 Oxford Economic Papers 52 (2000), 497–520

Evaluating public sector sponsored training in East Germany

By Bernd Fitzenberger* and Hedwig Prey†

* Department of Economics, University of Mannheim, 68131 Mannheim, Germany
† FAA, University of St. Gallen, Guisanstr. 92, 9010 St. Gallen, Switzerland

Support of training has been a very important instrument of active labor market policy in East Germany. This paper attempts an evaluation of the employment and wage effects of training supported by public income maintenance outside of a firm. After describing the labor market developments in East Germany, we illustrate the evaluation problem. Then, we estimate a simultaneous model for participation in training, employment, and wages. Taking account of selection effects before participation, our findings mostly suggest positive though only partially significant long-run effects of training on employment or wages.

1. Introduction

During the transformation process since unification in October 1990, the East German labor market has undergone fundamental changes. Support of training has been a very important instrument of active labor market policy attempting to increase productivity and to reduce unemployment. This study evaluates the impact of public-sector sponsored traning programs on individual employment probabilities and wages in East Germany from 1990–4. Based on a large panel data set, the 'Labor Market Monitor East', we mostly find positive results of these qualification programs when we account for possible selection effects using a regression-based difference-in-difference approach. In detail, the results show that, on the one hand, these programs were successfully targeted at persons with bad labor market experience but, on the other hand, were not well targeted at persons without any occupationsl certificate—a group which severely suffers from structural change and which is especially exposed to unemployment. Before the start of the program we observe, conditional on our set of regressors, significantly worse employment probabilities and wages of future participants than those of non-participants. After the program the relative position of trainees to non-trainees has improved—but not always significantly so. The literature often finds that participants of training programs have had a particularly bad labor market experience in the period immediately preceding participation. With regard to earnings this finding has been referred to as Ashenfelter's Dip which is notoriously difficult to model; see Ashenfelter (1978) and Heckman and Smith (1995a). Trying to avoid a possible bias emerging from the Dip we ground our interpretation on the

difference between long-run rather than short-run post- and pretraining effects. When we compare the results with those based on slightly different sample compositions or regression models, we observe some variations in the estimated program effects. However, the general positive picture does not vanish.

Our investigation is motivated by the question whether the large labor market programs the German Federal Bureau of Labor initiated in 1990 can be considered successful to combat unemployment and to raise wages. Despite the importance of labor market policies in East Germany and the massive financial transfers involved so far, only a small number of studies has attempted to evalutate the effects of such programs. To illustrate the background, the East German labor market has undergone fundamental changes since unification in October 1990. Unemployment occurred for the first time in 40 years and the structure of the active labor market population has changed considerably. This development induced large financial transfers from West to East Germany, partly through labor market policies. But also labor market policies themselves promoted structural changes in the East German labor market. One of the main instruments of labor market policy in East Germany was the implementation of qualification (training) programs. In the first year after unification, these programs were considered necessary to adjust the qualificational structure of the East Germany labor force to the requirements of a western style economy. A great variety of different institutions and programs for publicly supported occupational training were created, partly by means of new regulations in the work support act (*Arbeitsförderungsgesetz*) which were specifically designed for East Germany. In the meanwhile, the objective of labor market policy has changed and qualification programs in East Germany are apparently considered to be an important tool in fighting unemployment.

The remainder of the paper is organized as follows: Section 2 describes the extent of training in East Germany. Section 3 discusses the general evaluation problem with specific reference to descriptive statistics for East Germany from the Labor Market Monitor and introduces the econometric evaluation approach taken in this paper. Section 4 presents the estimation results. Section 5 concludes and the Appendix contains further estimation results.

2. Training programs in East Germany

The situation in the East German labor market right after the breakdown of the GDR was very untypical for western style economies. (Open) unemployment did not exist, jobs were created or maintained by the authorities to meet political goals rather than for economic considerations, and female labor force participation was very high. In the first years of transition, a large part of the economy was severely hit by the loss of its traditional eastern trading partners, and by the sudden exposure to western competition. Female labor market participation dropped considerably, and early retirment became a frequently chosen way of exiting employment. Despite this decline of the labor force, and despite large flows of

migrants from East to West Germany, the East German unemployment rate rose dramatically, from almost zero in 1989 to 16% in 1994.

In response to the East German unemployment problem, the German Federal Bureau of Labor, the *Bundesanstalt für Arbeit* (BA), implemented all types of active labor market policies at a huge scale. In the beginning, the BA financed mainly short-time work which was the quickest instrument to implement. In 1991, around 1.6m workers were working short-time corresponding to almost 20% of the East German labor force at that time; see German Council of Economic Experts *Sachverständigenrat zur Begutachtung der Gesamtwirtschaftlichen Entwicklung* (1996, pp. 275 and 347). The instrument short-time work was supposed to act only for the first months after unification until other, more productivity oriented instruments could be implemented, namely qualification programs (*Fortbildung und Umschulung*) and job-creation schemes (*Arbeitsbeschaffungsmaoßnahmen*). The institutional setup of these programs turned out to be rather complicated[1] but already in 1990 many persons entered these programs. The number of persons participating in job-creation-schemes or training programs reached a peack at roughly 880,000 participants in 1992, decreased thereafter to 450,000 persons in 1994, and is still at around 340,000 in 1997, see BA *Bundesanstalt für Arbeit*, Statistik–IIIa4–4221. Participation in training programs accounts for more than half of these numbers, and, in this respect, training programs are the most important instrument of active labor market policy in East Germany.

Public sector sponsored training consists mainly of continuous vocational training and retraining (*Fortbildung und Umschulung*, FuU in the remainder).[2] They refer to programs that require completed occupational education or some years of professional experience. Participation in FuU may lead either to an update or upgrade of existing occupational education or to a new occupational certificate that should significantly improve the trainee's labor market prospects. Most often the qualification program takes place in external institutions but it may also be located in the firm where a person is employed. The program can be full-time or part-time, and its duration has to be more than two weeks and is usually restricted to a maximum duration of two years for full-time programs.

Financial support by the BA is only granted, if the participatation in the qualification is supposed to significantly increase re-employment prospects of an unemployed person or to avoid a layoff which has already been announced to an employed person. The latter regulation was eased for East Germany until the end of 1995 assuming a global threat of unemployment. The support may include covering the costs of the program or even covering living expenses (*Unterhaltsfeld*)

[1] Especially the large number of training isntitutions and programs that were supported by the BA in 1990 and 1991 have often been criticized for their lack of concepts and qualified personnel, see also Brinkman and Völkel (1992, p. 268f).

[2] Other programs are: support of vocational training (*Förderung der beruflichen Ausbilding*) or institutional support.

for the participant during the qualification program. The payment of *Unterhaltsgeld*, which in the remainder we refer to as Public Income Maintenance, PIM, covers 67%/60% of a participant's last net monthly earnings with/without children or spouses in need of care. Expenditures for PIM have made up a considerable and increasing fraction of the BA's total budget for occupational training in East Germany. For instance, in 1992, the BA spent 11.28 bn DM for occupational training in East Germany, *cf. Institut der Deutschen Wirtschaft* (1995, p. 134), and more than half of this for PIM (6.0bn DM), *cf. Amtliche Nachrichten der Bundesanstalf für Arbeit* (1995, p. 315).

The importance and size of transfers that were spent on qualification programs motivated a number of microecnometric studies to evaluate some of the effects of these programs, e.g. Hübler (1994b, 1997), Pannenberg (1995), Fitzenberger and Prey (1997), Staat (1997), Eichler and Lechner (1996), and Lechner (1998). Though the empirical evidence presented in the literature is far from conclusive. The evaluations differ with respect to the data set on which the studies are based, the method used to account for a possible selection bias due to program participation, the time period, and the type of training program to be evaluated. Furthermore, the program effects turn out to be positive, not significant, or even negative. Both the diversity of the issue and the question of the adequate selection correction method presumably lead to the contradictory results.

This study presents further evidence on the effects of public sector sponsored training in East Germany while carefully addressing the selection issue. We investigate whether participation in a qualification program leads to better individual (re-)employment prospects and/or to higher wages of participants. Since we are interested in the effects of public labor market policy, we focus on those programs where the participants are supported by PIM because this is the main indicator of public involvement in our data set. To investigate a more homogeneous type of training program we also exclude those PIM-supported programs that take place in the firm (which are not many) as they are likely to be targeted at employed persons, practically oriented, and therefore very different from classroom-education which is typical for the qualification programs analyzed here.

3. The empirical evaluation of training effects

This section discusses the evaluation of training programs supported by public income maintenance as the problem to obtain an adequate estimate for the hypothetical situation of not having participated. Based on the Labor Market Monitor, the methodological issues involved are illustrated by means of descriptive evidence. We then develop our econometric estimation approach building on a simultaneous random effects model for participation, employment, and wages while accounting for attrition. Using variants of the pre-program test introduced in Heckman and Hotz (1989), we suggest a regression-based difference-in-differences estimator of the training effect.

3.1 Evaluation problem

A comprehensive evaluation of active labor market policy (ALMP) has to solve the following problems: does participation in training improve the labor market situation of the participant in a causal way? And does the success of a training program justify the private and public cost involved? This paper is restricted to the first issue. However, it has immediate consequences with regard to the last question since positive effects of training on the individual level (evaluated here) are necessary to render the last question meaningful.

3.1.1 *The data* Our empirical analysis uses the Labor Market Monitor (LMM, *Arbeitsmarktmonitor Ost*) for East Germany. The LMM is a panel survey starting in November 1990 and ending in November 1994. It was commissioned by the research institute of the BA (Federal Bureau of Labor) and covers 14,993 individuals living in East Germany, a representative sample of the active labor market population in 1990. The LMM contains a variety of questions about the individual's social and economic background as well as questions on the firm where a person is employed. Special attention is given to labor market participation, unemployment, and participation in active labor market programs. The first questionnaire also asked retrospectively about labor market status and size and sector of the firm in November 1989 so that some basic information about this year Zero is available as well. The time between each wave varies over the period 1990–4. However, we restrict our investigation to five time-equidistant waves in November each year. Further details on the dataset and comprehensive summary statistics are available upon request.

We evaluate the effects of public sector sponsored qualification programs involving training in an institution, which is not the firm where the participant is employed (QP–OFF $\equiv$ qualification program off the job), and which are supported during participation by public income maintenance (PIM) as part of ALMP. Between 80% and 90% of all participants in QP–OFF receive PIM. It should be noted that individuals are also asked in the dataset whether they participate in private training in the firm and that individuals often participate in more than one type of of qualification program during the time period under investigation. The following numbers illustrate this point based on the sample used for estimation purposes from 1990–4. On the one hand, among the 810 participants in QP–OFF–PIM, 339 individuals (41.9%) also participate in some other type of qualification program either in a firm where employed or without receiving income maintenance. On the other hand, the data comprise a total of 2,849 individuals participating in other types of qualification programs and a total of 4,352 individuals not participating in any qualification program.

3.1.2 *Obtaining an adequate control group* The main criteria considered to evaluate the success of training on the individual level are future employment and future

wages after participation. The estimation of the effects of training on these variables raises difficult methodological questions which are the subject of an intense debate in the evaluation literature, see among others Björklund (1989), LaLonde (1986), Heckman and Smith (1995b). Typically, a descriptive comparison of employment rates and earnings between former participants (treatment group) and the group of non-participants does not allow for causal inference on whether training increased (or decreased) future employment rates or future wages.

In order to evaluate the effect of training properly, one would have to compare the situation of the individual after participation with the situation of the same individual in the hypothetical case of not having participated (potential outcome approach). The evaluation problem consists of obtaining an adequate estimate for the hypothetical situation of not having participated. If employment rates of former participants lie above the estimated employment rates in the hypothetical situation of not having participated, then one can conclude that participation raises future employment rates. In fact, the validity of an evaluation hinges on the validity of the estimate of the success criterion in the hypothetical situation of not having participated. In particular, one has to recognize that ALMP is often targeted to specific groups of people, for intance, temporarily or permanently disadvantaged individuals. At the beginning of the transformation process in East Germany (1990–1), AMLP was aimed at improving the general skill level of the labor force then mainly being employed. Later in the transformation process unemployment increased significantly and then ALMP was mainly aimed at improving labor market prospects of unemployed individuals.

Therefore, it is unlikely that the labor market prospects of the group of non-participants are a good estimate for the labor market prospects of participants in the hypothetical situation of not having participated. Such a comparison is typically subject to a sample selection bias. An adequate evaluation has to find (or define) a control (comparison) group providing an adequate estimate of the success criterion in the hypothetical case of not having participated. Thus, an adequate control group is a group of individuals corresponding to the treatment group with respect to all aspects which influence the success criterion. A sample selection bias in the comparison between treatment and control group can occur with respect to observable and unobservable characteristics, *cf.* Heckman and Hotz (1989).

3.1.3 *Descriptive evidence on preprogram and postprogram comparisons* To illustrate the evaluation problem, we define various groups of individuals at a certain point of time relative to training participation, namely, those who

(i) participated in a qualification program at some time in the past (long-Run postprogram QP-effect: $D^{PostQP}_{Long\text{-}Run} = 1$);
(ii) finished a qualification program since the last interview (Short-Run post-program QP-effect: $D^{PostQP}_{Short\text{-}Run} = 1$);
(iii) will participate in a qualification program in the future and did not do so in the past (Long-Run preprogram effect: $D^{PreQP}_{Long\text{-}Run} = 1$);

(iv) will participate in a qualification program prior to the next interview and did not so in the past (Short-Run preprogram effect: $D^{PreQP}_{Short\text{-}Run} = 1$).

Accordingly, the dummy variables D, defined implicitly above, indicate whether and when the person participated in a qualification program in the past (Post QP) or whether the person will participate in the future (PreQP) and has not done so in the past.

The acronym *PreQP* relates to the preprogram-test advocated by Heckman and Hotz (1989) which involves a comparison of the success criterion for the participants and the control group under consideration before participation. The idea of the preprogram-test is that the success criterion for the group of participants and for the control group should be equal (apart from statistical variation) before participation. Only in this case, is it plausible that the control group provides an adequate estimate of the success criterion for participants in the hypothetical situation of not having participated.

Table 1 provides descriptive evidence on post- and preprogram comparisons between partcipants and non-participants based on estimating

$$y_{it} = \sum_{t=1990}^{1994} \beta_t \cdot DT_t + \lambda_{Pre} \cdot D^{PreQP}_{Long\text{-}Run} + \lambda_{S\,Pre} \cdot D^{PreQP}_{Short\text{-}Run} + \lambda_{S\,Post} \cdot D^{PostQP}_{Short\text{-}Run} + \lambda_{Post} \cdot D^{PostQP}_{Long\text{-}Run} \quad (1)$$

for employment and wages as left-hand-side variables. DT_t denote year dummies and $D.$ represent pre- and postprogram dummy variables for the groups defined above.

Starting with the postprogram comparisons ($D^{PostQP}_{Short\text{-}Run}$, $D^{PostQP}_{Long\text{-}Run}$), both former male and female participants exhibit significantly lower employment rates in the long run compared to non-participants. In contrast, wages do not seem to differ

Table 1 Descriptive evidence on pre-/postprogram comparisons between participants and non-participants by gender*

	Employment				Wages			
	men		women		men		women	
	coeff.	(t-stat)	coeff.	(t-stat)	coeff.	(t-stat)	coeff.	(t-stat)
PreQP-Long-Run	−0.206	(1.3)	−0.220	(2.4)	−0.013	(0.4)	−0.033	(1.4)
PreQP-Short-Run	−0.745	(2.9)	−1.007	(6.0)	−0.015	(0.2)	0.021	(0.3)
PostQP-Short-Run	−0.664	(3.1)	−0.562	(4.3)	−0.220	(3.0)	−0.118	(2.1)
PostQP-Long-Run	−0.575	(4.5)	−0.583	(7.2)	0.042	(1.2)	0.003	(0.1)

* Employment and wages (for the employed) are regressed (for employment by means of a probit regression) on time dummies (omitted here) and dummies for participation in the qualification program (QP-dummies) according to specification (3.1.3) in the main text. The sample consists both of persons not having participated in any qualification program and persons having only participated in QP-OFF-PIM.

between the two groups. Turning to the short-run differences after participation, both employment and wages are significantly lower than in the long run (note that the two coefficients have to be added since $D^{PostQP}_{Short\text{-}Run} = D^{PostQP}_{Long\text{-}Run} = 1$ in the short run), i.e. labor market prospects improve over time after participation. Even though the postprogram situation of participants is not better (and typically worse) than that of non-participants, it is clear that one cannot conclude that participation exhibits no positive (or even negative) effects because participants and non-participants can differ in observable and unobservable characteristics which influence employment and wages.

The preprogram comparisons in table 1 ($D^{PostQP}_{Short\text{-}Run}$, $D^{PostQP}_{Long\text{-}Run}$) indicate that such differences exist. Employment rates of future participants are considerably below the employment rates of the group of non-participants and this effect is particularly strong, when participation starts before the next interview. However, no significant wage differences can be found before participation. Possible explanations for the disproportionate deterioration of the labor market situation just before participation lie in the anticipation of participation via reduced search effort and the targeting of ALMP. Nevertheless, the long-run preprogram estimates indicate that participants and non-participants are not directly comparable. An adequate evaluation has to control both for observable and unobservable differences between the two groups. With respect to observable differences, further descriptive evidence not reported here (available upon request) suggests, for instance, that participants are on average younger than non-participants, are more likely to be females, and exhibit a higher education level. All these varibles are likely to affect the labor market situation of a person independently of participation.

3.1.4 *Nonexperimental approach to resolve the evalution problem* While focusing on an econometric nonexperimental evaluation approach in this paper, it has to be mentioned that a significant part of the evaluation literature advocates controlled social experiments, where participants and control group are drawn randomly from the same group of individuals in order to provide adequate control groups of non-participants, see e.g. LaLonde (1986).[3] However, Heckman and Smith (1995b) show that experiments can also involve a selection bias in the control group, for instance, in form of a substitution bias, when members of the control group, who were randomly rejected from participation, might try to participate in a similar program at another occasion. Since no experimental data are available for East Germany, this paper is restricted to a nonexperimental evaluation.

Based on the discussion in the previous subsection, a natural step is to control for the differences by means of an econometric model. The idea is that an econometric regression function defines an appropriate control group (reference level)

[3] Following this study, a large body of literature assesses the merits and weaknesses of nonexperimental evaluation methods by applying them to experimental data and constrasting the results to the estimates produced by the experiments.

after participation and postprogram dummies as in eq. (1) pick up the differences between participants and control group. At the same time, participation in the qualification program should be modeled itself in order to correct for the sample selection into the program. Applying the aforementioned preprogram test advocated in Heckman and Hotz (1989) by putting the dummy variables $D^{PostQP}_{Short\text{-}Run}$, $D^{PostQP}_{Long\text{-}Run}$ into the regression model allows to investigate whether the econometric model controls for the differences between participants and nonparticipants before participation. Our empirical results show that our econometric strategy is not successful in this respect, since we find significant preprogram effects. We will come back to this issue after presenting the econometric model and we will suggest a regression-based difference-in-differences esimator to make use of this finding.

3.2 Estimation approach

In the following, we develop our nonexperimental estimation approach, which encompasses the methodological considerations in the previous subsection.

3.2.1 *Econometric model* Our empirical analysis models the dummy variables panel attrition, $Y_{it,A} \equiv A_{it}$, participation in qualification, $Y_{it,Q} \equiv Q_{it}$, and employment status, $Y_{it,E} \equiv E_{it}$, and the wage variable $Y_{it,W} \equiv W_{it}$ within a simultaneous random-effects probit and tobit model. The underlying latent variables depend on various regressors and on an individual random effect which is invariant over time and which is supposed to take account of permanent unobserved heterogeneity across individuals. The relationship between the observable dummy variable $Y_{it,j}$ $(j \epsilon \{A, Q, E\})$ for an individual i at time t and the underlying latent variable $Y^*_{it,j}$ is given as follows

$$Y_{it,j} = \begin{cases} 1 & \text{if } Y^*_{it,j} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

and the wage is only observed when the individual is employed resulting in the following

$$W_{it} = \begin{cases} Y^*_{it,W} & \text{if } E_{it} = 1 \\ NA & \text{if } E_{it} = 0 \end{cases} \tag{3}$$

where NA stands for 'not available' and $Y^*_{it,W}$ denotes the wage if employed.

The attrition dummy, A_{it}, indicates whether a complete observation vector of the qualification, for the employment, and for the wage equation is available for individual i at time t ($A_{it} = 1$ available), thus, we allow for the case that an observation is unavailable for one wave but is again available for the next wave. This allows us to keep a considerable number of observations, which seems important given the severity of panel mortality, but it restricts the way how to model attrition (attrition cannot be a function of most lagged variables). The employment dummy, E_{it}, describes the employment status, where $E_{it} = 1$ indicates employment and $E_{it} = 0$ represents all non-employment states which we simply call unemployment.

Obviously, it would be of great interest to differentiate between the different types of employment states. This is, however, beyond the scope of this paper. The wage variable, $W_{i,t}$, denotes the logarithm of the net real wage obtained as the ratio of monthly earnings deflated by the Consumer Price Index and the reported hours worked.

The dummy variable, Q_{it}, capturing participation in the qualification program is modeled as an endogenous variable. The training effect on employment and wages is formally identified since only lagged but not current participation appears as regressor and since the time invariant random effect is assumed to take account of autocorrelation in the selection effects. To preview our results, we acknowledge that our selection correction approach fails since the preprogram tests show that significant differences remain between future participants and non-participants. We will come back to this issue in the next section.

Since people can participate in more than one type of qualification program, we base our analysis on different definitions of the treatment and the control group. Both for the treatment group and the control group, we distinguish whether a person in the respective group participates in other types of qualification programs during the period from 1990–4. Since different qualification programs can interact, thus leading to a contamination bias, and due to the possibility of a substitution bias in the control group (see Heckman and Smith, 1995b), our baseline estimate omits participants in other qualification programs both from the treatment group and the control group. In order to investigate the sensitivity of the results, we also estimate specifications including these types of persons, see Table 3 in Section 4.

It is not straightforward to determine current participation Q_{it} in a qualification program from the data. $Q_{it} = 1$ implies that participation has started or is going to start within the period preceding the next interview, or that the person has participated in qualification since the last interview but it is unclear whether qualification has ended by the time of the current interview. Unfortunately, we are unable to use information on the length of a qualification program. To capture the effect of participation on employment and wages, we introduce the postprogram dummy variables $D^{PostQP}_{Short\text{-}Run}$, and $D^{PostQP}_{Long\text{-}Run}$ as defined in the previous subsection) as regressors into the employment and the wage equation. Thus, we decompose the effect of participation into a temporary and a permanent component.

The model is built up recursively. Given a person has not attrited, $A_{it} = 1$, we model whether the individual participates in the qualification program, and only if the individual is not participating, $Q_{it} = 0$, we consider the employment and wage information. Thus, we allow for three labor market states, employed, $\{E_{it} = 1$ and $Q_{sit} = 0\}$, unemployed, $\{E_{it} = 0$ and $Q_{it} = 0\}$, and participating in qualification, $\{Q_{it} = 1\}$. In the latter case, we discard the information on employment and wages. However, the information on lagged employment status as a regressor is used for those individuals who participated in the previous period. Table 2 summarizes the contributions to the likelihood function where NA indicates that the variable is not considered in the likelihood function.

Table 2 Structure of contributions to likelihood function

A_{it}	Q_{it}	E_{it}	W_{it}
0	NA	NA	NA
1	1	NA	NA
1	0	0	NA
1	0	1	1

The latent variables $Y^*_{it,j}$ are assumed to depend on the set of regressor variables

$$Y^*_{it,j} = X'_{it,j}\beta_j + \sigma_j \cdot \varepsilon_{it,j} \tag{4}$$

where $X_{it,j}$ represents the vector of regressors, β_j the corresponding coefficient vector, $\varepsilon_{it,j}$ the standardized error term such that $Var(Y^*_{it,j}|X_{it,j}) = \sigma_j^2$. We assume the error vector $(\varepsilon_{it,j})_{t=0,\ldots,T;j\varepsilon\{A,Q,E,W\}}$ to be independently and identically distributed over the individuals, $i = 1, \ldots, N$, but we allow the components of the error vector for a given individual to be correlated over time and across equations. As conventional identifying assumption for the probit equations, we set $\sigma_j^2 = 1$ for $j \in \{A, Q, E\}$. The variance σ_W^2 in the wage equation has to be estimated. The correlation across j and t is supposed to be captured by an individual specific random factor α_i such that $\varepsilon_{it,j}$ can be decomposed

$$\varepsilon_{it,j} = \rho_j^{1/2} \cdot \alpha_i + (1 - \rho_j)^{1/2} \cdot u_{it,j} \tag{5}$$

where $u_{it,j}$ is an additional unsystematic random component which is independently and identically distributed over j and t.[4] The coefficient ρ_j (factor loading) is estimated and can be different across j. It measures the impact of the individual effect in equation j by the proportion of the variance which can be attributed to α_i.

We make the following additional assumptions: For all $i = 1, \ldots, N, t, t' = (0), 1, \ldots, 5, t \neq t'$, and $j, j' \in \{A, E, Q, W\}$, α_i, $u_{it,A}$, $u_{it,Q}$, $u_{it,E}$, $u_{it,W}$ follow a multivariate normal distribution and are uncorrelated (i.i.d.) across individuals with $Var(\alpha_i) = Var(u_{it,j}) = 1$ and $Cov(u_{it,j}, u_{it',j}) = Cov(u_{it,j}, u_{it,j'}) = Cov(u_{it,j}, u_{ti',j'}) = Cov(\alpha_i, u_{it,j}) = 0$ for $j \neq j'$.

For the derivation of the likelihood function, it is useful to define $z_{it,j}$ by

$$z_{it,j} \equiv X'_{it,j}\beta_j + \sigma_j \cdot \rho_j^{1/2} \cdot \alpha_i \tag{6}$$

Then, we obtain for the probability that the dummy variable $Y_{it,j}, j \in \{A, Q, E\}$, takes a value of unity

$$P(Y_{it,j} = 1|X_{it,j}, \alpha_i) = \Phi\left(\frac{z_{it,j}}{(1 - \rho_j)^{1/2}}\right) \tag{7}$$

[4] We were also trying to accommodate heteroscedasticity of the unsystematic part of the error term, $u_{it,j}$, in the probit equations by allowing for wave specific variances (only for one wave, the variance was set equal to unity as a necessary normalization). Estimation of such a model proved impossible due to numerical convergence problems.

and for the density of the wage tobit equation

$$f(W_{it}|X_{it,W}, \alpha_i) = (\sigma_W \cdot (1-\rho_j)^{1/2})^{-1} \cdot \varphi\left(\frac{W_{it} - z_{it,W}}{\sigma_W \cdot (1-\rho_j)^{1/2}}\right) \tag{8}$$

where Φ and φ denote cumulative and density of the standard normal distribution.

The derivation of the likelihood function makes use of the recursive nature of the model. The employment and qualification probabilities are allowed to depend on the previous employment status. We restrict the analysis to the employed population in November 1989, thus, we can avoid to model the employment status in period 0, $E_{i,0}$, see Heckman (1981). Since fall 1989 was the 'hour of zero' of German unification, it appears reasonable to assume there is a true initial state. In fact, out of 8,751 persons in our joint samples (men and women), 8,552 were employed in fall 1989.

The one-factor formulation of the dependency in the error term of individual i makes the model tractable, since only a standard one-dimensional integration is involved to obtain the individual contributions to the likelihood function. To evaluate the likelihood function, the random effect α_i is integrated out by means of Gauss–Hermitian quadrature techniques, *cf.* Butler and Moffitt (1982). However, the one-factor model restricts the dependency structure which is allowed for the error vector. Simulation methods, *cf.* e.g. Börsch-Supan and Hajivassiliou (1993), do not seem feasible for our problem at this point due to the complexity of the estimation and the number of observations.[5]

By means of conditioning on the random effect and other endogenous (past and present) dummy variables, the contributions to the likelihood function become simple cumulative normals or normal densities as described in eqs (7) and (8), respectively. The model assumes that the random effect accounts for a potential selectivity bias due to attrition, qualification, and employment in the respective other equations, and that therefore, conditional on the random effect, α_i, we have an i.i.d. error structure, cf. also Lechner (1995, section 3.3). For instance, conditional on the set of regressors, on the individual random effect, and on past employment status, participation in qualification and employment are independent of attrition, i.e., for $k = 0, 1$ and $j \in \{Q, E\}$

$$P(Y_{it,j} = k|A_{it}, X_{it,j}, E_{i,t-1}, \alpha_i) = P(Y_{it,j} = k|X_{it,j}, E_{i,t-1}, \alpha_i) \tag{9}$$

Due to the recursive structure of the model (see table 2), the contribution of the qualification probit can only be used when $A_{it} = 1$, that of the employment probit when $A_{it} = 1$ and $Q_{it} = 0$ and that of the wage tobit when $A_{it} = 1$, $Q_{it} = 0$, and $E_{it} = 1$. Estimation of the asymptotic variance–covariance matrix of the parameters reported in this paper is based on $\hat{V} = H^{-1} \cdot OPG \cdot H^{-1}$, where OPG denotes the

[5] The problem could be mitigated by allowing ρ_j to vary with t. Estimating the baseline model presented in the next section with time varying ρ_j does not change the estimated coefficients in an important way, however, the estimated Hessian cannot be inverted for numerical reasons even though there are no apparent convergence problems. Therefore, we report the results for ρ_j being constant.

outer product matrix of the gradient and H the Hessian of the estimated likelihood function evaluated at its maximum.[6]

3.2.2 *Preprogram test and difference-in-differences estimator* In addition to modeling a separate probit equation for the qualifiction program considered, we test and control for a potential selection bias by putting the short-run and long-run preprogram dummy variables $D^{PostQP}_{Short\text{-}Run}$, and $D^{PostQP}_{Long\text{-}Run}$ into the employment and the wage equation, see Heckman and Hotz (1989). If the model controls properly for selection bias, the estimated coefficients for these regressors must not be significantly different from zero. But, if we find significant preprogram effects, we conclude that, regarding their employment probability or their wage, participants differ from non-participants in a way which is unobservable for the econometrician.

Participants of training programs were often found to have had a particularly bad labor market experience in the period immediately preceding participation. It is very difficult to model these dynamics and to tell them apart from bad transitory shocks hitting non-participants. Thus, it is not surprising that we find significantly negative short-run and long-run preprogram effects implying that econometric conditioning does not define an appropriate control group after participation. Rather than stopping here, we reinterpret the coefficient estimate on the long-run preprogram dummy as a measure of the size of the permanent selection bias, i.e. the amount by which employment or wages of future participants differ permanently from non-participants even after controlling for the set of regressors and an unobserved random effect. Taking the long-run preprogram effect leaves us with a measure of selection bias which is likely to be exogenous to the short-run dynamics.

We suggest that the success of training should be evaluated relative to the coefficient for the long-run preprogram dummy, and not just relative to the value on the estimated regression function. This procedure represents a regression-based difference-in-differences[7] esimator of the impact of training, *cf.* Heckman *et al.* (1997) for a similar approach in a matching context. Analogously to the standard difference-in-differences estimator, where the program effect is evaluated by contrasting before-after-differences of participants and non-participants, the

[6] Alternatively, without being reported, inference is also based on $\hat{V}_1 = -H^{-1}$ and $\hat{V}_2 = OPG^{-1}$, cf. White (1982) and Hübler (1994a). The qualitative nature of the results does not depend on the variance estimate used.

[7] The difference-in-differences estimator is a popular stastistical tool in the evaluation literature, see Card and Sullivan (1988) and Heckman and Hotz (1989). The attractiveness of this estimator derives from contrasting average within-group-differences such that individual specific and time invariant effects cancel, *cf.* Heckmann and Hotz (1989). Analogous to the approach suggested in this paper, the standard difference-in-differences estimator should not use the short-run preprogram situation as the appropriate reference level.

regression function captures the before-after-difference both for participants and non-participants.

Finally, as a caveat on the use of preprogram dummy variables, it can be argued, that we are limited in using preprogram information since we neglect future pariticipation after the end of the observation period, i.e. after November 1994. To investigate this issue, we also estimate our model restricted to the time period from 1990–3 and we check whether the results on the preprogram dummies are sensitive to using the preprogram information for 1994.

4. Estimation results

Based on our econometric approach introduced in the last section, we estimate the employment and wage effects of public-sector sponsored training programs in East Germany. In the following, we will discuss the results of our estimates after briefly describing the sample definitions the various models are based upon.

4.1 Sample definitions

The data used for the evaluation stem from the Labor Market Monitor (LMM) described in Section 3.1.2. The program evaluation requires the choice of a comparison group, which we call the 'naive control group' as it controls for no other difference between trainees and non-trainees apart from treatment. This group consists of those individuals who did not take part in any qualification program during the time span 1989–94 covered by the LMM. A considerable number of persons participating in public sponsored training QP-OFF-PIM, had also participated in another type of training program during 1989 and 1994. To avoid possible sources of bias due to multiple program participation, the trainee sample of our baseline model only considers those persons who participated in QP-OFF-PIM but not in any of the other programs. Table 3 shows the sample composition for the estimated models. The selection of individuals for the baseline model fulfills two conditions: (i) either belonging to the naive control group or having participated in

Table 3 Differences in treatment group and control group across estimated models

	Persons participating in other qualification programs besides QP-OFF-PIM* incuded in	
models	**treatment group (participants in QP-OFF-PIM)**	**control group (non-participants in QP-OFF-PIM)**
(1)†, (3)–(5)	no	no
(2)	no	yes
(6)	yes	no

* training in an institution, which is not the firm where the participant is employed. The participants receive Public Income Maintenance during the training program.

† baseline model displayed in Table 5.

QP-OFF-PIM but not in any other qualification program; and (ii) being employed in November 1989, having answered to the most basic questions in the first wave of the LMM, and not being younger than 18 or older than 55 years. This leaves us with 2,414 men and 2,409 women, among which 146 men and 325 women belong to the treatment group (see models (1), (3)–(5) in Table 3). To control for a possible substitution bias, we also estimated one model with the inclusion of those persons who participated in other qualification programs (but not in QP-OFF-PIM) into the naive control group (model 2). We estimate another model where we augmented the QP-OFF-PIM treatment group by those persons who took part in QP-OFF-PIM and also in other types of qualification programs (model 6) to check for the effects of multiple program participation on the training impact.

4.2 Results

A serious problem with the Labor Market Monitor is panel mortality. For intance, among the 10,7541 persons who had answered the questionnaire in November 1990, only 3,354 individuals still responded in November 1994. Also, several questions were not answered at all or incorrectly by the interviewed individuals. To account for these effects on the parameter estimates, we model panel and answering behavior in a very simple way. Table 5 in the Appendix shows the full estimation results for our baseline sample. It turns out that for both women and men, the propensity of retention is a concave function of age peaking around 48 for men and 41 for women. Furthermore, the highly significant time dummies pick up increasing attrition over time.

Participation in training measures is modeled by the qualification probit. The estimates show that previous year's employment status is a very important determinant for participating in a training program with the trainee receiving public income maintenance by the BA (Federal Bureau of Labor). This is in contradiction to the often stated critique that training programs in East Germany were not well targeted at problem groups. The results indicate a significantly higher propensity to participate for previously non-employed persons. However, the coefficients of the human capital variables show that the probability to participate is significantly higher for individuals with a completed occupational certificate than for individuals without any certificate (the reference group).

Considering the dynamic probit estimates for employment, we find, not surprisingly, strong state dependence in employment. The probability of being employed in the next period is considerably higher when employed in this period. We also find a concave age profile with maxima at 38 years for men and 36 years for women. The results show significant effects of the occupational certificate, but there is no simple increasing relationship between occupational education and employment. Having a master occupational certificate (*Meister*) turns out to affect individual employment chances most positively in the male sample. In the female sample, a university degree is the best basis for individual employment chances. The overall development over time, captured by the time dummies,

shows a decrease in individual employment chances until November 1992 and a marked improvement for men from then onwards but only a small recovery for women.

The estimated wage equations exhibit a familiar concave age profile peaking at around age 37 for women and age 38 for men. Also, the human capital variables show a positive relationship between the occupational education and the individual wage rate. The results on the time effects showing no distinct increase from 1992 onwards are in sharp contrast to aggregate statistics where wages grow strongly over time in East Germany. Apparently, the aggregate wage trend corresponds to a selection of low-wage earners out of employment at the microlevel, i.e. a composition change among the employees. This effect could reflect the increasing importance of individual characteristics which are used when employers decide whether a person is (still) employable at the going wage rate. The wage rate itself was increased more or less exogenously, cf. e.g. Franz (1992).

Turning to the focus of this paper, how does participation in a qualification program QP-OFF-PIM affect future employment and future wages? The estimated training coefficients in the employment and the wage equation are shown in Table 4. For comparison, the baseline model is displayed in column (1). It is based on a sample of the naive control group and those persons who participated in QP-OFF-PIM but not in any other qualification program. Columns (2) to (6) represent the results of slightly different variants of our baseline model. In column (2), the naive control groups is enlarged by persons who participated in other qualification programs (but not in QP-OFF-PIM). Column (3) represents a model which considers the previous period's employment status to affect current wages. The underlying idea is that previous unemployment leads to lower wages in a new job—other things being equal—because of the loss of firm-specific human capital. In columns (4) and (5), we are interested whether the pre-program variables are affected by neglect of future participation after 1994. Thus, in model (4) we estimated our baseline model for the period 1990–3 only but were using the information about program participation in 1994. We compare this with model (5) where we estimated the model for the same period 1990–3 but now ignored the 1994-training information—just as in our baseline model where we do not know anything about individual training after 1994. Finally, for a model shown in column (6) we augment the treatment group by those who also participated in other types of qualification programs aside from QP-OFF-PIM. Here, we want to check whether a contamination bias in the treatment group is present due to multiple program participation.

4.2.1 *Training effects, baseline model (1)* The PreQP-Long-Run Dummy turns out to be significantly negative both in the employment and in the wage equation and for both genders. This dummy measures the difference in employment and wages between trainees and non-trainees even before the start of the program. According to the original motivation for the pre-program test, this would mean, that our

Table 4 Estimated QP-dummies across models* (1) to (6)

			Employment Men - Models			
QP-dummies	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
PreQP-Long-Run	−0.498	−0.576	−0.495	−0.463	−0.454	−0.288
	(3.0)	(3.5)	(3.0)	(2.8)	(2.7)	(2.0)
PreQP-Short-Run	−0.648	−0.697	−0.650	−0.670	−0.719	−0.726
	(2.1)	(2.3)	(2.2)	(2.2)	(2.4)	(3.3)
PostQP-Short-Run	−0.450	−0.412	−0.449	−0.378	−0.379	−0.476
	(2.0)	(1.8)	(2.0)	(1.3)	(1.3)	(2.8)
PostQP-Long-Run	−0.212	−0.294	−0.210	−0.411	−0.395	−0.068
	(1.4)	(2.0)	(1.4)	(2.2)	(2.2)	(0.6)
Net Program Effect†	0.286	0.282	0.285	0.052	0.059	0.220
	(1.3)	(1.3)	(1.3)	(0.2)	(0.3)	(1.2)
			Wage Men - Models			
QP-dummies	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
PreQP-Long-Run	−0.108	−0.115	−0.107	−0.099	−0.098	−0.093
	(3.5)	(4.0)	(3.4)	(3.2)	(3.2)	(3.8)
PreQP-Short-Run	−0.017	−0.026	−0.012	−0.022	−0.018	0.006
	(0.3)	(0.4)	(0.2)	(0.4)	(0.3)	(0.2)
PostQP-Short-Run	−0.132	−0.124	−0.096	−0.229	−0.229	−0.126
	(2.9)	(2.7)	(2.1)	(3.8)	(3.8)	(3.9)
PostQP-Long-Run	−0.064	−0.101	−0.043	0.002	0.002	−0.065
	(1.7)	(2.8)	(1.1)	(0.0)	(0.0)	(2.2)
Net Program Effect†	0.044	0.014	0.064	0.101	0.101	0.028
	(1.0)	(0.3)	(1.4)	(1.8)	(1.8)	(0.5)
			Employment Women - Models			
QP-dummies	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
PreQP-Long-Run	−0.402	−0.480	−0.399	−0.388	−0.388	−0.373
	(3.9)	(4.7)	(3.9)	(3.7)	(3.7)	(4.1)
PreQP-Short-Run	−1.040	−1.086	−1.040	−1.043	−1.132	−0.731
	(5.9)	(6.1)	(5.9)	(5.9)	(6.3)	(5.3)
PostQP-Short-Run	−0.320	−0.305	−0.319	−0.197	−0.197	−0.344
	(2.1)	(2.0)	(2.1)	(1.0)	(1.0)	(3.0)
PostQP-Long-Run	−0.216	−0.287	−0.214	−0.399	−0.388	−0.151
	(2.2)	(3.0)	(2.2)	(3.2)	(3.1)	(1.9)
Net Program Effect†	0.186	0.193	0.185	−0.011	0.000	0.222
	(1.4)	(1.5)	(1.4)	(0.1)	(0.1)	(1.9)
			Wage Women - Models			
QP-dummies	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
PreQP-Long-Run	−0.063	−0.110	−0.060	−0.056	−0.056	−0.048
	(2.4)	(4.4)	(2.3)	(2.2)	(2.1)	(2.1)
PreQP-Short-Run	−0.001	−0.013	0.000	0.000	−0.026	0.054
	(0.0)	(0.3)	(0.0)	(0.0)	(0.5)	(1.4)
PostQP-Short-Run	−0.120	−0.117	−0.082	−0.151	−0.151	−0.116
	(2.5)	(2.4)	(1.7)	(2.3)	(2.3)	(3.3)
PostQP-Long-Run	−0.040	−0.113	−0.008	−0.022	−0.022	0.008
	(1.1)	(3.2)	(0.2)	(0.5)	(0.5)	(0.3)
Net Program Effect†	0.023	−0.003	0.052	0.034	0.034	0.056
	(0.6)	(0.1)	(1.2)	(0.7)	(0.7)	(1.7)

* For choice of control group and treatment group see Table 3. The specification of models (2) to (6) corresponds to the baseline model (1) in Table 5 except for the following differences. (3) with lagged employment status in wage eq. (4) estimation for 1990–3 with preprogram dummy for 1994. (5) estimation for 1990–3 without preprogram dummy for 1994. t-statistics are in parenthesis.

† The Net Program Effect is defined as the difference between the PostQP-Long-Run and the PreQP-Long-Run coefficient.

model does not fully control for selection effects into the program. However, we use this regressor as an indicator of the selection: here, it is obvious that, other things being equal, future trainees in QP-OFF-PIM belong to a group which has significantly lower employment prospects and wages than non-trainees. Since we control for lagged employment status, age, sectoral affiliation, and occupational education, the remaining preprogram difference between trainess and non-trainees must result from other characteristics not observable by us.

Also the PreQP-Short-Run Dummy is negative and significant in the employment equation (but not in the wage equation). Thus, in addition to the long-run preprogram selection effect, there exists also a further drop in the future trainee's employment propensity in the period preceding the qualification program. Surprisingly the short-run preprogram dummy is not significant in the wage equation. Evidently Ashenfelter's Dip is not present with respect to wages regarding QP-OFF-PIM in East Germany. In the period following the program, the short-run PostQP-Effects are significantly negative again both in the employment and the wage equation and for both genders. However, in the long run, the negative employment effect becomes insignificant for men (but not for women). The long-run PostQP-Effect in the wage equation is still significantly negative for men at the 10% level, and for women our results do not show any long-run wage effect.

Central to this study is the re-interpretation of the preprogram variables according to a regression-based difference-in-differences approach; see Section 3. This means that we explicitly use the information of the preprogram dummies as an information about the permanent selection effects of the treatment group and relate the postprogram effects to this selectivity information. Accordingly, the net program effect is given by the difference of the long-run postprogram and the long-run preprogram dummy

$$(\textit{Net Program Effect}) = (\textit{PostQP-Long-Run}) - (\textit{PreQP-Long-Run})$$

Our estimates indicate a positive net program effect. Both for men and women, participation in QP-OFF-PIM (qualification programs in external institution while being supported through the provision of income maintenance by the BA) leads to more fortunate changes in individual employment and wages than employment and wage changes of non-participants—although trainees start from a worse position. We also formally test the significance of this difference. The t-statistics in the last rows of Table 4 show that the net program effect on employment and wages is typically not significant. While the t-statistics for wages are mostly not close to be significant, the numbers for employment (1.3, 1.4) nevertheless suggest that a positive long-run effect is likely to be present. Indeed, it should be noted that it is difficult to disentangle with great precision long-run pre- and postprogram effects based on only five annual observations.

4.2.2 *Sensitivity of the results, models (2) to (6)* When augmenting the naive control group by people who participated in other types of training which are not publicly supported (model 2), the long-run and pre- and post-program dum-

mies become larger (in absolute values) and the long-run post-program dummies (PostQP-Long-Run) become significant. A very intuitive explanation for this is that participants in other qualification programs, especially in those programs that are financially supported by the firm, belong to a positive selection. Then, the difference in employment and wages, even after controlling for observable characteristics, between the negative selection QP-OFF-PIM-treatment group and the comparison group are likely to be larger. However, this should not affect the difference-in-differences result. Indeed, this is the case for the employment effect both for women and men, but the net program effect on wages is smaller than in model (1). Again, the net program effect is positive but not significant as in the baseline model (1). The wage effect is consistent with participation in other types of training inducing higher wages (such a result was part of an earlier version of this paper).

Column (3) reports the result of a model where we included the lagged employment status into the wage equation. The motivation behind is that, after a spell of non-employment, an individual looses her or his firm-specific human capital and cannot expect to be paid the same wage as without the non-employment spell. The results confirm this. Although the long-run effects on wages of the qualification program are still insignificant as well as the net program effect, the difference between the long-run postprogram and the long-run pre-program effect increases. Since it is likely that the fraction of previously non-employed persons among the trainees is very high compared to the naive control group, a negative impact of the lagged employment status on wages might have been captured by the PreQP-Long-Run dummy in model (1). Thus, controlling for lagged employment status in the wage equation renders the net program effect of QP-OFF-PIM on wages larger.

The sensitivity of the results with regard to possible future program participation after 1994 is investigated in columns (4) and (5). Both estimates cover only the period 1990–3, thus excluding the year 1994. However, in model (4), the information about participation in QP-OFF-PIM in 1994 is still used whereas model (5) ignores it as it is done in our baseline model for the years after 1994. The results from Table 4 show that the coefficient estimates for the preprogram variables are very similar in both models. The omission of 1994 seems to change the estimated postprogram effects considerably which mostly reflect compensating changes in the division between short-run and long-run effects. This is neither surprising nor disconcerting for our interpretation of the results since one year less of data reduces considerably the information of the postprogram dynamics. Our point here is only to show that models (4) and (5) do not differ in their preprogram estimates.

The results of our baseline model's last variation are displayed in column (6). Here, the treatment group also includes persons who participate not only in QP-OFF-PIM but also in other training programs. It is striking that this change in treatment-group composition seems to have a different effect on the results by gender. Whereas both the net program effects on employment and wages decrease for men, the opposite is true for women. The difference between the long-run post-

program and the long-run pre-program effects for women turns out to be significantly different from zero. Contrasting this with our baseline model, this points to the existence of a contamination bias. Those men who participate in QP-OFF-PIM and in other types of qualification programs seem to belong to a 'bad selection' compared to those who only participate in QP-OFF-PIM. And the opposite is the case for women: those who also participate in other types of qualification programs seem to belong to a 'good selection' with respect to their employment and wage development. However, the net program effects are quite similar in size to the baseline model and, for women, they become significant.

5. Conclusions

This paper is concerned with the employment and wage effects of public sector sponsored training in East Germany during the transformation period 1990–4. Based on data from the Labor Market Monitor, we illustrate the methodological issues involved when attempting an evaluation study. The evaluation problem consists of finding an appropriate estimate for the hypothetical state of non-participation in training which the effects for participants can be compared with. Using descriptive evidence, we find considerable selection effects, when comparing simple descriptive statistics between former participants and non-participants.

We then estimate a simultaneous model of participation, employment, and wages, but our model is only partly successful in taking account of the selection effects which are present in the descriptive statistics. Ideally, the estimated equation should define the control group, i.e. the level which the program effect can be compared with, given all observable and unobservable characteristics considered. Significant pre-program effects remain in the estimated equations both for wages and employment, even after controlling for observable and unobservable characteristics. We interpret these preprogram effects as the reference level, which the program effect should be compared with.

Our findings indicate that programs involving public income maintenance that take place in external institutions are targeted at persons with bad labor market experience. We mostly estimate positive effects on future employment or future wages but we acknowledge that only few of the estimated net program effects based on a long run pre- and post-program comparison are significant. Nevertheless, when doing sensitivity tests, the positive employment effects appear quite stable across estimated models. Our results differ from findings obtained by Lechner (1996) using a matching estimator based on a different dataset. A topic for future research would be to investigate whether the differences are due to different data sets or different methods.

Acknowledgements

We are particularly grateful for the evaluations by two anonymous referees which led to significant improvements of the paper. We also benefitted from helpful comments by Lutz

Bellmann, Dieter Blaschke, Wolfgang Franz, Michael Lechner, and participants in seminars in Amsterdam, Kassel, Konstanz, and Mannheim. We are also grateful to the Institute for Employment Research (IAB) in Nürnberg for providing the data of the Labor Market Monitor. However, all errors are our sole responsibility.

References

Amtliche Nachrichten der Bundensanstalt für Arbeit (1995). *Jahreszahlen 1994, Bundesanstalt für Arbeit,* Sondernummer, Nürnberg.

Ashenfelter, O. (1978). 'Estimating the Effect of Training Programs on Earnings', *Review of Economics and Statistics*, **60**, 45–57.

Ashenfelter, O. and Card, D. (1985). 'Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs', *Review of Economics and Statistics*, **67,** 648–60.

Björklund, A. (1989). 'Evaluations of Training Programs: Experience and Proposals for Future Research', Discussion Paper No. FS I 89–13, Research Unit: Labour Market and Employment, Wissenschaftszentrum Berlin.

Börsch-Supan, A. and Hajivassiliou, V. (1993). 'Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variables Models', *Journal of Econometrics*, **58,** 347–68.

Brinkmann, C. and Völkel, B. (1992). 'Zur Implementation arbeitsmarktpolitischer Maßnahmen in den neuen Bundesländern', *Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit (IAB), Mitteilingen aus der Arbeitsmarkt- und Berufsforschung*, **3/92,** 260–76.

Butler, J.S. and Moffitt, R. (1982). 'A Computationally Efficient Quadrature Procedure for the One-Factor-Multinomial Probit Model', *Econometrica*, **50,** 761–4.

Card, D. and Sullivan, D. (1988). 'Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment', *Econometrica*, **56,** 497–530.

Eichler, M. and Lechner, M. (1996). *Public Sector Sponsored Continuous Vocational Training in East Germany: Institutional Arrangements, Participants, and Results of Empirical Evaluations*, *Institut für Volkswirtschaftslehre und Statistik. Beiträge zur angewandten Wirtschaftsforschung, N.* 549, University of Mannheim.

Fitzenberger, B. and Prey, H. (1997). 'Assessing the Impact of Training on Employment: The Case of East Germany', *Ifo-Studien*, **43,** 69–114.

Franz, W. (1992). '*Im Jahr danach—Bestandsaufnahme und Analyse der Arbeitsmarktentwicklung in Ostdeutschland*', in B. Gahlen, H. Hesse and H. Ramser (eds), *Von der Plan— zur Martwirtschaft*, Mohr-Siebeck, Tübingen.

Heckman, J.J. (1981). 'The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process', in C. F. Manski and C. McFadden (eds), '*Structural Analysis of Discrete Data with Econometric Applications*', MIT Press, Cambridge, MA.

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

Heckman, J.J. and Smith, J. (1995a). 'Ashenfelter's Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies', Research Report 9505, Department of Economics, University of Western Ontario.

Heckman, J.J. and Smith, J. (1995b). 'Assessing the Case for Social Experiments', *Journal of Economic Perspectives*, **9,** 85–110.

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

Hübler, O. (1994a). '*Schätzen und Testen von Probitmodellen*', *Allgemeines Statistisches Archiv*, **78,** 160–75.

Hübler, O. (1994b). '*Weiterbildung, Arbeitsplatzsuche und invididuelle Beschäftigung—Eine ökonometrische Untersuchung für Ostdeutschland*', *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, **114,** 419–47.

Hübler, O. (1997). '*Evaluation beschäftigungspolitischer Maßnahmen in Ostdeutschland*', *Jahrbücher fur Nationalökonomie und Statistik*, **216,** 21–44.

Institut der Deutschen Wirtschaft (1995). *Zahlen zur wirtschaftlichen Entwicklung in der Bundestrpublik Deutschland*, Deutscher Instituts-Verlag BmgH, Köln.

LaLonde, R. (1996). 'Evaluating the Econometric Evaluations of Training Programs with Experimental Data', *American Economic Review*, **76,** 604–20.

Lechner, M. (1995). *Ökonometrische Modelle für qualitative Paneldaten—Das Beispiel der selbständigen Erwerbstätigkeit in den neuen Bundesländern*, Studies in Labor Economics, **5,** Campus Verlag, Frankfurt and Main.

Lechner, M. (1996). *An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany, Institut für Volkswirtschaftslehre und Statistik. Beiträge zur angewandten Wirtschaftsforschung Nr.* 539–96, University of Mannheim.

Lechner, M. (1998). *Training the East German Labor Force*, Physica-Verlag, Heidelberg.

Pannenberg, M. (1995). Weiterbildungsaktivitäten und Erwerbsbiographie: Eine empirische Analyse für Deutschland, Campus Verlag, Frankfurt and New York.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (1996). *Jahresgutachten 1996/97: Reformen voranbringen,* Metzler-Poeschel, Stuttgart.

Statt, M. (1997). *Empirische Evaluation von Fortbilding und Umschulung*, Nomos Verl.-Ges., Baden-Baden.

White, H. (1982). 'Maximum Likelihood Estimation of Misspecified Models', *Econometrica*, **50,** 1–25.

Appendix

Table 5 Estimated baseline model (1)* for men and women

	men		women	
variable	**coeff.**	**(t-stat)**	**coeff.**	**(t-stat)**
Attrition equation				
Intercept	−1.375	(5.1)	−2.954	(10.8)
Age/10	1.130	(8.0)	2.030	(14.1)
$Age^2/100$	−0.118	(6.8)	−0.246	(13.7)
$DU\,M_{1991}$	−1.000	(28.0)	−0.878	(28.2)
$DU\,M_{1992}$	−1.345	(33.6)	−1.168	(32.4)
$DU\,M_{1993}$	−1.636	(38.9)	−1.375	(36.4)
$DU\,M_{1994}$	−1.751	(40.6)	−1.510	(38.7)
ρ_A	0.638	(40.0)	0.649	(44.6)
Qualification equation				
Intercept	−2.184	(2.3)	−1.686	(2.6)
$E_{i,t-1}$	−1.075	(10.0)	−0.982	(13.2)
Age/10	0.211	(0.5)	0.039	(0.1)
$Age^2/100$	−0.051	(1.0)	−0.036	(0.9)
Facharbeiter	0.366	(1.2)	0.490	(3.2)
Meister	0.465	(1.5)	0.796	(3.8)
Fachschule	0.899	(2.9)	0.416	(2.5)
University	0.811	(2.7)	0.316	(1.6)
$DU\,M_{1991}$	0.679	(4.6)	0.828	(7.0)
$DU\,M_{1992}$	0.743	(5.0)	0.896	(7.1)
$DU\,M_{1993}$	0.339	(2.1)	0.415	(3.1)
$DU\,M_{1994}$	0.237	(1.2)	0.193	(1.2)
ρ_Q	0.051	(1.8)	0.058	(2.1)
Employment equation				
Intercept	−1.963	(4.5)	−2.008	(5.3)
$E_{i,t-1}$	1.214	(11.7)	1.393	(16.6)
$E_{i,t-1} \times Sec_2$†	−0.010	(0.1)	0.209	(1.6)
$E_{i,t-1} \times Sec_3$ †	0.068	(1.0)	0.038	(0.3)
$E_{i,t-1} \times Sec_4$†	0.006	(0.1)	−0.046	(0.6)
$E_{i,t-1} \times Sec_5$†	0.054	(0.8)	−0.238	(3.8)
$E_{i,t-1} \times Sec_6$†	−0.211	(2.5)	−0.153	(2.8)
$E_{i,t-1} \times Sec_7$†	0.168	(1.7)	0.336	(2.7)
$E_{i,t-1} \times Sec_{8,9}$†	0.012	(0.1)	0.046	(0.7)
$E_{i,t-1} \times Public$‡	−0.038	(0.5)	0.085	(1.3)
$E_{i,t-1} \times Short$‡	−0.415	(4.8)	−0.541	(7.0)
Age/10	1.066	(4.8)	1.121	(5.6)
$Age^2/100$	−0.142	(5.2)	−0.157	(6.3)
Facharbeiter	0.284	(3.3)	0.144	(2.2)
Meister	0.406	(3.5)	0.500	(3.2)
Fachschule	0.289	(2.6)	0.382	(4.4)
University	0.150	(1.2)	0.568	(4.4)
Married	0.392	(6.0)	−0.031	(0.6)
$DU\,M_{1991}$	−0.185	(2.2)	−0.348	(5.2)
$DU\,M_{1992}$	−0.649	(7.8)	−0.492	(6.6)
$DU\,M_{1993}$	−0.383	(3.8)	−0.452	(5.7)

Continued

Table 5 *continued*

	men		women	
variable	**coeff.**	**(t-stat)**	**coeff.**	**(t-stat)**
Employment equation				
$DU\ M_{1994}$	−0.330	(3.2)	−0.364	(4.1)
PreQP-Long-Run	−0.498	(3.1)	−0.402	(3.9)
PreQP-Short-Run	−0.648	(2.2)	−1.040	(5.9)
PostQP-Short-Run	−0.451	(2.0)	−0.320	(2.1)
PostQP-Long-Run	−0.213	(1.4)	−0.216	(2.2)
ρ_E	0.144	(4.5)	0.096	(4.2)
Wage equation				
Intercept	2.907	(27.1)	2.893	(27.9)
Age/100	0.197	(3.7)	0.125	(2.3)
$Age^2/100$	−0.026	(3.8)	−0.017	(2.5)
Facharbeiter	0.072	(2.9)	0.004	(0.2)
Meister	0.135	(4.3)	0.085	(1.8)
Fachschule	0.194	(5.5)	0.243	(9.3)
University	0.467	(11.8)	0.501	(13.8)
$DU\ M_{1991}$	−0.016	(1.3)	−0.005	(0.4)
$DU\ M_{1992}$	0.123	(7.0)	0.197	(11.6)
$DU\ M_{1993}$	0.111	(5.7)	0.171	(9.3)
$DU\ M_{1994}$	0.132	(6.2)	0.193	(10.0)
PreQP-Long-Run	−0.108	(3.5)	−0.063	(2.4)
PreQP-Short-Run	−0.017	(0.3)	−0.001	(0.0)
PostQP-Short-Run	−0.132	(2.9)	−0.120	(2.5)
PostQP-Long-Run	−0.065	(1.7)	−0.040	(1.1)
ρ_W	0.069	(2.1)	0.027	(1.7)
σ_W	0.335	(37.6)	0.335	(35.9)

* See Table 3.

† Reference category: agriculture—Dummies are constructed such that coefficients sum up to zero.

sec2: mining, gas, energy
sec3: construction
sec4: metal and electrical manufacturing
sec5: other manufacturing
sec6: commerce
sec7: traffic, post, railways
sec89: banking, insurance, other services

‡ Public: working in the public sector; Short: short-time work