Assessing the employment effects of vocational training using a one-factor model

Andrén, Daniela; Andrén, Thomas

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Andrén, D., & Andrén, T. (2006). Assessing the employment effects of vocational training using a one-factor model. *Applied Economics*, *38*(21), 2469-2486. https://doi.org/10.1080/00036840500427577

ASSESSING THE EMPLOYMENT EFFECTS OF VOCATIONAL TRAINING USING A ONE-FACTOR MODEL

Journal:	*Applied Economics*
Manuscript ID:	APE-05-0037.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	12-Apr-2005
JEL Code:	J31 - Wage Level and Structure|Wage Differentials by Skill, Training, Occupation, etc. < J3 - Wages, Compensation, and Labor Costs < J - Labor and Demographic Economics, J38 - Public Policy < J3 - Wages, Compensation, and Labor Costs < J - Labor and Demographic Economics
Keywords:	vocational training, sorting, unobserved heterogeneity, unobserved heterogeneity, matching estimator

ASSESSING THE EMPLOYMENT EFFECTS OF VOCATIONAL TRAINING USING A ONE-FACTOR MODEL

Running title:

ASSESSING THE EMPLOYMENT EFFECTS OF VOCATIONAL TRAINING

Abstract

Matching estimators use observed variables to adjust for differences between groups to eliminate sample selection bias. When minimum relevant information is not available, matching estimates are biased. If access to data on usually unobserved factors that determine the selection process is unavailable, other estimators should be used. This study advocates the one-factor control function estimator that allows for unobserved heterogeneity with factor-loading technique. Treatment effects of vocational training in Sweden are estimated with mean and distributional parameters, and then compared with matching estimates. The results indicate that unobservables slightly increase the treatment effect for those treated.

Keywords: vocational training, sorting, unobserved heterogeneity, one-factor model, matching estimator

JEL Classification: J31, J38

I. Introduction

During the last decade, there has been an increasing international interest in active labor market programs (i.e., measures to raise employment that are directly targeted at the unemployed) among policy makers. This has resulted in a growing literature that estimates and quantifies the potential effects of those measures (see Kluve and Schmidt, 2002). This journal contains a number of contributions, using mainly micro data from European countries. For example, Bryant and Wilhite (1990) estimated the military experience and training effects on civilian wages by accounting for the length of time spent in the military and differentiating between that time and military training. Main (1991) estimated the effects of the youth training scheme (YTS) on employment probability using data from the Scottish Young People's Survey, and found that YTS may be operating as a successful manpower programme. Beenstock (1996) estimated the effect of training and the time to find a job in Israel and found that trainees exit unemployment at a much faster rate than comparable unemployed who did not receive training. Makepeace (1996) estimated a model in which the type of training undertaken is determined by the predicted lifetime earnings for each type of training, personal factors and educational attainment. Lifetime earnings have a positive effect on the allocation of individuals across training types, and there is a market for training in which the type of training undertaken responds to earnings incentives. Kraft (1998) investigated the effectiveness of labor market policy using data from Austria, France, Germany, Great Britain, Sweden and the United States and found that passive labor market policy had a negative, and active labor market measures a positive, effect on the number of persons employed. Groot and Van Den Brink (2000) found that formal work-related training increases employability in Netherlands.

Having access to Swedish data for the 1993-1997 recession period, this study aims to estimate the treatment effect of participating in a vocational training program 1993-1994 on the individuals' employment probabilities in the following year, 1995. Using a one-factor control function estimator allows us to study the heterogeneous treatment effect on discrete outcomes as a measure for the change in employment probability as a result of the treatment. Using the same set of control variables, the parameters estimated by the control function estimator are compared with the parameters estimated by the propensity-score matching estimator, as a mean to investigate the impact of controlling for unobserved factors.

In recent years, matching estimators have received substantial attention in evaluating social programs (see Heckman et al., 1997b, 1998a, and Heckman et al., 1998b). Matching estimator's uses observed variables to adjust for differences between groups that give rise to selection bias. However, when the analyst does not have access to the minimum relevant information, matching estimates are biased. Furthermore, having more information, but not all of the minimal relevant information in terms of variables, increases the bias compared to having less information (Heckman and Navarro-Lozano, 2003). Therefore, it is necessary to have access to a rich data set so that most of the usually unobserved factors that determine the selection process are observed. This is important since it is expected that unobserved factors such as aptitude and ambition are relevant components when an individual is being selected into a social program such as vocational training. If access to such data is unavailable, other estimators should be used. This paper advocates the one-factor control function estimator formulated by Aakvik et al. (2000). The one-factor model incorporates the selection process and allows unobserved factors to explain the outcome in each state as well as in the selection-process, using the factor-loading technique. Because the method of control function explicitly models omitted relevant variables rather than assumes that there are none, it is more robust to omitted conditioning variables than the matching estimator is. Furthermore, matching has the strong implicit assumption that the marginal participant in a given program gets the same return as the average participant in the same program, which makes the economic content more restrictive compared to the control function estimator. The structure of the one-factor model also makes it possible to derive both the mean and the distributional treatment parameters, where the latter parameter shows how the treatment effect is distributed. The distribution and functional form assumptions of the control function estimator are often exposed to critique (see Vella, 1998). However, the distributional assumption of the unobserved factor is easily relaxed by approximating it with a discrete point distribution (non-parametric). This allows for a comparison between the parametric and non-parametric assumptions of the non-observed factor.

The analysis of this study is done separately for the Swedish-born and the foreign-born, given that these two groups have different arrangements of characteristics, which determines the selection and treatment process. The foreign-born group is also much more heterogeneous compared to the Swedish-born group, which further emphasizes the importance of analyzing the groups separately.

The rest of the paper is organized in the following way: Section II presents the institutional settings and the main characteristics of the active labor market programs in Sweden for the analyzed period. Section III presents the econometric specifications. The data and main descriptive statistics for both treatment and control groups are presented in Section IV, and the results in Section V. Section VI summarizes the findings of the paper.

II. Institutional settings

Swedish labor market policy has two components: a (passive) *benefit system* that supports individuals while they are unemployed, and a range of (active) *labor market programs* (vocational and non-vocational) offered to improve the employment opportunities of the unemployed. The benefit system has two components: *unemployment insurance* (UI), and the *cash labor market assistance* (CA).[1] UI is the most important form; it is income-related and is available for 60 calendar weeks. The daily compensation is 75% of the previous wages (was 90% before July 1993). A part-time unemployed person registered at a public employment office and actively searching for a job is also eligible for unemployment benefits. CA was designed mainly for *new entrants* who are *not* members of any UI fund. Its compensation is *lower* than that of UI, and is paid (in principle) for a maximum of 30 calendar weeks.

The public employment offices have a central role in assigning job seekers to training courses. The employment office is responsible for providing information on different courses, eligibility rules, training stipends, etc.[2] Those eligible for training are *mainly* unemployed persons who are job seekers and persons at risk of becoming unemployed. One can also be eligible for other reasons. For example, the status of political refugee makes a foreigner eligible for training courses during the first three years in Sweden. Although there is no formal rule for the offer of labor market training being given to a person who has been unemployed for a long period, there are reasons to believe that this is often the case.[3] Since 1986, the time-period a trainee participates in a labor market program is considered equal to time spent on a regular job. Therefore, participation in a labor market program for five months counts as an *employment* spell, and thus qualifies for a renewed spell of unemployment compensation.

Originally, labor market training mainly consisted of vocational training programs. However, over time, schemes comprised of programs of a more general nature have grown more prevalent. During the 1990s, other education programs such as Swedish for immigrants and computer training were added to labor market training. This study focuses only on vocational training, which represented around 20% of all programs within active labor market policy in 1993-1994.

III. Econometric specifications

The fundamental issue of the evaluation problem is that a person is unable to be in two different labor market states at the same time. In the training context, for each trainee there is a hypothetical state of how he or she would have done without training. For each non-trainee, there is the hypothetical state of being a trainee. Our point of departure is the index sufficient latent variable model (Heckman, 1979) that postulates a standard framework of potential outcomes and a selection mechanism for the choice of state:

$$Y_1^* = X\beta_1 + U_1, \quad Y_1 = 1 \text{ if } Y_1^* \geq 0, Y_1 = 0 \text{ elsewhere,} \tag{1}$$

$$Y_0^* = X\beta_0 + U_0, \quad Y_0 = 1 \text{ if } Y_0^* \geq 0, Y_0 = 0 \text{ elsewhere,} \tag{2}$$

$$D^* = Z\beta_D + U_D, \quad D = 1 \text{ if } D^* \geq 0, D = 0 \text{ elsewhere.} \tag{3}$$

For a given individual, Y_1^* represents a latent variable for the propensity to be employed in the training state,while Y_0^* represents a latent variable for the propensity to be employed in the non-training state. X is a matrix of observed characteristics explaining the outcomes of the two potential states. Each state also has an unobserved stochastic component represented by U_1 and U_0. Equation (3) defines the selection decision, with D^* being a latent variable for the propensity to participate in a vocational training program, Z a matrix of observed characteristics and U_D a vector of unobserved components that explain the selection decision between the two states.[4] The remaining vectors β_1, β_0, and β_D are unknown parameters that are to be estimated.

Within this framework, there are two separate problems to deal with: 1) how to recover the unobserved marginal densities, $f(Y_1 \mid X)$ and $f(Y_0 \mid X)$, using information from the observed conditional densities, $f(Y_1 \mid X, D=1)$ and $f(Y_0 \mid X, D=0)$; and 2) under what conditions it is possible recover the full bivariate density, $f(Y_1, Y_0 \mid X)$, using the recovered marginal densities. We follow Aakvik et al. (2000) and deal with both of these problems using the assumption of a one-factor structure on the unobservables. The assumed factor structure is unobserved and needs further assumptions regarding its distribution. We consider two frequently used distributions: the continuous normal distribution and the discrete mass-points distribution, which will be discussed in the following sections.

The one-factor assumption is based on the idea that for a particular individual there is some unobserved factor out there that is common to the two states, as well as to the selection mechanism. It could be ambition, motivation, or some other idiosyncratic quality that is important both when searching for a job and when being selected into a program. With this common factor, it is possible to connect the training state, the non-training state as well as the selection into the states, and thereby being able to recover the full unconditional distribution for the problem. This is of special interest since the full distribution may be used to answer several important policy-oriented questions.

A. The normal one-factor model

The one-factor model makes specific assumptions about the structure of the unobservables. The assumed error terms in equations (1)-(3) are defined and decomposed in the following way:

$$U_1 = \rho_1 \xi + \varepsilon_1,\ U_0 = \rho_0 \xi + \varepsilon_0,\ U_D = \rho_D \xi + \varepsilon_D, \tag{4}$$

where ξ constitutes the common unobserved "ability" factor and ρ_i, $(i = 1, 0, D)$, the factor loadings, unique for each equation.

The factor structure assumption for discrete choice models was introduced in Heckman (1981) and produces a flexible yet parsimonious specification, while making it possible to estimate the model in a tractable fashion. The following normality assumption is imposed: $(\xi, \varepsilon_1, \varepsilon_0, \varepsilon_D) \sim N(0, I)$, where I is the identity matrix. This implies that $(U_1, U_0, U_D) \sim N(0, \Sigma)$, with all components in the covariance matrix, Σ, recovered by the factor loadings, and normalizations made by the normality assumption. Conditioning on ξ, the likelihood function for the one-factor model has the form:

$$L = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \Pr(D_i, Y_i \mid X_i, Z_i, \xi_i) dF(\xi_i) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \Pr(D_i \mid Z_i, \xi_i) \Pr(Y_i \mid D_i, X_i, \xi_i) dF(\xi_i).$$

Since ξ is unobserved, we need to integrate over its domain to account for its existence, assuming that $\xi \perp (X, Z)$. Since the probabilities in the likelihood function are conditioned on ξ, an unobserved factor essential for the selection to training, we have $(Y_1, Y_0) \perp (X, Z, \xi)$, which implies that $\Pr(Y_i \mid D_i, X_i, \xi_i) = \Pr(Y_i \mid X_i, \xi_i)$. This means that both the selection probability and the outcome probabilities are unconditional probabilities in the likelihood function, which reduces the computational burden. We estimate the parameters of the model using maximum likelihood technique, with a Gaussian quadrature to approximate the integrated likelihood.[5]

Identification of the parameters of the model is insured by the exclusion restrictions and the joint normality assumption for the unobserved components of the model. The normalization and the joint normality imply that the joint distribution of (U_1, U_0, U_D) is known and defined by the one-factor structure.

B. The discrete one-factor model

An alternative way of defining the factor structure is to assume that the unobserved factor component can be represented (or approximated) by a number of discrete mass-points. Since Heckman and Singer (1984) proposed this method, to allow for unobserved heterogeneity in duration models, it has been used extensively in the applied literature. Mroz (1999) provides a useful overview of the theoretical basis of the method. It is assumed that the distribution of the unobserved factor can be approximated by a step function given by $\Pr(\xi = \eta_j) = p_j$, $j = 1, 2, ..., J$, with $0 \leq p_j \leq 1$ and $\sum_1^J p_j = 1$. With this distribution the likelihood function is given by

$$L = \prod_{i=1}^{N} \sum_{j=1}^{J} \Pr(D_i, Y_i \mid X_i, Z_i, \xi) \Pr(\xi = \eta_j) .$$

To ensure that the sum-up criteria is fulfilled in the estimation of the mass-points, p_j, we define the probabilities using the cumulative distribution function of the extreme value distribution, which also restricts the mass-points to positive numbers less than one.[6] In order to identify the model, two problems have to be solved. First, the location of the support-points η_j, is arbitrary. The easiest way to solve this is to set one of the support-points to a specific number. Second, the scale of the discrete factor is undetermined. Normalizing one of the factor loadings could solve this problem. In our analysis, we choose to restrict the range of the support-points. We use two points of support in the empirical analysis: one is normalized to zero, i.e., $\eta_1 = 0,$ and the other to one, i.e., $\eta_2 = 1.$

The non-parametric identification of the distribution of the unobserved factor depends on the correlation between the selection equation and the two state equations. However, if no such dependency exists, there would be no need to model the selection, and other methods could be used. It is also essential to have at least three points to peg the distribution with, which in our case is achieved by the use of three equations over which the unobserved factor works (see Heckman, 1981). For a formal proof of the identification for this kind of model, see Carneiro et al. (2003) and Heckman and Taber (1994), and for a discussion of the conditions under which the discrete factor model is identified, see Mroz (1999).

C. Treatment parameters

There are three parameters commonly estimated in the literature: 1) the average treatment effect (ATE), 2) the mean treatment on the treated (TT), and 3) the marginal treatment effect (MTE). The second two parameters are modified versions of the first parameter, and they all represent the mean values of the population under investigation. Estimating a structural model and thereby recovering the full density of the latent variables involved, allow one to determine the distributional effects corresponding to each of the mean effects. The distributional effects offer information about the distribution of the treatment effects, such as the share of the treated that benefits from the program, and the share that is actually worse off participating in the program, etc.

When the outcome variables are discrete and represent a measure for employment, the probability of the events has to be formed. The ATE parameter is therefore defined as the difference in mean probabilities between the two states and across the individuals.

In order to incorporate the unobserved factor, it has to be integrated out over the assumed distribution.[7]

$$\text{ATE}(X,Z) = \int_{-\infty}^{\infty} \left[\Phi(X\beta_1 + \rho_1\xi) - \Phi(X\beta_0 + \rho_0\xi) \right] dF(\xi) . \tag{5}$$

The TT parameter answers the question of how much a person who participated in training gained compared to the case where no training took place. TT is a modified version of ATE in the sense that it considers the conditional distribution of ξ, relevant for those who participated in a program. The parameter is defined as:[8]

$$\text{TT}(X, D=1) = \int_{-\infty}^{\infty} \left[\Phi(X\beta_1 + \rho_1\xi) - \Phi(X\beta_0 + \rho_0\xi) \right] dF(\xi \mid D=1, X, Z). \tag{6}$$

The MTE parameter measures the treatment effect for individuals with a given value (u) of U_D, i.e. the unobserved component of the selection equation,[9] and it is defined in the following way:

$$\text{MTE}(X, U_D = u) = \int_{-\infty}^{\infty} \left[\Phi(X\beta_1 + \rho_1\xi) - \Phi(X\beta_0 + \rho_0\xi) \right] dF(\xi \mid X, U_D = u). \tag{7}$$

When $U_D = 0$, MTE = ATE.

However, these are not the only useful parameters. Heckman (1992), Heckman et al. (1997a) and Heckman and Smith (1998) emphasized that many criteria for the evaluation of social programs require information on the distribution of the treatment effect. For example, questions such as “Among those treated, what percentage benefits from the program and what percentage is hurt by it?” can only be answered by the distributional parameter. In this study, we estimate the distributional parameters for TT, which is defined in the following way:

$$\begin{aligned}\mathrm{TT}_{\text{dist}} &= [Y_1 - Y_0 = 1 \mid X, Z, D = 1] \\ &= \int_{-\infty}^{\infty} \Phi(X\beta_1 + \rho_1\xi)(1 - \Phi(X\beta_0 + \rho_0\xi))dF(\xi \mid D = 1, X, Z).\end{aligned} \tag{8}$$

The distributional treatment parameter, $\mathrm{TT}_{\text{dist}}$, predicts the probability of the event that $Y_1 - Y_0 = 1$, which is interpreted as a successful treatment in the sense that with training the individual received employment, i.e. $Y_1 = 1$, while with no training, no employment would have been received, i.e. $Y_0 = 0$. This gives us the possibility to predict the probability of three different events: 1) the successful event, $Y_1 - Y_0 = 1$; 2) the unsuccessful event, $Y_1 - Y_0 = -1$; and 3) the indifferent event, $Y_1 - Y_0 = 0$. In order to detrmine the predicted probabilities for the remaining events, expression (8) must be elaborated accordingly.

IV. Data

The data analyzed in this paper come from two longitudinal databases, the Swedish Income Panel (SWIP) and Händel, which contain information on personal characteristics, earnings, incomes and unemployment history. SWIP has two components: a sample of people that represents 1% of the Swedish-born population, and another sample that represents 10% of the foreign-born. SWIP is a database of *individual incomes*, built on a stratified random sample drawn (by Statistics Sweden) from the 1978 register of total population (RTB). People from *this* initial sample were followed over time with repeated yearly cross-sections. Additionally, to each consecutive year, a supplementary sample of individuals were added to each cross-sectional unit to adjust for migration in such a way as to make each and every stratified cross-section representative of the Swedish population with respect to each stratum.

Income information is provided by the Swedish tax-register, which also includes information about those who do *not* pay income tax.

Händel is a register-based longitudinal event history database that contains information on *all* persons registered at the *public* unemployment offices. Its observation period starts in August 1991 and (in this paper) ends in December 1997. Händel has a *multiple spell* structure which provides exact information for the starting and ending dates of registered unemployment spells for each individual (with detailed information about the searching and program episodes that compose each spell). In addition to providing other information related to spells and episodes (e.g., the occupation unemployed people are looking for, the amount of desired labor supply, the location of a possible job, the reason for ending the registration spell, etc.), it provides information about personal characteristics of the job seekers (age, gender, citizenship, education, etc.). The main characteristics of this database are those components that allow us to identify the labor market trainees and counterfactuals. We construct treatment and comparison groups for both Swedish- and foreign-born. The selection steps are presented in Appendix A1 and A2, and Tables A1 and A2 in the Appendix present the descriptive statistics of the treatment and comparison groups, stratified by country of birth into Swedish-born and foreign-born. The variable specifications were chosen to be as parsimonious as possible, yet to include variables that are relevant and available. Nevertheless, the minimum relevant information for the selection to training was unavailable, which made it essential to control for unobservables. However, having access to a valid instrument is still an important requirement.

One of the key variables in our analysis is the discrete dependent indicators for employment. We construct these variables using information from both Händel and

SWIP databases. Händel provides information about both the date and employment status at the beginning and the end of each unemployment spell. Unfortunately, this information is not enough to compute the employment duration for a particular year. Therefore, we also use a variable on annual income from SWIP. Controlling for both unemployment dates and employment status, persons were considered to be employed if their annual earnings were at least 40,000 SEK.[10] This level was decided after analyzing the percentage of the employed by various ceiling levels, and the figure corresponds to an average of around 3.5 months of full time work, which functions as a threshold level for being considered to be employed in the analysis.

Another important variable when dealing with control function estimators is the exclusion restriction, or the instrument, that drives the potential effect of a training program. We use the rate of unemployment measured at the municipal level. A change in the local (municipal) unemployment rate is expected to have a significant impact on the demand for social programs that are directed towards groups of unemployed, such as vocational training programs.

When the local unemployment rate increases, the overall propensity to participate in training increases and, with some delay, the policy induced supply of programs meets the demand in order to reduce the open unemployment rate. This causal relationship drives the covariance between unemployment rate and training status.

On the other hand, when the unemployment rate increases, the number of vacancies decreases, which means that the number of employment opportunities for those unemployed are reduced. This reduction decreases the likelihood of finding a new job. Hence, there is causal relationship between unemployment rate and employment opportunities as well, at a given point in time. However, when the training period covers two years (1993-1994) and the employment probability is to be determined one year later (1995), the statistical relationship is reduced. Furthermore, if the local unemployment rate for 1991 is used as a proxy for the rate in 1993, then the relationship with the employment probability in 1995 is very close to zero, and no statistical relationship can be determined. Since the statistical relationship with the training status remains (i.e. is significant), it is expected that the local unemployment rate works satisfactory well as an exclusion restriction or instrument for the selection to vocational training.

V. Results

A. The One-Factor model

This Section reports the results of the one-factor model for 1995, i.e., one year after the training period. Table 1 presents the parameter estimates for the three equations and for three versions of the model: no unobserved factor (NoF), normal unobserved factor (NF), and discrete unobserved factor (DF) for the Swedish-born people. Although the goodness of fit for discrete choice models in general is fairly low, Pseudo R^2 indicates that the fit for both the NF and DF models is quite good, predicting probabilities that are 31-32% better than a model using only constants.[11] The likelihood ratio test indicates that the unobserved factor has a significant effect on the performance of the model.

<Insert Table 1 here>

In the NF and DF models, the constants are replaced by the *factor loadings*, which are designed to capture the effect from unobserved heterogeneity, such as aptitude or ambition or any other relevant factor that is left out of the model. For the DF model, the factor loadings are significant only for the employment equation for the treated and the selection equation, while for the NF model, the factor loading is significant only for the selection equation.[12] For the selection equation, the NF model estimated a factor loading effect that is two times stronger than the value estimated by the DF model.

Since the factor loadings are parts of the covariances of the model, the sign of the factor loadings is important when determining the stochastic relationship between U_1, U_0, and U_D. The factor loading of the employment equation for the treated multiplied by the factor loading of the selection equation represents the covariance between U_1 and U_D. Since this covariance is positive, the selection to training is positive, which indicates that the employment probability is greater for the selected group of trainees compared to what it would have been if the selection to training had been random.

The factor loading of the employment equation for the non-treated multiplied by the factor loading of the selection equation represents the covariance between U_0 and U_D. Since this covariance is negative (but not significant), the selection to non-treatment is positive.[13] This implies that the employment probability of non-treated is higher compared to what it would have been if the selection had been random.

The other estimated parameters differ in sign and size both across models and across equations. For all three equations, having children younger than 18 is the only variable for which all models estimated a significant positive effect. The estimated effect is much larger for the treated (about 0.43) than for the untreated (about 0.22), and much smaller for the selection equation (the NoF and DF models estimated an effect of 0.11, while the NF model estimated an effect of 0.198).

Women are expected to have a lower probability to be employed than men. Except for the DF model for the employment equation of the untreated, all models estimated a significant gender effect for all equations. The estimated effect is much weaker for the untreated (i.e. -0.05) than for the treated (-0.245 for the NoF model and -0.28 for the other two models). In other words, for the untreated, there is a relatively small difference in the probability of getting a job between women and men. Women have also a lower probability of being selected into a training program than men do: the effect estimated by the NF model is much stronger (-0.302) than the effect estimated by the other two models (-0.166 by the NoF model and -0.187 by the DF model).

The age effect estimated by the NoF model is not significant. The other two models estimated a significant *positive* effect for the untreated and a significant *negative* effect for the selection equation. In other words, the probability of being selected into training decreases with age, while for the untreated, the probability of getting a job after one year increases with age.

For both treated and untreated, all three models estimated that those who have high school education have a higher probability of getting a job than those with lower levels of education (the effect estimated by the NoF model for the untreated is not significant). For the selection equation, the estimated effect by all models is *negative*, suggesting that those with a high school education have a lower probability of being selected into the training than those with lower levels of education.

Having a college education is estimated to increase the probability of getting a job after one year for both treated and untreated (but for the treated, only the NoF model estimated a significant effect). Moreover, having a college education is estimated to decrease the probability of being selected into a training program. The NF model estimated a stronger effect (-1.038) than the other two models (-0.672 by the DF model and -0.588 by the NoF model). The fact that the positive effect of a college education is significant for the untreated but not for the treated, might suggest that the non-treated searched for, or even accepted, jobs to a higher extent already when their treated peers were still participating in the programs. Even though training is aimed at people with a low education, about 15% of the trainees have some sort of college education, which indicates that their education did not pay off in the way it was intended. It is reasonable to believe that the unemployed with a college degree have a higher reservation wage compared to those with lower levels of education, which therefore reduces their employment opportunities. Another explanation is that being an unemployed college graduate and participating in a training program might give negative signals to potential employers, thereby reducing the employment probability.

Living in a city region is estimated, by all three models, to *decrease* both the probability of getting a job for the untreated, and the probability of being selected into training. Even though the estimated effects are not significant for the treated, all three models suggest that living in a city region is estimated to *increase* their probability of getting a job.

Local unemployment rate has a positive and significant effect on the probability of being selected in the training. This is expected since it is the unemployment rate that drives the program participation rate. That is, if the unemployment rate increases, more people are sorted into vocational training. Having a college degree and living in a city region turn out to have a positive relation with the selection to training. Furthermore, it is statistically unrelated with the employment probability. For the Swedish-born, this component therefore constitutes the second part of the exclusion restriction in the specification. The concentration of those who are college educated is larger in city regions, which implies that they to a larger extent enter into vocational training programs in those regions. The NF model estimated a stronger effect for both of the exclusion restricting variables compared to the other two models.

Table 2 reports the parameter estimates of the one factor model for the foreign-born people. The level of the goodness of fit for the model is comparable to the level for the Swedish-born people, the results indicating that the NF and DF models perform 34-35% better than the model that contains only constants. The likelihood ratio test indicates that the unobserved factor has a significant effect on the performance of the model, indicating that unobservables are important for the foreign-born as well.

<Insert Table 2 here>

As discussed earlier, the sign of the factor loadings gives an important indication of the sorting structure of the unemployed into the two states. Since the factor loadings of the employment equation for the treated and the selection equation are positive in both the NF and DF models, the covariance between the unobservables of the two equations is positive, which means that the selection to training is positive. That is, the employment probability is greater for the selected group of trainees compared to what it would have been if the selection to training had been random. However, the overall effect is a function of both the observed and the unobserved components.

The age effect is significant only in the selection equation estimated by the NoF and DF models, and suggests that the probability of being selected into a training program decreases with age.

The estimated effect of gender is significant only in the selection equation (without the NF model), which shows that women have a lower probability of being selected into a training program than men do. For the employment equations, the gender effect estimated by all three models is not significant. However, the estimates show that treated women have a lower probability of getting a job after one year compared to men, while untreated women have a higher probability.

The estimated effect of educational level for the untreated is significant for all three models, and shows that those who have high school or college education have a higher probability of getting a job than those with lower levels of education. The effects of both high school and college education are not significant for the treated. For the selection equation, all models suggest that those with a high school education have a higher probability of being selected into a training program than those with lower levels of education. The estimates are not significant for the NF model.

Living in a city region is estimated by all three models to decrease both the probability of being selected into a training program, and the probability of getting a job for the untreated. The estimated effects are not significant for the treated.

All three models suggest that having children *increases* the probability of getting a job for both treated and untreated, but *decreases* the probability of being selected into a training program. However, the parameter estimated by the NF model is not significant.

Important variables when analyzing foreigners are the country of origin, and duration in the host country since immigration.[14] The parameter estimates for the country of origin suggest that people born in a country outside Europe are a subgroup with particular problems. The groups with the bigger negative effect were those from Arab and African countries. For all three equations, being born in one of these countries are the only variables for which all models estimated a significant negative effect. Being born in one of these countries decreases the probability of being selected into a training program, and also the probability of getting a job regardless of participating in training or not.

For the trainees, with the exception of these two variables and the variable "has children", the rest of the observed characteristics have no significant effect on the employment probability. Hence, for those who participated in training, country of origin was the major factor for the probability of receiving a job one year after the training period.

Number of years in the country has a significant effect for the *untreated*, suggesting that for this group the relatively new immigrants have a higher probability of getting a job than those who have lived in Sweden for more than ten years. Compared with those who have been residents for more then ten years, people who have been residents for less than ten years are more likely to get a job (the probability is even higher for those who have been residents for less than six years). Local unemployment rate has a positive effect on the probability of being selecting into the training, just as for the Swedish-born group.

B. Mean and distributional treatment effects

Table 3 reports the mean treatment effects based on the estimated parameters in the three models. There is a relatively big difference across the models and also between Swedish-born and foreign-born. For example, the ATE parameters estimated by the three models are almost the same for Swedish-born and foreign born, but the size of the parameters estimated by the NF model is much higher than the parameter estimated by the other two models.

<Insert Table 3 here>

In the first year after the training, the ATE parameter is negative for both Swedish- and foreign-born people, suggesting a negative effect of training for a randomly chosen individual from the population. This estimate is in accordance with the literature on Swedish data that primarily reports either negative or non-significant effects from training.[15] This is not of special concern, ATE being a hypothetical parameter that is of less interest from a policy point of view since publicly funded training is seldom aimed at the total population but at a selected group with problems finding jobs.

The TT parameter is of more interest, since the employment probability of the two states is adjusted by the probability of being treated. For Swedish-born, the TT parameter is positive and significant for the NF model, while it is not significant but yet positive for the DF model. The NoF model estimated a negative (almost zero) parameter that is not significant. For the foreign-born, the TT parameter is very small but not significant for any of the three models. In conclusion, one could say that the effect of training is zero or slightly positive for the Swedish-born.

The last effect, $\mathrm{TT}-\mathrm{MTE}(u=0)$, gives a measure for the sorting gain generated from the selection process. The marginal treatment effect estimated here represents the treatment effect for those on the margin of being selected into the training, as predicted by the model. The sorting gains are positive and significant for both Swedish- and foreign-born when controlling for unobserved heterogeneity. When the factor is assumed normal, the effect is larger for both groups. The sorting gain is larger for Swedish-born, with an almost double size compared to the foreign-born.

For both Swedish- and foreign groups, when no factor loading is included in the model, the estimates are not significant for any of the parameters, and they are very close to zero. That the NoF model generates the same result for all parameters comes as no surprise since it does not account for potential selection bias. When no selection bias is present, ATE and TT effects are the same, which implies that the sorting gain should be zero.

Table 4 presents the estimates for the distributional treatment effects with respect to the treatment on the treated. We have three measures: 1) the share that gained from training (or positive effect); 2) the share that lost from training (or negative effect); and 3) the share with no effect at all.

<Insert Table 4 here>

The distributional assumptions used here seem to be of less importance for the estimated effects since they are very close to each other for both Swedish- and foreign-born. For the Swedish-born trainees, 21-25% gained from the training, while 18-19% lost from it, and 57-59% had no effect from training (which means that they either would have received a job without the training, or they would not have received a job in any case). For the foreign-born trainees, we have a similar situation, but with somewhat larger numbers for those who gained from training (24-27%) and those who lost from it (24-25%), and a lower number (50-51%) for those who had no effect from training.

Table 5 presents correlation measures that illustrate to what degrees observed and unobserved factors are associated with each other. For the Swedish-born, most correlation coefficients are significant. There are only the relations between the unobserved components of the treated and the untreated states, and between the unobserved components of the untreated and training states that are not significant. The component of the training state, on the other hand, is related to the unobservables of the selection equation. This confirms the presence of a sorting structure, which shows that those most likely to gain from training go to training, as driven by components that the analyst has no access to. Another interesting correlation is the one between the selection and the treatment effect. The linear relationship between the observables only, is stronger than their relationship when the unobservables are included.

<Insert Table 5 here>

For the foreign-born, the picture is somewhat different. The level and significance of the correlation measures differ, and when using discrete factor approximation, none are significant, even though the signs of the measures in most cases are the same for the two models.

C. Other estimators in the literature

The mean treatment effects presented in the previous sections will now be compared to our own matching estimations, using the same variable specification as in the factor model, and to results from the previous literature. Our own estimations are based on three different propensity-score matching estimators: two cross-sectional matching estimators and a difference in difference matching estimator (see Heckman et al. 1997b, 1998a, and Heckman et al., 1998b).

The matching estimator is of special interest here since the identifying assumption imposed requires that the outcomes are independent of the treatment choice given the observed variables, which is the conditional independence assumption restriction. This assumption is relaxed in the factor model by instead allowing for unobserved heterogeneity, which is essential in explaining the selection. The matching model estimator can therefore be seen as a special case of the one-factor model, where it is assumed that the conditional independence assumption holds if an unobserved random variable is included in the conditioning set (Aakvik et al., 1999). Using the method of matching, we estimate the ATE and TT parameters.[16]

When testing for significance of the matching estimates we use the usual variance formula for the variance of differences in means. A potential problem with this is that it ignores the components of the variance due to the estimation of scores. Asymptotically, the part due to the estimation of the scores goes away due to the faster convergence of the parametric propensity score model. Additionally, Heckman, Ichimura and Todd (1997) present Monte Carlo estimates that show that this component of the variance matters even for samples of moderate size. However, Eichler and Lechner (2001), who compared the simple estimator with the bootstrap, suggest that it can be ignored with samples in the 1000s. We follow the last study's suggestion on this point since we use a sample of around 1000 individuals.

Table 6 presents the estimates together with simple mean differences in probabilities between the two outcome equations.

<Insert Table 6 here>

For the Swedish-born, the simple mean differences have very low values and none are significant for the three consecutive years. Furthermore, the size of the estimates is decaying over time. The matching estimators show the same picture, and are similar in size (around 3%).

For the foreign-born, the situation is slightly different. The simple mean differences are much larger than the estimates from the matching estimates, and the effect is growing from the first year to the second year. None of the three matching estimates is significant. The point estimates are lower than for the Swedish born, which also is the case for the factor model.

The overall conclusion is that vocational training has no effect on the employment probability when unobserved factors are left out. This picture is also partly confirmed by the previous studies of treatment effects of labor market training in Sweden during the 1990s, whose results tend to give a picture of initial negative effects moving towards zero effects (see Calmfors et al., 2002).

Larsson (2003) evaluated Swedish youth programs in 1992-1993 for individuals aged 20-24 using propensity score matching, and found negative and significant effects on the employment probability when measured one year after completed training. Okeke (2001) analyzed register and survey data on a stratified sub-sample of participants in labor market training using propensity score matching, and found a positive and significant effect on the employment probability six months after the completion of training. Richardson and van den Berg (2001) analyzed a 1% random sub-sample of all who become openly unemployed during the 1993-2000 period, using a bivariate duration model investigating the unemployment duration. They found a negative and significant effect that vanished within two months after the training ended. Sianesi (2002) analyzed adult individuals entitled to unemployment benefits who registered at employment offices for the first time in 1994. Using matching estimators, she found negative and significant effects on the employment rates up to 30 months, but no significant effects afterwards.

VI. Summary and conclusions

We estimated a one-factor model that allows for unobserved heterogeneity using the factor loading technique within the framework of full information maximum likelihood. The model was estimated with different distributional assumptions for the unobserved factor, in order to detect possible differences in the training effect due to the distributional assumption of the factor. The structural model allowed us to estimate both mean treatment effects and distributional treatment effects, focusing on those who participated in training.

We investigated how the effect is distributed across the participants and explored the relationship between selection into training and the employment probability. This has been done for Swedish-born and for foreign-born separately, focusing on people participating in a labor market program in Sweden during 1993-1994. The effect on employment probability has been evaluated for the following year.

The treatment effect on employment probability for the Swedish-born is driven by being a man, having a high school education, having children younger than 18, and a heavy load of the unobserved factor. The predominant component is the loading factor, which has a larger effect on the outcome then the other components. The ATE parameter is negative for the first year after training, suggesting a negative effect from training for a random chosen individual. The TT parameter is positive and indicates that the participation in training increased the employment probability by around 7%. The fact that TT>ATE indicates that the selection into training is positive. The distributional parameter suggests that around 22% gained from training, while 20% were harmed by it. The estimated values of the NF and the DF models are very similar for the marginal effects, even though the factor loadings are non-significant in the outcome equations in the NF model. The treatment parameters are much larger in absolute terms for the NF model. However, the TT parameter is only significant for the NF model. Comparing the distributional parameters, only small differences could be found. The sorting effect due to unobservables is significant for both models, yet much larger for the NF model.

The treatment effect on the employment probability for the foreign-born is driven by factors such as having children younger than 18, and being from an Arab or an African country. The unobserved factor has a positive effect on the employment probability, but is significant only for the treated. The mean treatment parameters show a negative effect for the average treatment effect and no effect for the treatment on the treated; yet the sorting gain is positive and significant. The distributional treatment parameter shows that after the first year, around 26% gained from training, while 24% were harmed by it. The NF model generated larger effects than the DF model.

When comparing the NF and DF models, a clear distinction appears when comparing the estimates for the mean treatment parameters. The NF model tends to generate larger and slightly positive effects, while the DF model is closer to the matching estimates, i.e. small and non-significant. One should keep in mind that the non-parametric distribution of the DF model is approximated by just two mass-points, which was a number that the present data could handle. This limitation should be kept in mind when analyzing the results from the DF model.

Since we estimated a positive and significant effect of the sorting gain for both models, it is clear that the conditional independence assumption does not hold, which means that the matching estimates of this study are biased. This suggests that another estimator that is more robust to unobserved heterogeneity should be used, and therefore it proves that the one-factor model estimates of this study are preferred to the matching estimates.

Acknowledgements

We acknowledge useful comments from Arthur van Soest, and financial support from The Swedish Research Council. Thomas is also grateful for generous financial support from the Jan Wallanders and Tom Hedelius Foundation.

References

Aakvik, A. (1999) Five Essays on the Microeconometric Evaluation of Job Training Programs, University of Bergen, Dissertation.

Aakvik, A., J. J. Heckman, and E. J. Vytlacil. (2000) Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs, NBER Technical Working Paper 262.

Beenstock, M. (1996) Training and the time to find a job in Israel, *Applied Economics* 28, 935-46.

Bryant, R. and A. Wilhite. (1990) Military experience and training effects on civilian wages. *Applied Economics* 22, 69-81.

Calmfors, L., A. Forslund and M. Hemström. (2002) Does Active Labour Market Policy Work? Lessons from the Swedish Experience, Institute for International Economic Studies, Stockholm, Seminar paper No. 700.

Carneiro, P., K. Hansen, and J. J. Heckman. (2003) Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice, *International Economic Review* 44, 361-422.

Edin, P.-A. and O. Åslund. (2001) Invandrare på 1990-talets arbetsmarknad (Immigrants in the Labor Market during the 1990s), in SOU 2001:54 *Ofärd i välfärden*, Stockholm: Fritzes.

Eichler, M. and M. Lechner. (2001) Public Sector Sponsored Continuous Vocational Training in East Germany: Institutional Arrangements, Participants, and Results of Empirical Evaluations (pp. 208-253), in: Riphahn et al. (Eds.) *Employment Policy in Transition: The Lessons of German Integration for the Labor Market*, Heidelberg: Springer.

Eriksson, M. (1997) To Choose or Not to Choose: Choice and Choice Set Models, Umeå University, Umeå Economic Studies No. 443.

Groot, W. and H. M. Van den Brink. (2000) Education, training and employability, *Applied Economics* 32, 573-81.

Heckman, J. (1978) Dummy Endogenous Variables in a Simultaneous Equations System, *Econometrica* 46, 1931-60.

Heckman, J. (1979) Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-61.

Heckman, J. (1981) Statistical Models for Discrete Panel Data, in: C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. M.I.T. Press.

Heckman, J. (1992) Randomization and Social Program Evaluation (pp. 201-230), in Charles Manski and Irwing Garfinkel (Eds.), *Evaluating Welfare and Training Programs*, Cambridge, MA: Harvard University Press.

Heckman, J. (1997) Instrumental Variables: A Study of Implicit Behavioural Assumptions Used in Making Program Evaluations, *Journal of Human Resources* 32, 441-62.

Heckman, J. and B. Singer. (1984) A Method for Minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* 52, 271-320.

Heckman, J. and C. R. Taber. (1994) Econometric Mixture Models and More General Models for Unobservables in Duration Analysis, NBER Technical Working Paper 157.

Heckman, J. and J. Smith. (1998) Evaluating the Welfare State, in S. Strom (Ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

Heckman, J., J. Smith, with the assistance of Nancy Clements. (1997a) Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts, *Review of Economic Studies* 64, 487-535.

Heckman, J., H. Ichimura, and P. Todd. (1997b) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program, *Review* *of Economic Studies* 64, 605-54.

________ (1998a) Matching as an Econometric Evaluation Estimator, *Review of Economic Studies* 65, 261-94.

________ (1998b) Characterizing Selection Bias using Experimental Data, *Econometrica* 66, 1017-98,

Heckman, J., R. LaLonde and J. Smith. (1999) The Economics and Econometrics of Training Programs, in: O. Ashenfelter and D. Card (Eds.), *Handbook of Economics*, Volume III, Amsterdam: North Holland.

Heckman, J., J. Tobias, and E. Vytlacil. (2000) Simple Estimators for Treatment Parameters in a Latent Variable Framework with an Application to Estimating the Return to Schooling, NBER Working paper 7950.

Heckman, J. and S. Navarro-Lozano. (2004) Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models, *Review of Economics and Statistics* 86, 30-57.

Judd, L. K. (1998) *Numerical Methods in Economics*. The M.I.T. Press.

Kluve, J. and C. Schmidt. (2002) Can Training and Employment Subsidies Combat European Unemployment, *Economic Policy* 17, 450-97.

Kraft, K. (1998) An evaluation of active and passive labour market policy, *Applied Economics* 30, 783-93.

Larsson, L. (2003) Evaluation of Swedish Youth Labor Market Programs, *Journal of Human Resources* 38, 891-927.

Main, B. G. M. (1991) The effects of the Youth Training Scheme on employment probability, *Applied Economics* 23, 367-72.

Makepeace, G. H. 1996. Lifetime earnings and the training of young men in Britain, *Applied Economics* 28:6, 725-35.

Mroz, T. (1999) Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome, *Journal of Econometrics* 92, 233-74.

Okeke, S. (2001) Arbetsmarknadsutbildningen 1999- effekter för individen. Stockholm: The National Labour Market Board, Ura 2001:7.

Richardson, K. and G. van den Berg. (2001) The Effect of Vocational Employment on the Individual Transition Rate from Unemployment to Work, *Swedish*

Economic Policy Review 8, 177-213.

Sianesi, B. (2002) Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults during the 1990s, Uppsala: The Institute for Labour Market Policy Evaluation, Working Paper 2002:5.

Vella, F. (1998) Estimating Models with Sample Selection Bias: A Survey, *Journal of Human Resources* 33, 127-72.

Åtgärdsundersökning. (1998) En Uppföljning av Deltagare som Avslutat Konjunktureberoende Åtgärd Fjärde Kvartalet 1996. Stockholm: The National Labour Market Board, Apra 1998:3.

APPENDIX

A1 The construction of the treatment group

Given the available information, we selected individuals who fulfill the following criteria:

1) they completed *one* vocational training program during 1993-1994;
2) they did not participate in any program during 1991-1992 and 1995;
3) they were 20-60 years old at the time the program started. The age selection was done considering the following two aspects: 1) in general people are allowed to participate in a vocational training program if they are at least 20 years old; 2) we would like all individuals to be under the mandatory retirement age (65 years) in the last year (1997) of the analyzed period.

Applying these selection filters to Händel and merging this sample with the SWIP database, the size decreases to 1,099 persons: 534 Swedish-born and 565 foreign-born.

A2 The construction of the comparison group

Given the available information and the selection criteria for the treatment group, we construct a comparison group using the following filters:

1) they were unemployed *at least* 30 days in 1993 or *at least* 30 days in 1994. This filter was designed in a such way that there is a minimum unemployment spell during the training period (1993-1994) when people could qualify for starting a labor market program;
2) they did not participate in any vocational program during 1991-1995;
3) they were 20-60 years old at the time the program started.

After merging the sample of non-participants from Händel with the SWIP database, a sample of 12,327 persons was obtained: 5,776 Swedish-born and 6,551 foreign-born. The first filter was imposed in order to harmonize the unemployment behavior between the treated and the untreated. The objective was to form two groups with comparable unemployment characteristics.

Table A1 Mean values of the demographic characteristics, 1993

	Swedish-born		Foreign-born	
	Treatment group n = 534	**Comparison group** n = 5776	**Treatment group** n = 565	**Comparison group** n = 6551
Women	0.45	0.49	0.43	0.46
Age	34.46	31.30	36.44	35.42
	(10.2)	(9.86)	(9.55)	(9.09)
Age groups				
19-25 years	0.21	0.38	0.14	0.15
26-45 years	0.61	0.49	0.67	0.68
46-55 years	0.17	0.13	0.19	0.17
Married	0.38	0.27	0.49	0.47
Municipality groups				
Stockholm	0.11	0.16	0.18	0.32
Göteborg	0.08	0.09	0.12	0.13
Malmö	0.06	0.06	0.08	0.08
Other	0.75	0.69	0.62	0.47
Country of origin				
Nordic (excl. Sweden)			0.40	0.32
Western countries			0.09	0.07
East Europe			0.12	0.08
South Europe			0.07	0.09
Arab countries			0.13	0.18
Africa			0.10	0.13
Other			0.09	0.13
Years in Sweden			11.04	9.95
			(7.30)	(7.00)
0-5 years			0.29	0.34
6-10 years			0.14	0.18
> 11 years			0.56	0.48

Note: Standard deviations are reported within parentheses only for quantitative variables. The rest of the variables are all dummies (taking a value of 1 for the mentioned category, and 0 otherwise). This holds true for the next table as well.

Table A2 Education and unemployment characteristics

	Swedish-born		Foreign-born	
	Treatment group n = 534	**Comparison group** n = 5776	**Treatment group** n = 565	**Comparison group** n = 6551
Years from last degree	9.41	7.27	13.10	10.75
	(10.78)	(10.01)	(18.02)	(15.68)
Education groups				
Low	0.23	0.21	0.28	0.40
Medium	0.64	0.57	0.54	0.41
High	0.13	0.22	0.18	0.19
Days of unemployment by year				
1990	1.22	2.16	2.07	4.34
	(12.05)	(19.79)	(20.62)	(28.30)
1991	33.94	37.57	43.07	43.93
	(70.38)	(73.27)	(81.41)	(81.49)
1992	121.42	103.81	135.49	107.92
	(134.99)	(127.81)	(140.36)	(134.58)
1993	244.73	226.53	263.65	210.55
	(131.06)	(112.46)	(125.95)	(136.41)
1994	261.59	253.48	286.43	219.35
	(129.92)	(149.67)	(116.10)	(140.98)
Share with employment				
1995	0.73	0.70	0.54	0.51
1996	0.71	0.68	0.57	0.48
1997	0.72	0.71	0.57	0.48
Municipal unemployment rate (%)				
1991	7.49	7.11	7.13	6.76
	(2.01)	(2.07)	(2.06)	(2.07)
1993	14.47	13.97	14.05	13.61
	(2.43)	(2.55)	(2.49)	(2.58)
1995	15.88	15.26	15.35	14.86
	(2.85)	(2.91)	(2.79)	(2.91)

Table 1 Parameter estimates for Swedish-born

	NoF model			NF model			DF model		
	P.E.	S.E	M.E.	P.E.	S.E	M.E.	P.E.	S.E	M.E.
Employment equation-treated									
Factor	-	-	-	0.257	0.194	0.095	0.453**	0.231	0.156
Age	0.058	0.079	0.019	-0.014	0.059	-0.005	-0.009	0.058	-0.003
Woman	-0.245*	0.128	-0.082	-0.287**	0.132	-0.106	-0.288**	0.130	-0.099
Education (CG: lower)									
High School	0.388**	0.172	0.131	0.282*	0.146	0.104	0.278*	0.144	0.096
College	0.363*	0.212	0.122	0.234	0.218	0.086	0.281	0.208	0.097
Children	0.433***	0.131	0.145	0.463***	0.132	0.171	0.432***	0.132	0.149
City region	0.072	0.171	0.024	0.004	0.143	0.002	0.026	0.141	0.009
Employment equation-untreated									
Factor	-	-	-	-0.095	0.276	-0.032	-0.161	0.171	-0.053
Age	0.109	0.019	0.036	0.104***	0.013	0.035	0.126***	0.020	0.041
Woman	-0.056***	0.035	-0.018	-0.058*	0.035	-0.019	-0.049	0.035	-0.016
Education (CG: lower)									
High School	0.285	0.045	0.094	0.263***	0.037	0.089	0.315***	0.048	0.104
College	0.338***	0.052	0.112	0.326***	0.048	0.110	0.366***	0.055	0.121
Children	0.223***	0.039	0.073	0.232***	0.040	0.078	0.221***	0.039	0.073
City region	-0.163***	0.036	-0.054	-0.168***	0.036	-0.057	-0.157***	0.036	-0.052
Selection equation									
Factor	-	-	-	1.434***	0.178	0.101	0.711***	0.091	0.104
Age	-0.055	0.022	-0.008	-0.082**	0.037	-0.006	-0.084***	0.023	-0.012
Woman	-0.166**	0.046	-0.025	-0.302***	0.085	-0.021	-0.187***	0.049	-0.027
Education (CG: lower)									
High School	-0.121***	0.053	-0.018	-0.187**	0.095	-0.013	-0.173***	0.056	-0.025
College	-0.588**	0.087	-0.089	-1.038***	0.170	-0.073	-0.672***	0.091	-0.098
Children	0.112***	0.049	0.017	0.198**	0.087	0.014	0.119**	0.051	0.017
City region	-0.399**	0.057	-0.061	-0.694***	0.110	-0.049	-0.477***	0.060	-0.069
City region & College	0.317***	0.125	0.048	0.617***	0.217	0.043	0.361***	0.132	0.052
Local unemployment	0.063**	0.006	0.009	0.113***	0.014	0.008	0.084***	0.007	0.012
a of mass-point P_1	-	-	-	-	-	-	-0.143	0.121	-
LL model	-5734			-5741			-5692		
LL constants				-8479			-8376		
LL no factors				-5750			-5734		
LR test for no factor				18			84		
Pseudo R^2				0.32			0.31		
AIC				5764			5716		

Notes: CG means comparison group; P.E. means parameter estimate; S.E. means standard error; and M.E. means marginal effect. The marginal effects are means and are defined as the analytical derivatives averaged over the unconditional distribution over X. The estimate is significant at the 1% (***), 5% (**), or the 10% (*) level. The estimated coefficient *a* reported in this table is used to compute the mass-point $P_1 = \exp(a)/(1-\exp(a))$. LL stands for Log likelihood. LR represents the likelihood ratio test that tests the model specification against the specification with no factor. $\mathrm{AIC} = -\mathrm{LL} + k,$ where k represents the number of estimated parameters. These notes also apply to Table 2.

Table 2 Parameter estimates for foreign-born

	NoF			NF			DF		
	P.E.	S.E	M.E.	P.E.	S.E	M.E.	P.E.	S.E	M.E.
Employment equation-treated									
Factor	-	-	-	0.336***	0.124	0.115	0.622**	0.308	0.223
Age	0.042	0.063	0.016	-0.011	0.075	-0.003	-0.016	0.068	-0.006
Woman	-0.145	0.113	-0.055	-0.191	0.135	-0.065	-0.183	0.125	-0.066
Education (CG: lower)									
High School	0.034	0.131	0.013	0.012	0.121	0.004	-0.036	0.137	-0.013
College	0.134	0.170	0.051	0.125	0.175	0.043	0.106	0.176	0.038
Has children	0.343***	0.115	0.129	0.328**	0.128	0.113	0.340***	0.120	0.122
City region	-0.067	0.115	-0.025	-0.136	0.150	-0.047	-0.094	0.123	-0.033
Country of origin (CG: Nordic)									
East Europe	-0.190	0.199	-0.071	-0.223	0.216	-0.076	-0.218	0.212	-0.078
West Europe	-0.169	0.184	-0.063	-0.182	0.193	-0.062	-0.211	0.197	-0.075
South Europe	-0.178	0.219	-0.067	-0.264	0.256	-0.091	-0.243	0.236	-0.087
Arab countries	-0.433**	0.191	-0.163	-0.544**	0.259	-0.187	-0.530**	0.223	-0.191
Africa	-0.716***	0.210	-0.271	-0.848***	0.313	-0.292	-0.839***	0.259	-0.301
Other nations	-0.085	0.202	-0.032	-0.175	0.238	-0.060	-0.148	0.216	-0.053
Years since immigration (CG: >10)									
0- 5 years	-0.213	0.179	-0.081	-0.256	0.198	-0.088	-0.255	0.192	-0.092
6-10 years	0.093	0.135	0.035	0.078	0.141	0.027	0.074	0.141	0.026
Employment equation-untreated									
Factor	-	-	-	0.093	0.456	0.034	-0.195	0.163	-0.071
Age	-0.018	0.019	-0.006	-0.016	0.014	-0.006	0.001	0.019	0.000
Woman	0.036	0.033	0.013	0.036	0.032	0.013	0.046	0.034	0.017
Education (CG: lower)									
High School	0.205***	0.036	0.075	0.211***	0.046	0.077	0.221***	0.037	0.081
College	0.339***	0.045	0.125	0.343***	0.051	0.125	0.349***	0.046	0.127
Has children	0.191***	0.033	0.070	0.191***	0.034	0.069	0.193***	0.034	0.071
City region	-0.069**	0.032	-0.025	-0.072**	0.036	-0.026	-0.062*	0.033	-0.023
Country of origin (CG: Nordic)									
East Europe	0.046	0.066	0.017	0.047	0.066	0.017	0.060	0.066	0.022
West Europe	-0.287***	0.062	-0.105	-0.286***	0.062	-0.104	-0.281***	0.062	-0.103
South Europe	-0.206***	0.059	-0.075	-0.208***	0.061	-0.076	-0.193***	0.029	-0.071
Arab countries	-0.699***	0.052	-0.257	-0.703***	0.058	-0.257	-0.683***	0.021	-0.250
Africa	-0.799***	0.057	-0.294	-0.803***	0.063	-0.294	-0.780***	0.055	-0.285
Other nations	-0.231***	0.053	-0.085	-0.233***	0.053	-0.085	-0.212***	0.053	-0.077
Years since immigration (CG: >10)									
0- 5 years	0.196***	0.047	0.072	0.196***	0.047	0.072	0.203***	0.047	0.074
6-10 years	0.289***	0.038	0.106	0.291***	0.041	0.106	0.296***	0.039	0.109
Selection equation									
Factor	-	-	-	0.658***	0.225	0.089	0.112***	0.041	0.016
Age	-0.076***	0.022	-0.011	-0.092	0.058	-0.012	-0.081***	0.022	-0.012
Woman	-0.148***	0.045	-0.022	-0.178	0.114	-0.024	-0.151***	0.045	-0.022
Education (CG: lower)									
High School	0.185***	0.048	0.027	0.221	0.137	0.030	0.181***	0.048	0.026
College	0.047	0.064	0.007	0.056	0.083	0.007	0.044	0.064	0.006
Has children	-0.090*	0.046	-0.013	-0.109	0.083	-0.015	-0.091**	0.046	-0.014
City region	-0.411***	0.045	-0.061	-0.492*	0.282	-0.067	-0.418***	0.045	-0.062
Country of origin (CG: Nordic)									
East Europe	-0.097	0.084	-0.014	-0.116	0.121	-0.015	-0.101	0.084	-0.015
West Europe	0.112	0.078	0.016	0.133	0.120	0.018	0.109	0.078	0.016
South Europe	-0.256***	0.086	-0.038	-0.306	0.201	-0.042	-0.259***	0.086	-0.038
Arab countries	-0.310***	0.072	-0.046	-0.372	0.227	-0.051	-0.315***	0.072	-0.046
Africa	-0.327***	0.078	-0.048	-0.392	0.241	-0.053	-0.335***	0.078	-0.049
Other nations	-0.350***	0.077	-0.052	-0.420*	0.254	-0.057	-0.356***	0.077	-0.052
Years since immigration (CG: >10)									
0- 5 years	-0.092	0.068	-0.013	-0.110	0.102	-0.015	-0.095	0.068	-0.014
6-10 years	-0.035	0.053	-0.005	-0.041	0.067	-0.006	-0.036	0.053	-0.005
Local unemployment	0.054***	0.006	0.008	0.064*	0.036	0.009	0.056***	0.006	0.008
a of mass-point P_1	-	-	-	-	-	-	-0.125	0.278	-
Log likelihood	-6543			-6539			-5917		
L-L constants				-9849			-9131		
L-L no factors				-6549			-5945		
LR test for no factors				20			56		
Pseudo R^2				0.34			0.35		
AIC				6586			5964		

Table 3 Mean treatment parameters

	Swedish-born			Foreign-born		
Parameter	NoF	NF	DF	NoF	NF	DF
ATE	-0.0180	-0.139*	-0.051*	-0.002	-0.124*	-0.007
TT	-0.0003	0.066*	0.024	0.023	-0.009	0.034
TT–MTE(u=0)	0.0177	0.205*	0.075*	0.025	0.115*	0.041*

Note: * indicates significance at the 10% level, and the standard errors used for the test of significance were determined using the delta method.

Table 4 Distributional treatment parameters

	Swedish-born			Foreign-born		
Parameter	No Factor	Normal Factor	Discrete Factor	No Factor	Normal Factor	Discrete Factor
Positive effect	0.201	0.248	0.217	0.261	0.241	0.272
No effect	0.598	0.570	0.590	0.502	0.510	0.491
Negative effect	0.201	0.182	0.193	0.237	0.249	0.237

Table 5 Correlation indices

	Swedish-born		Foreign-born	
Correlations	Normal Factor	Discrete Factor	Normal Factor	Discrete Factor
Corr[$Z\beta_D$, $X(\beta_1 - \beta_0)$]	0.316*	0.261*	-0.035	-0.093
Corr[U_D, $U_1 - U_0$]	0.198*	0.231*	0.093*	0.055
Corr[$Z\beta_D + U_D$, $X(\beta_1 - \beta_0) + (U_1 - U_0)$]	0.200*	0.230*	0.086*	0.048
Corr[U_1, U_0]	-0.023	-0.065	0.030	-0.101
Corr[U_D, U_0]	-0.077	-0.092	0.051	-0.021
Corr[U_D, U_1]	0.203*	0.239*	0.175*	0.058

Note: * indicates that the estimate is significant at the 10% level.

Table 6 Parameter estimates from the matching models

	Swedish-born		Foreign-born	
Parameter	Estimate (%)	t-test	Estimate (%)	t-test
Mean difference $_{1995}$	2.68	1.35	2.65	1.21
Mean difference $_{1996}$	1.95	0.96	8.90*	4.08
Mean difference $_{1997}$	-0.29	-0.14	8.13*	3.72
Cross-sectional matching ATE	1.53	0.36	0.19	0.06
Cross-sectional matching TT	2.99	1.09	-0.18	-0.06
Diff-in-Diff matching TT	3.18	0.94	-1.41	0.39

Note: * indicates that the estimate is significant at the 10% level.

Notes

[1] We present the structure and rules of the system valid during 1993-1994, the period analyzed by this study.

[2] Eriksson (1997) carried out an informal telephone interview with Swedish officials, and found that during the contact between the unemployed and the administrator, ambition and motivation of the unemployed were important for recruitment to a training program. Åtgärdsundersökning (1998) interviewed individuals who participated in a program in 1997. This survey showed that 60% of the participants took the initiative to participate in the training program (i.e., by getting informed about different courses and programs from ring binders, billboards, and/or computer terminals available at the unemployment office).

[3] As many unemployment spells are short, a reasonable strategy for officials at labor market offices is to concentrate training offers to people with longer unemployment spells and others who can be assumed to have difficulties being employed without such efforts. Okeke (2001) reports an average waiting time before starting a training program of three months.

[4] When selecting into vocational training, two main decision-makers are involved, i.e. the program administrator and the unemployed. The equation should be seen as a measure for the combined effort of the two with respect to the involved variables, since several decisions easily may be represented by only one index.

[5] We use Gauss-Hermite quadrature to evaluate the integrals in the model, using five evaluation points. Points and nodes are taken from Judd (1998).

[6] The first mass-point is defined as $P_1 = \exp(a)/(1-\exp(a))$, where "$a$" is estimated. In order to receive the mass-point, one has to apply the formula.

[7] Note that ATE(X, Z) does not depend on Z, so that $\mathrm{ATE}(X, Z) = \mathrm{ATE}(X)$. We choose to include Z to emphasize that the estimated values of β_1, β_0, ρ_1, and ρ_0 depend on Z, because the selection equation is estimated jointly with the two outcome equations.

[8] $\mathrm{dF}(\xi \mid D=1, X, Z) = \mathrm{dF}(\xi \mid D=1, Z)$. By Bayes' rule, $\mathrm{dF}(\xi \mid D=1, X, Z) = \Phi(Z\beta_D + \rho_D \xi)\mathrm{dF}(\xi) / \Phi(Z\beta_D / \sigma_D)$, which is used in (6).

[9] See Heckman (1997), Heckman and Smith (1998), and Heckman et al. (2000).

[10] Assume that an individual has a wage rate of 50 SEK per hour. With an annual income of 40,000 SEK, he or she would be working 800 hours per year, which roughly corresponds to 5 months of full-time work. If instead the wage rate were 100 SEK per hour, the corresponding figure would be 2.5 months of full-time work. We believe that the true number of full-time equivalence lies somewhere in between these two numbers. In May 2004, 100 SEK = 10.74 EUR.

[11] Pseudo R^2 is a goodness of fit measure defined as $1\text{-}1/[1+2(\log L_1\text{-}\log L_0)/N]$, with N being the number of observations used in the estimation. The measure is based on a model estimated only with the factors of the models, because there are no ordinary constants included in the model.

[12] The statistical significance refers to a significance level of 10% or better. This is applied throughout the paper, unless otherwise stated.

[13] Non-trainees have lower values of U_D, which corresponds to a lower probability to participate in training. Since σ_{0D} is negative, it follows that they have higher values of U_0, which corresponds to an increased employment probability compared to what the employment probability would have been if the selection were random.

[14] Edin and Åslund (2001) describe the labor market situation in Sweden for foreign-born, and find that the immigrants as a group have a weak position in the labor market, especially since large groups came to Sweden as refugees during the 1990s.

[15] See Calmfors et al. (2002) for a survey of the evaluation of active labor market programs in Sweden.

[16] The matching estimator used in the study is the average nearest neighbor estimator, using one neighbor. When estimating the propensity score used in the matching procedure, we use a parametric probit. The choice of variables in the probit model is the same as in the factor model for comparability reasons. Both the balancing score and match of propensity distribution are fulfilled. More details with estimates and statistics about the matching procedure may be received from the authors on request.