assumption.

One substantial difference compared to the results obtained by matching, however, is an improvement in the employment effects of youth practice. Table A6 reports a practically zero short-term effect and a significantly positive effect in the long run, whereas the effects obtained from the matching framework are clearly more negative. Consequently, the difference between the employment effects of practice and training is more obvious in Table A6. Moreover, the long-term earnings effect of labor market training is estimated to be significantly negative by OLS, whereas matching obtains a zero effect. These differences are presumably explained by the parametric restrictions underlying the OLS and probit estimations. Matching allows for heterogeneity in the treatment effects in a more flexible way.

A second approach applied to the continuous dependent variable—earnings—is a multinomial generalization of the classical Heckman two-stage model presented by Lee (1983) and called a polychotomous selectivity model. The Lee model is similar to other selectivity models in that it is designed to adjust for both observed and unobserved selection bias. Thus, it does not require the conditional independence assumption to be valid. However, it rests on other strong assumptions, among them linearity in the outcome variable and joint normality in the error terms.

The results are shown in Table A7. The multinomial logit model underlying the inverse Mill's ratios is exactly the same as the one used to estimate the propensity

36. A more extensive account of these approaches and the results is available from the author on request.

scores. The local employment office variables are now only assumed to affect the selection into programs but not earnings and are thus excluded from the earnings equation.

The results show fewer negative effects of the programs than matching. The difference between the effects of practice and training is also diminished. The long-term effect of labor market training is estimated to be less favorable than in the short run. Drawing conclusions about the existence of unobserved heterogeneity is not straightforward, however, because the standard errors for the parameter estimates for selection adjustment terms are very large. The precision of the estimates of the treatment effects is also low. In sum, the results seem to suggest that there may be some unobserved heterogeneity between the samples that implies a (moderate) negative bias in the estimated program effects presented in previous sections of this paper, but the evidence is not unequivocal.

VII. Conclusions

The purpose of this study has been to evaluate labor market programs for youth in Sweden using three measures of effectiveness: post-program annual earnings, employment probability, and probability of entering regular education. More precisely, the programs evaluated are youth practice and labor market training. The age group examined is 20–24.

Identification of the average treatment effects is based on the *conditional independence assumption* (CIA), whereby participation in the various treatments, including the no-treatment state, is independent of the post-program outcomes conditional on observable exogenous factors. The results from the main analysis suggest that both youth practice and labor market training have negative short-term effects on earnings and employment, where ''short-term'' refers to one year after the start of a program. Two years after the start of a program, however, the effects are no longer as obvious; most estimates for employment and earnings are statistically insignificant at the 5 percent level. As regards the third measure of effectiveness, the probability of regular education, the results show no significant effects of youth practice, whereas labor market training may have had a negative effect, at least in the short run. Finally, a comparison of the two programs suggests that practice was better—or ''less harmful''—than training.

How robust are these results? Beginning with the question of identification, neither the preprogram outcome tests nor the comparison with results from other methods seem to give any reason to seriously doubt the plausibility of the CIA in this context. The traditional two-stage selectivity model does indeed yield somewhat different results for both programs than matching, at least in the short run, but the point estimates are nevertheless negative. Moreover, drawing conclusions from such methodological comparisons is not straightforward. Because no direct test for the fundamental identifying assumptions is available, it is ultimately up to the reader to judge the results by weighing in the institutional setting and the available data.

Sensitivity analysis in the matching framework confirms the presumed importance of controlling for preprogram earnings and unemployment, as well as education and experience in the propensity estimation: Excluding any of these variables changes the estimated program effects, which generally become more positive. As concerns the choice of model, matching on the conditional propensities obtained from binomial logit estimations yields results very similar to those obtained by the multinomial logit model.

The effects are shown to be heterogeneous for various types of individuals, however. The effects are more favorable—less negative—for women than for men, and the effect of labor market training on earnings and employment seems to have been somewhat less negative for those who took a vocational course than for participants in nonvocational courses of a more preparatory nature.

An attempt to control for variation with respect to the business cycle suggests an additional source of heterogeneity. The results from separate analyses of the individuals enrolling in the programs during 1992, 1993, and 1994 show that the effects are more positive, the later the start of a program, and thus the better the business cycle.

Hence, the results from the sensitivity and heterogeneity analyses suggest that in the total sample of 4,000 individuals, there are subsamples for which the effects are not as negative as they are for the aggregate, on average. Moreover, a plausible explanation, provided by the sensitivity analysis, for the negative or nonexistent earnings and employment effects of youth practice is that participants put less or no effort into finding a job during the program, despite regulations requiring active job search. This hypothesis is also supported by the fact that already after two years the effects exhibit quite remarkable improvements.

All in all, neither of the youth programs seems to work as intended. In an international perspective, this is not surprising. Surveys on existing evaluation studies by Martin (1998) and Heckman, LaLonde, and Smith (1999) show that most of the OECD countries have failed in active labor market programs for the youth. What is the reason for these poor effects?

The results for youth practice might be explained by insufficient planning and followup, as pointed out in several implementation studies, as well as by low-qualified tasks that did not provide any human capital accumulation. Moreover, the results from the analysis of business-cycle variation may suggest that these problems were more severe when the program was relatively new. Given that search activity was very low during program participation, it seems to be more or less expected that the effect did not turn out to be positive.

An explanation for the negative results of labor market training requires more than what is suggested for youth practice, however. The program has existed for decades, and thus ''start-up problems'' are not the answer. Furthermore, excluding the program period still produces significantly negative effects. One potential explanation is that the courses simply do not fit the employers' requirements for labor, and that training thus has both professional and regional ''lock-in'' effects on participants.

What policy conclusion can be drawn from these results? To find the answer, recall the interpretation of the nontreatment state described in Section III. The institutional setting in Sweden implies that basically all the unemployed are assigned to labor market programs, given that they are unemployed long enough. Consequently, the group of nonparticipants collected from the database *does not* represent a world without active labor market programs; when deciding not to participate these individuals know that they can—and probably will—enter a program at a later stage.

Thus, *it is incorrect to draw the conclusion that participants would have been better off had there been no programs at all*. Instead, my results suggest that it was better to wait and postpone the decision to participate.[37] The results may also be interpreted as a good mark for the local employment offices' job-seeking service for the openly unemployed. Moreover, they suggest that workplace practice is more effective than pure training, a result also found in several other Swedish studies.

37. The timing of programs, in the sense of whether it is better to participate early or late in an unemployment spell, is an interesting field for a future study. An example, along with a detailed discussion of the problems in identifying the no-treatment state, is provided in Sianesi (2002). Unfortunately, the data available for my study do not provide enough information to convince me that identifying such effects is possible.

Appendix 1

Descriptive Statistics and Estimation Results

Table A1
Descriptive Statistics of Registration Records and Preprogram Characteristics

	Non (1)	YP (2)	LMT (3)
Registration with ES			
Mean	Nov–92	Dec–92	July–92
Median	Nov–92	Nov–92	May–92
Assigned/true duration of preprogram unemployment in days (mean)	67.6	121.5	112.6
Preprogram unemployment at least four months (percent)	16.3	42.4	35.8
Preprogram unemployment at least 270 days (percent)	0.3	6.8	8.1
Annual earnings one year before registration (mean)	74,700	50,900	70,400
Assigned/true program start			
Mean	Feb–93	April–93	Nov–92
Median	Jan–93	March–93	Sept–92
Duration of program in days (mean)	—	146.6	131.3
Number of observations	2,024	1,657	606

Notes: Program start of nonparticipants is a hypothetical date randomly assigned by the procedure described above. Duration of preprogram unemployment of nonparticipants is based on this hypothetical date. SEK 100 ≈ USD 10.8 (June 2002).

Table A2
Descriptive Statistics of Selected Individual Characteristics

	Non (1)	YP (2)	LMT (3)
Age (mean)	22.75	21.46	22.38
Female (percent)	44	44	37
Non-Nordic (percent)	4	5	13
Regional characteristics (percent)			
Forest county	21	21	26
City county	41	29	35
Other county	39	50	39
Education (percent)			
Compulsory	14	12	18
High school 1–2 years	41	41	40
High school 3–4 years	31	39	34
University	14	9	7
Specific education[a] (percent)			
No	42	51	52
Yes	58	49	48
Experience[a] (percent)			
None	34	45	40
Some	32	35	34
Good	34	21	26
Number of observations	2,024	1,657	606

Notes: Age is an approximation for the age when registered with the Employment Service as openly unemployed. It is calculated as the difference between the year of registration and year of birth (precise data on dates of birth were unavailable). *Compulsory education* also includes individuals with less than the legally required 9–10 years. *High school education* is divided into two groups depending on duration.

a. *Specific education* and *experience* refer to the qualifications required for the job applied for, with the variables based on information given by job seekers when entering the Employment Service records. For individuals who have applied for several jobs, and thus have reported various levels of education and experience, I have used the observation with the highest level of experience. Information on both education and experience is *missing* for approximately 16.1 percent of the complete sample.

Table A3

Descriptive Statistics from Local Employment Offices, Expressed as Deviations from the Contemporary Country Mean (Percentage Points)

	Non (1)	YP (2)	LMT (3)
Share of program participants of all registered unemployed	−1.84	−0.77	−0.30
Share of youth practice of all program participants	−0.60	1.28	−0.99
Share of labor market training of all program participants	−0.05	−1.27	1.48
Number of observations	2,024	1,657	606

Notes: The figures in the table were calculated as follows. For each local employment office and each month, I calculated the three various ''share of *something*'' variables. Next, I took the difference from the country mean in the same month. I then took the mean of these deviations for each of the three groups. Thus, −1.84 in the first row of Column 1 shows that the local employment offices of nonparticipants were less inclined to assign individuals to programs than all offices in the country on average.

Table A4
Results from the Multinomial Logit Estimations

	Youth Practice			Labor Market Training		
	Coefficient (1)	Standard Error (2)	*RRR* (3)	Coefficient (4)	Standard Error (5)	*RRR* (6)
Constant	−57.7	8.45	—	0.17	9.81	—
Personal characteristics						
Female	0.15	0.08	1.17	−0.14	0.10	0.87
Age	5.59	0.76	2.68	−0.04	0.88	0.96
Age^2	−0.14	0.02	0.87	−0.00	0.02	1.00
Non-Nordic	0.24	0.18	1.27	1.22	0.18	3.38
Regional characteristics						
Forest county	−0.14	0.11	0.87	0.35	0.13	1.42
City county	−0.61	0.09	0.54	−0.18	0.12	0.83
Education[a]						
High school 1–2 years	0.28	0.13	1.33	−0.16	0.15	0.86
High school 3–4 years	0.23	0.13	1.25	−0.07	0.16	0.93
University	0.18	0.18	1.19	−0.52	0.23	0.60
Specific education[2]						
Yes	−0.27	0.09	0.76	−0.15	0.12	0.86
Missing	−0.19	0.14	0.82	0.03	0.18	1.03
Experience[b]						
Some	−0.11	0.11	0.90	0.00	0.14	1.00
Good	−0.37	0.12	0.69	−0.40	0.16	0.67
Preprogram labor market status						
Duration of preprogram unemployment (days)	0.01	0.00	1.01	0.01	0.00	1.01
Earnings one year before registration (in SEK 10,000)	−0.04	0.01	0.96	0.00	0.01	1.00
Local employment office variables[c]						
Share of program participation of all registered unemployed	1.94	0.47	6.98	2.57	0.59	13.0
YP of all program participation	0.88	0.37	2.40	0.58	0.47	1.79
LMT of all program participation	0.77	0.21	0.93	1.28	0.59	3.58
Missing	0.77	0.21	2.16	0.90	0.23	2.45
Log likelihood: −3,559.3, LR chi2 (38): 1440.9, Pseudo R2: 0.1683						

Notes: Nonparticipants are used as the reference category. Columns 1 and 4 report coefficients β^{YP} and β^{LMT}, and Columns 2 and 5 show the standard errors of the estimated coefficients. Bold type indicates statistical significance at the 5 percent level. Relative risk ratios (*RRR*) in Columns 3 and 6 report the exponentiated value of the coefficient, $\exp(\beta^{YP})$. It is interpreted as the relative probability (or risk) ratio for a one-unit change in the corresponding variable, when risk is measured as the risk of the category relative to the reference category. *Age* is an approximation for the age when registered with the Employment Service as openly unemployed.
a. Compulsory education is the reference level.
b. *Specific education* and *experience* refer to the qualifications required for the job applied for, with the variables based on information provided by job seekers when entering the Employment Service records. For individuals who have applied for several jobs, and have thus reported various levels of education and experience, I have collected the observation with the highest level of experience. The dummy variable *Missing* indicates the observations for which both *education* and *experience* are missing (approximately 16.1 percent of the complete sample). The reference level is no specific education or experience.
c. The variables from the local employment offices were computed as deviations from the contemporaneous country mean. Missing observations are set to zero, and denoted by the dummy variable *Missing* equal to one.

Table A5
Predictive Power of the Multinomial Logit Model

Predicted Outcome	True Outcome Non	YP	LMT	Total
Non	1,541	491	332	2,364
	(76.1%)	(29.6%)	(54.8%)	(55.1%)
YP	461	1,153	233	1,847
	(22.8%)	(69.6%)	(38.5%)	(43.1%)
LMT	22	13	41	76
	(1.1%)	(0.8%)	(6.8%)	(1.8%)
Total	2,024	1,657	606	4,287
	(100%)	(100%)	(100%)	(100%)

Table A6
Results from a Linear Regression/Probit Analysis

	YP–Non (1)	LMT–Non (2)	YP–LMT (3)
OLS Regression			
Earnings one year after program start (SEK)	**−10,350**	**−23,830**	**13,480**
	(−4.43)	**(−8.03)**	**(4.43)**
Earnings two years after program start (SEK)	90	**−11,680**	**11,770**
	(0.03)	**(−3.05)**	**(2.94)**
Probit			
Employment within 12 months after program start (percentage points)	−0.03	**−0.10**	**0.07**
	(−1.89)	**(−4.29)**	**(2.77)**
Employment within 24 months after program start (percentage points)	**0.04**	−0.00	0.05
	(2.32)	(−0.09)	(1.87)
Studies within 12 months after program start (percentage points)	**−0.02**	**−0.05**	**0.04**
	(−2.15)	**(−4.04)**	**(2.53)**
Studies within 24 months after program start (percentage points)	−0.01	**−0.03**	0.02
	(−0.88)	**(−2.11)**	(1.44)

Notes: **Bold type** indicates statistical significance at the 5 percent level. Results for the probit model are reported as marginal changes dF/dx. t-values in parentheses. The marginal change is defined as a change in probability due to a one-unit change in the covariate, $dProb(E = 1)/dx$ or $dProb(S = 1)/dx$. Thus, −0.01 in the last row in Column 1 should be interpreted as follows. A change in the dummy variable for youth practice from 0 to 1 implies a one percentage point decrease in the probability of entering studies within 24 months after the start of the program.

Table A7
Results from the Estimation of Lee's Selectivity Model

	Earnings One Year After Program Start	Earnings Two Years After Program Start
YP–Non	−4,310	2,040
	(−0.76)	(0.26)
LMT–Non	−12,940	−21,350
	(−0.86)	(−1.02)
YP–LMT	8,640	23,380
	(0.56)	(1.09)
Selection adjustment terms		
λ_1	−6,340	−2,050
	(−1.14)	(−0.28)
λ_2	−6,870	6,250
	(−0.73)	(0.48)

Notes: Standard errors were calculated using a White heteroscedasticity robust variance estimator. *t*-values in parentheses.

Appendix 2

Matching Algorithm

The matching algorithm, estimators, and covariance matrixes applied in this paper follow Lechner (2001). The procedure is outlined below.

1. Collect the participant samples and the largest possible sample of nonparticipants, and randomly assign the start of the program dates for nonparticipants from the distribution of participants (by month). Eliminate all nonparticipants assigned a date after their actual de-registration from open unemployment.

2. Specify and estimate a multinomial discrete choice model to obtain the (estimated) propensities $P(T=0|X)$, $P(T=1|X)$, $P(T=2|X)$. Test for omitted variables in a binomial framework. Compute the conditional probabilities $P^{m|ml}(X)$.

3. Common support: Eliminate all observations outside the defined common support.

4. Apply the following procedure to match each observation in group $T=m$ with an observation in the comparison group, $T=l$:

   (i) Choose an observation from group m, and remove it from that pool.

(ii) Find an observation in group l that is as close as possible to the one collected in step (i) in terms of predicted probabilities. The distance can be measured by a Mahalanobis distance metric. Alternatively, base the proximity on the conditional probability $P^{m|ml}(X)$. Do not remove that observation so that it can be used again.

(iii) Repeat (i) and (ii) until there is no observation left in group m.

(iv) Repeat (i)–(iii) for all combinations of m and l.

5. Test for the balance of the covariates. If the covariates are not balanced, refine the specification of the discrete choice model, and repeat steps 2–4.
6. Use the comparison groups formed in 4(iv) to compute the respective conditional expectations by the sample mean. Note that the same observation may appear several times in the sample.
7. Compute the estimates of the treatment effects using the results of Step 6, and compute their covariance matrix.

References

Ackum, Susanne. 1991. ‘‘Youth Unemployment, Labor Market Programs and Subsequent Earnings.’’ *Scandinavian Journal of Economics* 93(4):531–41.

Ackum Agell, Susanne. 1996. ‘‘Arbetslösas sökaktivitet’’ (Search Activity of the Unemployed). *Aktiv Arbetsmarknadspolitik (Active Labor Market Policy)*, SOU 1996:34. Stockholm: Fritzes.

Angrist, Joshua, Guido Imbens, and Donald Rubin. 1996. ‘‘Identification of Causal Effects Using Instrumental Variables.’’ *Journal of the American Statistical Association* 91:444–455.

Aschenfelter, Orley. 1978. ‘‘Estimating the Effect of Training Programs on Earnings.’’ *Review of Economics and Statistics* 60:47–57.

Carling, Kenneth, and Katarina Richardson. 2001. ‘‘The Relative Efficiency of Labor Market Programs: Swedish Experience from the 1990’s.’’ IFAU Working Paper 2001:2. Uppsala: Institute for Labour Market Policy Evaluation.

Dehejia, Rajeev, and Sadek Wahba. 1999. ‘‘Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs.’’ *Journal of the American Statistical Association* 94:1053–62.

Dahlberg, Matz, and Anders Forslund. 1999. ‘‘Direct Displacement Effects of Labour Market Training Programmes: The Case of Sweden.’’ IFAU Working Paper 1999:7. Uppsala: Institute for Labour Market Policy Evaluation.

Edin, Per-Anders, and Bertil Holmlund. 1991. ‘‘Unemployment, Vacancies and Labor Market Programs: Swedish Evidence.’’ In *Mismatch and Labor Mobility*, ed. Fiorella Padoa Schioppa, 405–48. Cambridge: Cambridge University Press.

Eriksson, Maria. 1997. ‘‘Comparison of Compensatory and Non-compensatory Models for Selection into Labour Market Training.’’ Umeå Economic Studies 439. Umeå: Umeå University.

Gerfin, Michael, and Michael Lechner. 2002. ‘‘Microeconometric Evaluation of the Active Labour Market Policy in Switzerland.’’ *Economic Journal*. Forthcoming.

Greene, William. 1993. *Econometric Analysis*. New Jersey: Prentice Hall.

Hallström Nils-Eric. 1994. ''Ungdomspraktikens implementering. En utvärdering av ungdomspraktikens genomförande i åtta kommuner i tre län'' (Implementation of Youth Practice. An Evaluation of Youth Practice in Eight Municipalities in Three Counties). EFA report 28. Stockholm: Ministry of Labor.

Heckman, James, and Joseph Hotz. 1989. ''Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.'' *Journal of the American Statistical Association* 84(408):862–80.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1998. ''Matching As An Econometric Evaluation Estimator.'' *Review of Economic Studies* 65:261–94.

Heckman, James, Hidehiko Ichimura, Petra Todd, and Jeffrey Smith. 1998. ''Characterizing Selection Bias Using Experimental Data.'' *Econometrica* 66(5):1017–98.

Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. ''The Economics and Econometrics of Active Labor Market Programs.'' In *Handbook of Labor Economics, Volume III*, ed. Aschenfelter Orley, and David Card, 1865–2097. Amsterdam: Elsevier Science.

Holland, Paul. 1986. ''Statistics and Causal Inference.'' *Journal of the American Statistical Association* 81:945–70, with discussion.

Hotz, Joseph, Guido Imbens, and Julie Mortimer. 1999. ''Predicting the Efficacy of Future Training Programs Using Past Experience.'' NBER Technical Working Paper, 238. Cambridge, Mass.: National Bureau of Economic Research.

Imbens, Guido. 2000. ''The Role of Propensity Score in Estimating Dose-Response Functions.'' *Biometrica* 87(3):706–10.

Korpi, Tomas. 1994. *Escaping Unemployment. Studies in the Individual Consequences of Unemployment and Labor Market Policy*. Ph.D. Thesis. Stockholm University: Swedish Institute for Social Research.

Layard, Richard, Stephen Nickell, and Richard Jackman. 1991. *Unemployment, Macroeconomic Performance and the Labor Market*. Oxford: Oxford University Press.

Lechner, Michael. 1999. ''Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification.'' *Journal of Business & Economic Statistics* 17:74–90.

———. 2001. ''Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption.'' In *Econometric Evaluation of Labour Market Policies*, ed. Lechner, Michael, and Friedhelm Pfeiffer, 43–58. Heidelberg: Physica/Springer.

Lechner, Michael, and Ruth Miquel. 2001. ''A Potential Outcome Approach to Dynamic Programme Evaluation—Part I: Identification.'' Discussion Paper 2001–07. St. Gallen: University of St. Gallen.

Lee, Lung-Fei. 1983. ''Generalized Econometric Models with Selectivity.'' *Econometrica* 51(2):507–12.

Martin, John. 1998. ''What Works among Active Labor Market Policies: Evidence from OECD Countries' Experiences.'' Labor Market and Social Policy—Occasional Papers 35. Paris: OECD.

Regnér, Håkan. 1997. *Training at the Job and Training for a New Job: Two Swedish Studies*. Ph.D. Thesis. Stockholm University: Swedish Institute for Social Research.

Rosenbaum, Paul, and Donald Rubin. 1983. ''The Central Role of the Propensity Score in Observational Studies for Causal Effects.'' *Biometrika* 70(1):41–55.

———. 1984. ''Reducing Bias in Observational Studies Using Subclassification on the Propensity Score.'' *Journal of the American Statistical Association* 79:516–24.

———. 1985. ''Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score.'' *American Statistician* 39:3–38.

Rubin, Donald. 1977. ''Assignment to Treatment Group on the Basis of a Covariate.'' *Journal of Educational Statistics* 2:1–26.

Rubin, Donald, and Neal Thomas. 1992. ‘‘Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Covariates.’’ *Biometrika* 79(4):797–809.
Schröder, Lena. 1995. ‘‘Ungdomars etablering på arbetsmarknaden— från femtiotalet till nittiotalet’’ (Establishment of Youth on the Labor Market—from the Fifties to the Ninetees). EFA report 38. Stockholm: Ministry of Labor.
Sianesi, Barbara. 2002. ‘‘Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults During the 1990s.’’ IFAU Working Paper 2002:5. Uppsala: Institute for Labour Market Policy Evaluation.