Discussion Paper

Discussion Paper No. 95-27

Effects of Continuous Off-the-job Training in East Germany after Unification

von Michael Lechner

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Labour Economics, Human Resources and Social Policy Series

Effects of Continuous Off-the-job Training in East Germany after Unification

by

Michael Lechner*)

Universität Mannheim and

Zentrum für Europäische Wirtschaftsforschung (ZEW)

December 1995

Address for correspondence:
Michael Lechner
Lehrstuhl für VWL, insb. Ökonometrie
Universität Mannheim
D-68131 Mannheim
Germany
Email: lechner@haavelmo.vwl.uni-mannheim.de

***Acknowledgement**
Part of this paper was written while I was visiting the Center for European Studies (CES), Harvard University, Cambridge, in the academic year 1994 / 1995. Financial support from the CES and the Deutsche Forschungsgemeinschaft (DFG) is greatefully acknowledged. The paper benefited much from ideas disseminated in the lectures by Guido Imbens and Donald Rubin. I thank Klaus Kornmesser for competent help with the data and Irene Bertschek, Bruno Crepon, François Laisney, Markus Pannenberg, Friedhelm Pfeiffer, Viktor Steiner and participants of workshops at the University of Freiberg, Germany and the IFS, London, UK for helpful comments and suggestions on a previous draft of this paper. Furthermore, I thank Viktor Steiner also for supplying a valuable piece of data. All remaining errors are my own.

Abstract

Retraining the labor force to match the demands of a modern economy is an important task during the transition process from a centrally planned to a market economy. This need is particular pressing in East Germany, because the transition process is much faster there than in the rest of Eastern Europe. Therefore, substantial resources are devoted to this purpose.

This paper analyses the impact of continuous off-the-job training in East Germany from the point of view of the individuals who were in the labor force before German unification in 1990. It answers questions about the average gains from participating in a specific type of training. Typical outcomes considered to measure these gains are income, employment status, job security and expected career prospects.

The methodology used for the empirical evaluation is the potential outcome approach to causality. This approach has received considerable attention in the statistical literature over the last 15 years and it has been recently rediscovered by the econometric literature as well. Here, it is adapted to allow for important permanent and transitory shocks that influence the decision to participate in the training as well as future labor market outcomes.

The empirical results are based on the first five waves of the Socio-Economic Panel (GSOEP)-East (1990-1994). This panel data set has the advantage that the fourth wave contains a special survey on continuous training and that it allows to keep track of individual behaviour on a monthly, respectively yearly, basis.

The econometric analysis focuses on off-the-job training courses that began after unification. Although it is obviously too early to evaluate the long-run implications, the results suggest that at least in the short-run there are no positive effects.

1 Introduction

Retraining the labor force to match the demands of a modern economy is an important task during the transition process from a centrally planned to a market economy. This need is particularly pressing in East Germany, because the transition process is much faster than in the rest of Eastern Europe. Therefore, substantial resources are devoted to this purpose, and the need for an evaluation of the results of the work-force training efforts is obvious.

This paper concentrates only on one particular aspect of the training part of the active labor market policy, namely off-the-job training. It tries to identify the average individual gains to the workers of the former GDR participating in off-the-job training between July 1990 and December 1992 compared to the hypothetical state of nonparticipation. Furthermore, the paper addresses the issue whether the gains, if any, are the same for the whole population, or whether there are specific groups of individuals for which they are substantially different. The targets of the evaluations are labor market outcomes after the completion of the training, such as current or expected income, labor market status, and career prospects. It is in the nature of the subject, that when this research was undertaken in 1994/5 only short-run effects could be identified.

In typical evaluations of work-force training programs, outcomes measured for the sample undergoing the training are compared to outcome measures for a *comparable* group, sometimes called control group, that does not get the training. In most social experiments such a group consists of individuals who apply for the program, but are denied participation by randomization, for instance. Such experiments are feasible in some countries, such as the US, but are rejected mainly for ethical reason in others, such as Germany. In an observational study, that is a study not based on experimental data, the researcher should find individuals who are identical to trainees regarding all *relevant* pre-training attributes except for not having obtained the training. Since typically such individuals cannot be easily identified, additional assumptions have to be invoked to adjust for their dissimilarity - in some sense - and avoid potentially serious *sample selection biases*.[1]

Various model-based procedures are suggested in the econometrics' literature in order to avoid such biases. Ashenfelter and Card (1985) and Lalonde (1986), the latter compares different estimates for various nonexperimental control groups with results obtained for an experimental control group, come -among others- to the conclusion that the results are highly sensitive to the different stochastic assumptions made about the selection process. Both papers conclude that the econometric adjustment procedures are unreliable, and hence that social

---

[1] Holland (1986) and Heckman and Hotz (1989) provide extensive and excellent discussions on these issues.

experiments are necessary to evaluate training programs. Yet, on the one hand, even when social experiments are available, evaluations based on them may have other undesirable features.[2] On the other hand, as Heckman and Hotz (1989) correctly observe, the only case you expect adjustment procedures based on different assumptions about the source of the sample selection bias to lead to the same results, is the very case when there is no bias. Consequently, these authors suggest test procedures to chose methods suitable for the particular problem analyzed. Recently, Dehejia and Wahba (1995a, 1995b) - using an approach very similar to the one chosen here - reevaluate the Lalonde (1986) data. They can replicate the experimental results very closely by using nonparametric techniques, partly to be discussed later. It seems that this issue is not yet settled.

Project (or treatment) evaluation and the related need for a definition of causality have a history in the statistics' literature as well. This literature does not put so much emphasis on modeling specific aspect of various distributions. Instead, it stresses the need for nonparametric solutions to the identification problem, and - once it is solved - on nonparametric estimation of the causal effects. Rubin (1974) seems to be the first in explicitly suggesting a model of potential outcomes (outcomes if trained and outcomes if not trained for the same individual). It clarifies the fact that the individual causal effect of training - defined as the difference of the two potential outcomes for example - is never identified. This model and possible necessary identifying assumptions for objects like average causal effects show a close resemblance to the experimental context and emphasize the importance of some sort of randomization as an identifying assumption. It is a useful device to point out that testing methods alone are insufficient, because of a basic lack of identification due to the unobservability of the counterfactual outcome. This has to be overcome by plausible, generally untestable assumption that usually depend heavily on the problem analyzed and the data available. As long as these identifying assumptions do not generate overidentifying restrictions, there is nothing that can be tested, and hence the conclusions by Heckman and Hotz (1989) have to be considered with care.

In the following sections I will try to convince the reader that the prototypical statistical approach - appropriately adjusted for this specific application - is more suited to the particular problem analyzed here than the model-based approaches. Accordingly, the empirical results are obtained by using the

[2] E.g. Manski and Garfinkel (1992), and papers therein (in particular Garfinkel, Manski and Michalopoulos, 1992), but for a forceful defence of experiments see e.g. Burtless and Orr (1986). However, in this particular case the a priori assumption that additional off-the-job training would be benefical was not in question. Therefore, the cost in terms of time 'lost' for conducting an experiment ahead of any program, which thus would have to be delayed for a considerable time, appeared to be prohibitively high.

potential outcome approach to causality as a general framework to define causal effects of off-the-job training on individual actual and expected post-training labor market outcomes. The paper argues that due to the specific situation in East Germany after unification and the rich data at hand, the assumption that the outcomes and the assignment mechanisms are independent conditional on observed attributes, including monthly pre-training employment status, is plausible. Hence, this assumption solves the identification problem inherent in causal analysis.

Since the identification problem is at the center of every causal analysis, the paper contains a considerable part on nonparametric identification of the causal effects in this setting. Nonparametric methods that are direct extensions of the matched pair methods suggested by Rubin (1979) and Rosenbaum and Rubin (1983, 1985) are then used for estimation. I will argue that this approach most probably reduces the bias of the estimated causal effects to a minimum.

The results in this paper do not confirm previous positive findings of the effectiveness of work-force training in East Germany.[3] Although there are only few studies conducted so far, they differ in many respects ranging from the database to the implementation of the evaluation. However, they share two common features that are absent from this work: they do not use an explicit causality framework, and they are based on modeling the distributions of the outcome variables given certain covariates. This paper explicitly avoids imposing these kinds of restrictions in general and puts emphasis of the particular notion of causality behind the results.

The paper contributes to the ongoing discussion of the effectiveness of the training in East Germany by understanding the participation decision as well as by identifying empirically important factors related to it, before obtaining evaluation results for several outcome measures related to the actual and prospective individual position in the labor market. On a methodological side, standard procedures taken from the statistical literature are extended to allow an accommodation of the specific problems encountered in this study and to exploit monthly information on the employment status which could be particularly valuable.

The paper is organized as follows: the following section outlines some basic features of the East German labor market after unification. This significant aspect of the economic environment is important to understand the processes leading possibly to an individual participation in training courses. Additionally, it is also important for the interpretation of pre- and post-training labor market outcomes. Section 3 introduces the longitudinal data used in this study and

[3] E.g. Fitzenberger and Prey (1995), Pannenberg and Helberger (1994) and the references therein.

presents several characteristics of the sample chosen. It is based on the first five years (1990-1994) of the Socio-Economic Panel study for East Germany. All computational aspects of the evaluation are discussed in Section 4, which consists of four subsections. The first subsection details the causality framework used and discusses particular conditions for the identification of average causal effects. The following subsection identifies factors influencing (potential) labor market outcomes as well as training participation. It argues that the respective identification condition is met and discusses the methodology as well as the results of the estimation of a binary choice model for training participation. The third subsection shows that transitory shocks just prior to training, measured on a monthly basis, play an important role for the participation probability. An adaptation of a matching approach is suggested which allows for these important factors to be included in the choice of the control population. The final subsection defines the outcomes, gives details of the suggested nonparametric estimation approach, and shows several aspects of the results. Section 5 concludes. Appendix A contains additional information about the data used. Appendix B illustrates some dangers of misinterpretations of evaluation results based on matches constructed whithout using the monthly unemployment information. Finally, Appendix C consists of several more technical parts concerning the econometric methods.

2 Some features of the East German labor markets

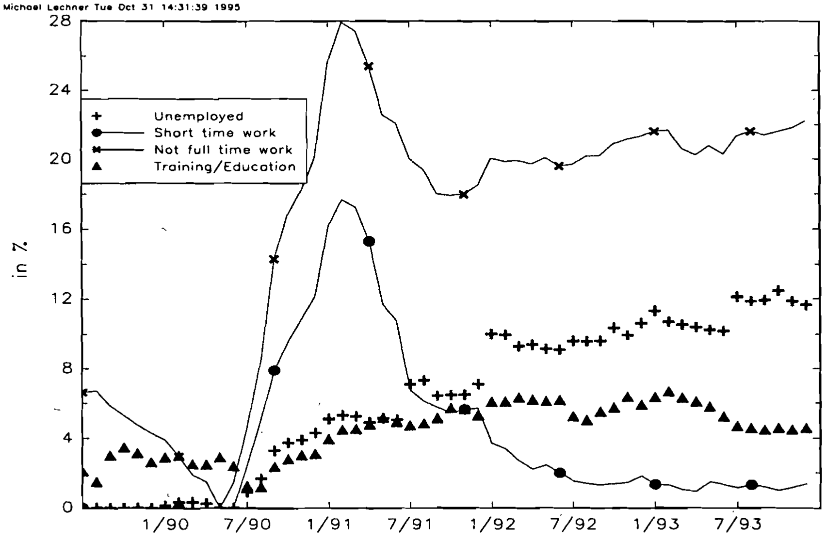
Unification came as a shock to the East German labor markets.[4] The transformation from the previous centrally planned economic system to a West-German-type market economy led to considerable disequilibria in the labor market.[5] Figure 1 shows monthly pre- and post- unification developments for various indicators, such as unemployment, involuntary short-time work (IST, "Kurzarbeit") and full-time employment. Figure 2 depicts gender differences for the sample indicators. Both figures describe the population that I am most interested in: individuals not younger than 20 and not older than 50 (1990). They worked full-time just before unification, lived in East Germany at least until 1994, and are not in bad health conditions. These people constitute the active working population of the late GDR and they are too young to consider (regular) retirement in the next years after unification.

---

[4] This section is based - unless indicated otherwise - on information contained in Statistisches Bundesamt (1994), DIW (1994), Bundesanstalt für Arbeit (1994a, 1994b), Bundesministerium für Bildung und Wissenschaft (1994), and Bundesminister für Arbeit und Sozialordnung (1991).

[5] The sharp decrease in fertility rates well below West German levels is an indication that decisive changes occurred not only in the labor market, but also in many other important aspects of daily life (see Conrad, Lechner and Werner, 1995).

Figure 1: Labor market states



Note: Own calculations based on GSOEP (1990-1994) using panel sampling weights; population is full-time working in June 1990, 20 - 50 years old (1990) and always responding.

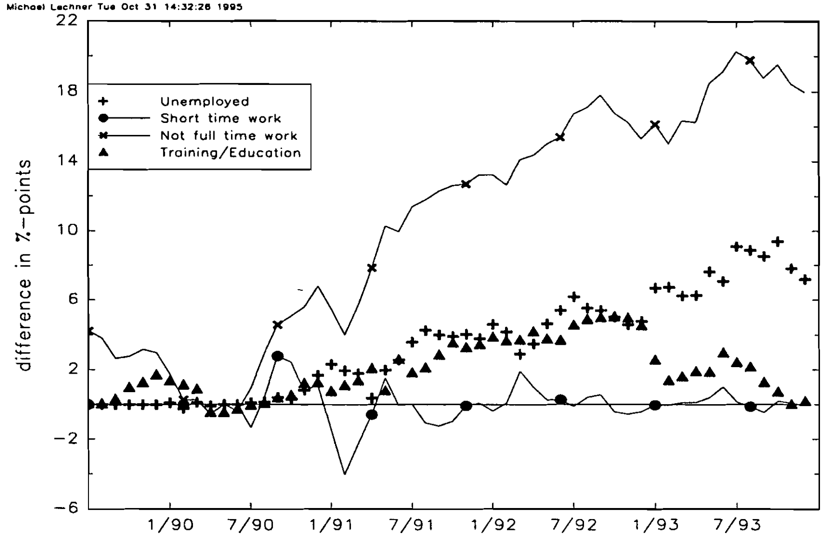
Figure 1 shows that for this population full-time employment (100 minus share not full-time-employed; denoted by ⋇) declines from 100% in mid 1990 to about 70% in early 1991 and than stabilizes at around 80%.[6] A very significant proportion of the early fall is absorbed into involuntary short-time work IST (•), which means a reduction of working hours in the firm accompanied by a subsidy from the labor office to compensate employees for the otherwise occurring income loss. In particular in the first year after unification this reduction of working hours could be substantial.[7] However, IST was only temporarily an important tool of active labor market policy. It was unimportant after 1991. As a result of the decline of IST after early 1991 as well as of the worsening general labor market conditions, the unemployment rate (+) - below 2% before unification (total population) - increased steadily up to about 12 % in the end of 1993.[8] Finally, the number of people taking part in some kind of job

[6] Full-time work includes Make-Work-Programs (ABM) which account for about 5-10% of full-time employment. After the decline of IST, it could be seen a substitute for it.

[7] In the total population in 1991 (1992, 1993) about 56% (48%, 34%) employees on short time work worked less than 50%, and 27% (26%, 23%) worked less than 25% of their usual hours.

training (Δ) also increased steadily after unification. It reached a proportion of about 5% in 1992 (of those full-time employed in 1990) and fell thereafter.

Figure 2: Labor market states: Female - male differences



Note: Own calculations based on GSOEP (1990-1994) using panel sampling weights; population is full-time working in June 1990, 20 - 50 years old (1990) and always responding to survey questions.

Figure 2 shows the difference of the above ratios for women as compared to men. Large differences appear in particular regarding full-time work, but the unemployment as well as the job training rates are significantly higher for women than for men, too. It is also clear that these gender gaps are not just a temporary phenomenon after unification, but it seems that large and permanent differences have emerged. It is perhaps not surprising that women experience more labor market problems than men, because nonparticipation rates in West Germany are much higher than they have been in the GDR and the East German institutional framework after unification is very similar or the same as its West German counterpart. Since unification the East German economy operates under institutional conditions that are very similar to the West German

[8] Unemployment and IST numbers are lower than the official rates, because of the age restriction and because different definitions of the populations appearing in the denominator of the ratios.

institutional arrangements, which are associated with these relatively low participation rates in the West.[9]

While employment went down, wages increased considerably after unification. The yearly average of wages for blue collar workers was about 47% higher in 1993 than in 1991. For white collar male employees the respective increase was about 66%, and for the female employees it was about 59%.[10]

To smooth the transition to a market economy and to adjust the East German stock of human capital to the needs of the new economic system, various levels of the state and its agencies, in particular the labor offices, conducted an active labor market policy. This policy not only provided significant funds for training and retraining opportunities (about 26 bn DM until 1993), but also supplied subsidies for IST (14 bn DM) and make-work-programs (Arbeitsbeschaffungsmaßnahmen, ABM, 26 bn DM). However, a discussion of the latter two policies is beyond the scope of this paper.

The following brief description of the continuous training in East Germany is based on official data from the labor office. Therefore, it concentrates on types of measures that are in some way subsidized by means provided by the Work Support Act (Arbeitsförderungsgesetz, AFG). In particular in East Germany, they form the biggest and most important part of the continuous training and retraining taking place after unification. There are three broad types of training and retraining that are supported: (i) continuous training to increase skills within the current profession (CT), (ii) learning a new profession, and (iii) employers are subsidized for a limited period to provide on-the-job training for individuals facing difficult labor market conditions in order to allow them to familiarize themselves with the new job. The focus of this paper is on the first group, which accounts for about two thirds of all participants in these subsidized courses, but I also include nonsubsidized courses in this area, because with to the available data this differentiation is difficult.

In an increasing number of cases (1991: 53%, 1993: 84%) the labor office does not provide the training, but pays for it, when certain conditions are met. These conditions are related to the employment history, the approval of the course by the labor office, and the potential termination of unemployment or the avoidance of a possibility to become unemployed soon. The last principle has been applied using a broad interpretation in East Germany, so that it includes more groups of the labor force than in the West. The payments cover in most cases almost all the costs for the provision of the course as well as usually more

[9] There are several other issues explaining these gender differences, but they are beyond the scope of this paper.

[10] Source: Bundesministerium für Arbeit und Sozialordnung (1994).

than two thirds of the previous net income. Less than 3% of these courses are provided by the employer.

Table 1: Participants in continuous training (CT) subsidized by the labor office

	Entering CT[1)]		Unemployed before entering CT[1)]	Leaving CT[1)]		Total employment[2)]
year	men	women		men	women	
1990	n/a	n/a	n/a	n/a	n/a	8,820,000
1991	252,352	377,304	51% [3)]	n/a	n/a	7,219,000
1992	201,120	389,896	77 %	151,498	292,590	6,344,000
1993	68,489	113,103	74 %	110,970	202,928	6,128,000

Notes: 1) BA (1994b), 2) Bundesministerium für Sozialordnung (1994);
3) Includes other types of training with higher unemployment shares in 1993;
n/a: not available.

Table 1 shows the number of participants entering and exiting training courses.[11] Women are more likely to participate in CT, mainly because their unemployment probability is higher than for men. Note also the high proportion of people who were unemployed before the start of CT and the dramatic fall in the number of course entrants in 1993. The latter is due to an accumulation of previous entrants who have not yet finished their courses, as well as to a cut in the budget for CT.

3 Data

The sample used for the following empirical analysis is drawn from the German Socio-Economic Panel (GSOEP), which is very similar to the US Panel Study of Income Dynamics (PSID). About 5000 households are interviewed each year beginning in 1984. A sample of just under 2000 East German households was added in 1990. The GSOEP is very rich in terms of socio-demographic information, in particular concerning current and past employment status. The attrition and item nonresponse rates seem to be reasonable low for such a panel study: the attrition rate for the East German sample (1990-1994) is 26% for households and 29.3% for individuals. For a more comprehensive English language description of the GSOEP see Wagner, Burkhauser and Behringer (1993).

[11] Missing entries and the lack of gender-differentiation for 2 columns are due to insufficient data.

A very useful characteristic of this panel survey is the availability of monthly information between yearly interviews. This covers different employment and income states. The information is obtained by retrospective questions about what happened in particular months of the previous year. Figure 3 shows an example for this type of 'calendar' that will figure prominently in the following empirical analysis. Although the monthly calendar contains also questions about training and retraining the level of aggregation of the many different types of training is too high for my purposes. For example, a distinction between on-the-job and off-the-job training is not possible. Therefore, the training information is taken from a special survey on continuous training included in the 1993 survey.

Figure 3: Selected items of the retrospective questions about employment status in the 1993 questionnaire (calendar)

	1992											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
full-time employed												
...												
registered unemployed												
...1)												

Note: 1) Other states include part-time work, vocational training and retraining, education, and out-of-the-labor-force, among others (see Infratest Sozialforschung, 1990, 1991, 1992, 1993,1994).

This special survey contains specific questions about the last three continuous training courses that were either completed in the last three years or are still going on at the time of the interview. The information provided for these courses includes the starting month of the training, the (approximate) duration, the number of weekly hours, its objective, whether it took place during working hours, and finally whether some kind of certificate of participation considered useful for future job applications has been obtained. Considerably more information is provided for the one particular course that the respondents consider to be the most important one for their own professional careers. However, the use of this information in an evaluation exercise could lead to biased results, since the 'unproductive' courses are screened out by the respondents.[12] Additionally, there is another problem related to the use of this special survey: about 19% of training participants attended more than 3 courses.

[12] This could be an empirically important consideration, because more than 60 % participated in more than one course, and of those 47 % stated that all courses were of equal importance.

No information is available on these additional courses. However, I conjecture that the 'lost courses' have been rather short and/or began very early (that is before unification) to fit into the three year time span used by the special survey. Hence, they are unimportant for this study. Another data problem relates to an imprecise measurement of the duration and, therefore, also of the ending date of the training, because there is only categorical information available.[13] In the empirical analysis the monthly durations are computed by using the mid-point of the duration intervals multiplied by the appropriately rescaled hours per week. The computation of the ending dates of the courses uses the end of the duration interval instead, to avoid attributing a part of the training to the post-training period. However, this problem is reduced by combining the information in the calendar variables (Figure 3) with the special-survey variables to adjust the duration and ending dates.

To be able to use the special survey as well as information concerning the employment status in the GDR, a balanced sample of all individuals born between 1940 and 1970 who responded in all four waves is selected. The upper age limit is set to avoid the need of addressing early retirement issues. Since the population of interest is the one that formed the labor force of the GDR, it is required that all selected individuals work full-time just before unification. Furthermore, the self-employed in the former GDR (2%), which form a very different group compared to employees,[14] are not observed taking part in off-the-job training, so they are deleted from the sample. Additionally, individuals reporting severe medical conditions are not considered either, because evaluating the specific kind of training they receive would be beyond the scope of this paper.

Table 2 displays some selected descriptive statistics for those who received off-the-job training (OFT) and those who did not receive it. Table A.1 in Appendix A gives a complete description of all variables used in the empirical analysis. Individuals who did not complete OFT until Dec. 93 are deleted from the sample (see below for further discussion of this issue). The definition of OFT used in this table and all the following empirical analysis is the following: The purpose of the course is qualification other than retraining for a different profession with a duration of more than three months. Its duration is 16 hours or more, or longer than one week. Furthermore, it does not take place during regular working hours (if employed). The purpose of the definition is to obtain a not so heterogeneous group of trainees by excluding very short courses, on-the-job-training and retraining for a different profession. Those are all very different kinds of training with very heterogeneous objectives and very different

[13] Categories: 1 day, up to 1 week, up to 1 month, up to 3 months, up to 1 year, up to 2 years, more than 2 years.

[14] See also Lechner (1993) and Lechner and Pfeiffer (1993).

selection rules. Note that this definition does not exclude the possibility that OFT-participants receive some other kind of training before or after OFT-participation.

Table 2: Descriptive statistics of selected socio-economic variables

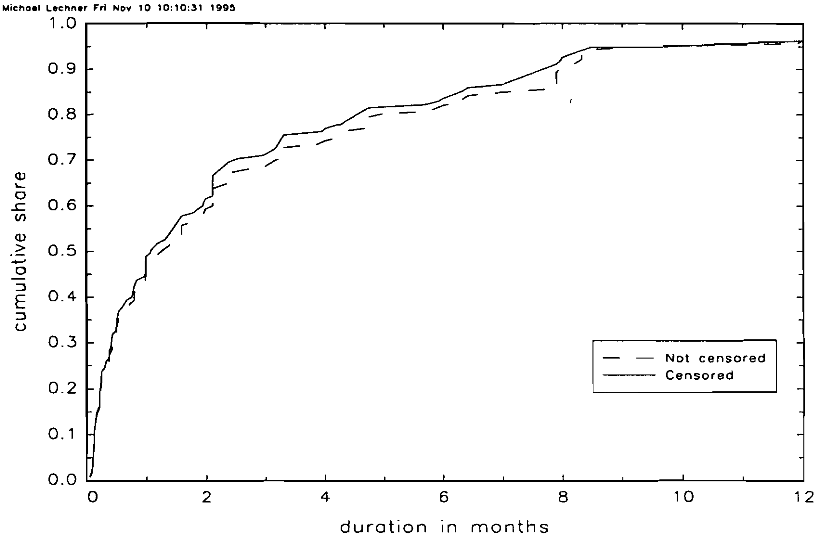
	No OFT (1105 obs.)	OFT (122 obs.)
Variable	mean or share*) in %	mean or share*) in %
Age	35.2 years	35.4 years
Gender: female	42	64
Federal states (Länder) in 1990		
Berlin	7	13
Mecklenburg-Vorpommern	10	6
Sachsen-Anhalt	20	15
Years of schooling (highest degree)		
12	17	31
10	60	63
8 or no degree	22	6
Highest professional degree in 1990		
university	11	25
engineering, technical college	16	33
skilled worker	65	34
Job position in 1990		
highly qualified, management	19	43
skilled blue and white collar	57	40

Note: *) Mean of indicator variable x 100 in subpopulation.

Table 2 shows clearly that OFT-trainees are not a random sample from the population of interest. There does not appear to be a large age difference, but there are far more women in OFT than men. Regarding schooling degrees, professional degrees and job positions in 1990 a very similar pattern appears. Individuals who accumulated more human capital and who reached a higher job position in the former GDR are more likely to seek and obtain OFT. Furthermore, regional aspects seem also to be of importance: Individuals living in East Berlin are more likely to be observed taking part in OFT than for example people living in Sachsen-Anhalt. As Section 5 will show, participants in OFT are also more likely to be unemployed or on IST before the beginning of the course as compared to nonparticipants in the same period of time.

Figure 4 shows the sample distribution function for the duration of OFT. It appears that about 50% of the courses have a duration of one month or less. Only a very small proportion (less than 5%) of the courses last longer than 12 months. Therefore, the typical censoring problem - due to the omission of courses not completed by the date of the 1994 interview - will have in no respect serious consequences for the attempted evaluations (see also Figures A.1 and A.2 in Appendix A).

Figure 4: Empirical distribution functions for durations of training



Note: *Censored* (dashed line) refers to the sample subject to a selection rule that requires the course to be completed by Dec. 1993. The remaining part (12 to 18 months) of the cdf is omitted. For a complete pdf see Appendix A.

The goal of 4 % of the courses was retraining for another profession (maximum duration less than 3 months - otherwise excluded, because different type of training is assumed), another 37 % was intended to qualify for promotion, and 71 % intended to adjust skills to new circumstances.[15] 85 % of the individuals obtained a certificate that they could use when applying for another job. Finally, about 30 % of the OFT participants stated that they were either unemployed or out of the labor force during OFT. Note that this number is less than the official

[15] 19 % had another objective. Numbers add to more than 100, because categories are not exclusive.

unemployment rates reported in Table 1. The reason for this is not entirely clear. It is however very likely that OFT in this sample includes several types of (short) courses that are not funded by the employment office and that could be attended parallel to a job. Also self-reported unemployment in retrospective surveys might be lower than indicated by official records.

As mentioned before, all information about costs to the individual and received subsidies are only available for the one course the individual believes is the most important one for the own career. Nevertheless, the following statistics provide information about these issues. About 16% of the individuals declared that they obtained financial support, such as a continuation of their wage or salary, by their employer. 44% obtained such a financial support from the labor office, whereas about 42% declared that they received nothing. This implies that the definition of OFT used here includes a substantial part of courses that were not subsidized by the labor office. 35% of those participants getting this kind of support would not have participated in OFT otherwise. About 60% had no costs at all for OFT. For those participants who had costs, the median is 300 DM and the mean is 800 DM. 72% paid less than 1000 DM. Another issue that arises in this context is the portability of the acquired human capital when changing jobs. When the individually most valuable course is OFT, 6% of participants state that they acquired nonportable skills and 23% limited portable skills. 39% obtained skills that are portable to a high degree, and 32% acquired completely portable skills.

4 Econometric methodology and empirical implementation

This section begins with a brief discussion of causal modeling and the restrictions that are used to identify the training effects. Subsection 4.2 shows that this identifying assumption is reasonable for the problem analyzed in this study and the data at hand. Then it discusses the estimation and test framework as well as the results of the estimation of the probability of OFT participation. Subsection 4.3 is devoted to specific issues related to the chosen nonparametric estimation approach. Finally, subsection 4.4 contains the econometric methods used for and the results of the actual evaluation. Several technical aspects are relegated to Appendix C.

4.1 Causality, potential outcomes, identification and the propensity score

"What is the average gain for OFT participants compared to the hypothetical state of nonparticipation?" This question is at the center of the empirical analysis of this paper. It refers to potential outcomes or potential states of the world, which never occur. The underlying notion of causality requires the researcher to determine whether participation or nonparticipation in OFT effects the respective outcomes, such as income or employment status. This is very different from asking whether there is an empirical association, typically related to some kind of correlation, between OFT and the outcome. Therefore, I do not try to answer the question whether OFT is associated with a higher income for example, but whether the effect of OFT is a higher income (does OFT cause a higher income in this sense?).[16] The framework that will serve as guideline for the empirical analysis is the potential-outcome approach to causality. Rubin (1974) seems to provide the first explicit suggestion of that framework. This idea of causality is very much inspired by the set-up of experiments in science. Its main building blocks for the notation are *units* (here: individuals, i), for which I will assume that they belong to the large population defined in the previous section, *treatment* (participating in OFT or not participating in OFT) and potential *outcomes*, which are also called *responses* (income, labor market states, either at a particular time, or at a particular span of time after having completed OFT). [17] Y_i^t and Y_i^c denote the outcomes (t denotes treatment, c denotes control, ie. no treatment).[18] Additionally, denote variables that are unaffected by treatments -called *attributes* by Holland (1986)- by X_i. It remains to define a binary *assignment* indicator S_i, which determines whether unit i gets the treatment ($S_i = 1$) or not ($S_i = 0$). If the unit participates in OFT the actual (observable) outcome (Y_i) is Y_i^t, and Y_i^c, otherwise. This notation points to the fundamental problem of causal analysis. The causal effect, for example defined as difference of the two potential outcomes, can never be estimated, even with an infinite sample, because the *counterfactual* (Y_i^t or Y_i^c) to the observable outcome (Y_i) is never observed. However, it is the important contribution of this literature to show under what conditions objects like average causal effects can be identified from a sample of the population.

As emphasized for example by Rubin and others, in order that the model's representation of outcomes is exactly adequate, the *stable-unit-treatment-value assumption* (SUTVA) has to be invoked for all members of the population. SUTVA implies that the value of the potential outcomes for unit i will be the same no matter what mechanism is used to assign OFT and no OFT to unit i and

[16] See Holland (1986) and Sobel (1994) for an extensive discussion of concepts of causality in statistics, econometrics, and other fields.

[17] Since the group aggregated in 'not OFT' is very heterogeneous, the reader may rightly wonder whether a disaggregation would be more informative. Similar considerations apply to the aggregation of the OFT group. In particular for the latter, the current aggregation is mainly driven by sample size considerations. In future work this kind of extension will be attempted with a larger or more specific data set. However, note that the causal estimands can be interpreted as the average effect of the different OFT courses weighted by their distribution in the treated population.

[18] As a notational convention big letters indicate quantities of the population or of members of the population and small letters denote the respective quantities in the sample. The units of the sample (n=1,...,N) are supposed to stem from N independent draws in this population.

no matter what treatments the other units receive (e.g. Rubin, 1986, 1991).[19] Furthermore, there should be no unrepresented treatments. A particular important case of the independence of the outcomes from the assignment is when individuals are randomly assigned to the treatment (*randomization*).[20]

The framework can be seen as a helpful device to design 'informative' social experiments, or - if this is not possible or not desirable - to set up the problem under investigation in such a way that it approximates closely the design of an experiment, and to point out possible departures. Unfortunately, in economic applications there are typically many such possible departures. In this particular case there might be worries that the treatments are too broadly aggregated, because they do take place at different times and have different durations (to be addressed later). Furthermore, there might be interactions between individuals through the market mechanism, because the supply (total number) of treated units should have some impact on their labor market outcomes when trained as well as when not trained.

Another advantage of this approach is that it enforces clear distinctions for three different stages of the empirical analysis: the set-up of the problem using an appropriate notation, the assumptions necessary for the identification of the quantities of interest, and the final estimation stage.

Finally, the potential outcome approach to causality emphasizes the need to explicitly choose a control group and discuss its characteristics. Ideally, members of this control should be like *clones* of the members of treatment group. This means that they should be identical in all aspects effecting the training decision as well as the potential outcomes (technical definitions of similarity appear later). If it is not possible to find such individuals, additional assumptions have to be invoked to - in some sense - adjust for their dissimilarity.

Before briefly discussing more aspects of this framework, a quick comparison with standard econometric approaches is in order. When a typical regression approach is used, based on modeling particular moments of the potential outcomes (e.g. Heckman and Hotz, 1989, Heckman and Robb, 1985, Maddala, 1983), the same issues as mentioned above need to be addressed to make causal instead of associational inference. The wording will then invoke assumptions relating unobserved error terms to regressors. One tends to speak about various sorts of *exogeneity*, *functional forms*, and *distributional assumptions*, etc., to overcome *selectivity* and *endogeneity* problems. I think that this indirect

---

[19] This part of SUTVA can be relaxed in many ways, some of them will be discussed below.

[20] Recently, this framework has also received attention in the econometric literature, e.g. Angrist and Imbens (1992) and Imbens and Angrist (1994).

approach is likely to hide important issues related to the causal or noncausal nature of the intended inference. Furthermore, basing identifying assumptions on unobservables of the assumed models has the 'advantage' of immunifying ones work - at least in some respects - from criticisms. It is generally easier to defend some assumptions on unobserved, unknown and anyhow artificial things like error terms - only in rare cases has the researcher a precise idea what the error term really embodies - than on substantive relationships between important components of the analysis, such as assignment mechanisms. This paper goes the latter route, and, consequently, I hope that it should attract much more informed criticism based on the real problem at hand. Therefore, this criticism can be used in subsequent revisions to obtain more reliable results than the discussion of relationships between unobserved error terms and explanatory variables would ever yield. Finally, another way - perhaps a bit too caricatural - to put some of the aspects mentioned above in perspective is "... the primary justification for model-based repeated sampling inference appears to be its richness of mathematical results rather than its practical relevance" (Rubin, 1991, p. 1225).

Although there is no answer to the question whether a particular individual gains from training, in the following I try to answer questions of the sort "What is the *average* gain for those individuals participating in OFT - or subgroups of them - compared to potential nonparticipation of these individuals?" Using the previous notation the estimand of interest, which is the average causal effect of OFT, is denoted by θ^0 and defined in equation (1):

$$\theta^0 := E(Y^t - Y^c|S=1) = E(Y^t|S=1) - E(Y^c|S=1). \qquad (1)$$

The short hand notation $E(\cdot|S=1)$ denotes the mean in the population of all units i who participate in training denoted by S=1. If the objective is to draw inference only in a subpopulation of S=1, defined by attributes contained in X, then this and the following expressions are changed in an obvious way.

The question now is how this expression can be identified from a large random sample of the population. The problem is the term $E(Y^c|S=1)$, because the pair $(Y_i^c, S_i = 1)$ is not observed for any individual. Much of the literature on causal models in statistics and selectivity models in econometrics is devoted to find reasonable (depending on the problem at hand) identifying assumptions to predict the unobserved expected nontreatment outcomes of the treated population by somehow using the observable nontreatment outcomes of the untreated population. If participation in OFT would have been decided by a random number generator (random assignment), then the potential outcomes would be independent from the assignment mechanism and it would be true that $E(Y^c|S=1) = E(Y^c|S=0)$. In this case the untreated population could be used as control group, which implies that the expectations of their observable outcome would be equal to $E(Y^c|S=1)$. Given a large enough sample, the corresponding sample moments converge towards these population moments under standard regularity conditions. However, a brief look at Table 2 shows that the assumption of random assignment is hardly satisfied. There appear to be several variables which influence assignment as well as outcomes (gender, schooling, etc.).

Using the law of iterated expectations to rewrite the crucial part of equation (1) as:

$$E(Y^c|S=1) = E[E(Y^c|S=1, X=x)|S=1], \qquad (2)$$

leads to another identifying restriction, called random assignment conditional on a covariate (Rubin, 1977). The assumption is that the assignment is independent of the potential outcomes conditional on the value of a covariate or attribute (CIA). If this assumption is true, then $E(Y^c|S=1, X=x) = E(Y^c|S=0, X=x)$, and the quantity $E[E(Y^c|S=0, X=x)|S=1]$ $[= E(Y^c|S=1)]$ can be estimated in large samples using respective sample analogues. Note that the outer expectation operator is with respect to the distribution of X in population of participants (S=1). The next section will show that this powerful restriction is reasonable in the context under investigation. The important task will be to identify (and observe) all variables that could be correlated with assignment and potential outcomes. This implies that there is no important variable left out which influences outcomes as well as assignment given a fixed value of the relevant attributes.[21] There are many different possible other restrictions (e.g. Angrist and Imbens, 1991, Imbens and Angrist, 1994, Heckman and Hotz, 1989, Heckman and Robb, 1985), but this one appears to be the most fundamental in its close resemblance of the experimental context.

Rosenbaum and Rubin (1983) showed that if CIA is valid, then the estimation problem simplifies further. Let $P(x) = P(S=1|X=x)$ denote the propensity score that is defined as the nontrivial probability ($0 < P(x) < 1$) of being assigned to the treatment conditional on the possibly high dimensional vector of characteristics x. Furthermore, let b(x) a function of attributes such that $P[S=1|b(x)] = P(x)$, or in their words, the balancing score b(x) is at least as 'fine' as the propensity score. Their most important result is that if the potential outcomes are independent of the assignment mechanism conditional on X=x,

[21] In the language of regression-type approaches such a variable would lead to simultaneity bias.

then they are also independent of the assignment mechanism conditional on b(X)=b(x), hence:

$$E[Y^c|S=1, b(X)=b(x)] = E[Y^c|S=0, b(X)=b(x)]. \tag{3}$$

Hence, $E(Y^c|S=1) = E\{E[Y^c|S=0, b(X)=b(x)]|S=1\}$ can be used for estimation. The major advantage of this property is the reduction of the dimension of the (nonparametric) estimation problem. The disadvantage is that the probability of assignment - and consequently any balancing scores that reduce the dimension of the estimation problem - is unknown to the researcher and has to be estimated. However, this estimation may lead to a better understanding of the assignment process itself. Details of this estimation are relegated to Section 4.2.2. That section will also discuss a particular form of a balancing score 'finer' than the propensity score that is especially useful for the specific problems encountered in this evluation study.

4.2 Estimation of the propensity score

4.2.1 Variables potentially influencing the training decision and outcomes

Variables that might influence the decision to participate in OFT as well as future potential outcomes should be included in the conditioning set X and, therefore, in the propensity score to avoid biased estimates of the causal effects. Variables only influencing the participation decision may also be included to increase efficiency. To judge what variables this might be, it is necessary to have a definition of OFT (see Section 3) as well as of the potential outcomes. Typical outcomes considered are gross monthly income for individuals employed or unemployment benefit for the unemployed, employment status, such as full-time employment, unemployment, involuntary short-time work, expected unemployment and expected changes in job positions in the next two years. Two concepts of timing are used for these outcomes, which specify either a date or a specific time span after the completion of the course (see Section 4.4.1 for details).

In the following, I identify reasons for participation in OFT by supposing that individuals maximize some sort of future utility, or more precisely, the difference between the present values of future income streams for both states. It seems plausible that at least factors influencing both income and participation in OFT can be identified in this fashion. It is not necessary to develop any formal behavioral model in any detail. Considering the broad building blocks of such a model is sufficient to identify potentially important attributes.[22] In principle one would like to condition directly on these expected income (utility) streams, but since they are unobserved, they have to be decomposed into the cost of OFT and the additional returns of OFT. These factors have to be uncovered, because they are potentially important determinants of the training decision.[23]

There are at least two hypotheses why income with OFT should be higher than without it, everything else being equal. First of all, the additional human capital should increase individual productivity and, therefore, workers should be able to obtain higher wages. Secondly, OFT can act as a signaling device for an employer who has incomplete information on the worker's productivity. Participation in OFT might signal in particular higher motivation, and the successful completion of longer OFT courses may also signal higher ability, and hence the employer may be prepared to compensate for the expected higher productivity. In the first case the additional human capital will yield returns - ignoring effects on pensions - until retirement, or until it is depreciated. This implies at least for older individuals that the remaining period until retirement could be smaller than the depreciation period for the human capital. Therefore, age should not increase the participation probability, but should most likely decrease it. The magnitude of the effect of age under the signaling hypotheses depends crucially on the ability of the employer to learn quickly the true productivity of the worker. When the signal is too positive, employers will try to adjust wages towards true productivity, et vice versa. People sending the 'wrong' signal will only gain a temporary advantage until the employer understands their true productivity. However, by getting employed due to a too positive signal, they may still obtain additional experience that may increase their income as well as employment prospects until retirement. This implies again a negative impact of age on OFT participation.[24] Another factor is how the individual subjectively estimates the own future income streams. For this analysis it is not so important to formulate the exact type of expectation formation as long as it is known what kind of subjective expectations about the own labor market prospects the individual holds. Fortunatelly, this information is available on a yearly basis on the GSOEP.

---

[22] For an introduction in this field of labor economics the interested reader is referred to any modern text book, such as Ehrenberg and Smith (1994).

[23] Note that for these considerations, it does not matter how the labor market really works, but how the individual (and/or the labor office) believes it to work. There might be substantial differences between actual and expected outcomes, when considering that individuals are used to the rules of the command type economy of the former GDR. Furthermore, the high speed of changes after unification makes correct predictions difficult.

[24] The only qualification is that older individuals could in principle retire before the employer learns their true productivity, so that for these people P(x) should not decrease with age.

It is useful to divide the potential costs of OFT for the individual in two broad groups: direct costs and indirect or opportunity costs. Potential direct costs depend mainly on the availability of subsidies (see Section 2 for details). Although direct costs should in principle not have much influence on future outcomes, the labor office tends to give subsidies to individuals with comparatively low (nontraining) labor market prospects, as estimated by the labor office. Therefore, there may be an important indirect effect of the labor market prospects on the potential outcomes through the potential costs. The opposite reasoning applies to employer-sponsoring, which, however, is not important for OFT. Opportunity costs basically consist of lost income and / or leisure. Since the marginal utility of leisure should be lower during non-full-time work (a larger amount is available), the actual labor market status can be an important factor for its own. It may also differ across individuals according to tastes, as well as other socioeconomic factors such as marital status, or the perceived actual (present) utility of time spent in training. The labor office, as well as possibly an employer, provides subsidies to make up most of the foregone earnings under similar conditions that apply to direct costs, so that the same reasoning as before is appropriate.

The above analysis has identified age, labor market prospects, actual labor status, and other socioeconomic characteristics as major factors that could potentially influence the employment decision. Before going in more details about the groups of variables used in the empirical analysis, I will discuss more fundamental issues concerning the admissibility of variables in the conditioning set. Additionally, I will state two assumptions which are very important in that respect for the particular situation in East Germany after unification, because they make CIA a powerful and justifiable assumption in this context.

From the discussion in the previous section the difference between attributes that cannot be influenced by the treatment and outcomes should be clear.[25] It should be also clear that the conditioning variables should be attributes (which could include the expected potential outcomes if they were observed) in order to obtain unbiased estimates of the causal effects. These variables do not change over time, change over time independently from the treatment, or they are dated before any action is taken regarding training participation. The latter point is important: consider an employee accepting a job that pays less than a comparable job with another firm, but offers the possibility of obtaining employer sponsored OFT. The pre-training income on this job cannot be considered as an attribute or exogenous variable, because it already contains an effect of the future treatment. Conditioning on this kind of pre-OFT variable will in this case almost certainly lead to an upward bias in the estimate of the effect of OFT. It is this kind of reasoning that lead to doubts of the exogeneity

[25] The former would be called *exogenous variables* in regression language.

of many job-related variables. Thus, it could make CIA an untenable assumption in many cases when long term planning is involved (this might be conjectured for OFT in West Germany for example).

However, the specific situation in East Germany before and after unification makes CIA a far more plausible assumption. The first hypothesis is that the complete switch from a centrally planned economy to a market economy in mid 1990, accompanied by a completely new incentive system, invalidates such long term plans. It was generally impossible for East German workers to predict the impact and timing of this system change.[26] Even when it was partly correctly foreseen, it was generally impossible to adjust behavior adequately in the old system. This assumption, which seems to be highly realistic, allows me to use all pre-unification variables as attributes.

An additional assumption will be invoked which is related to the condition of the labor market in the rapidly contracting East German post-unification economy. Figure 1 and 2 show that the labor market is characterized by rapidly and continuously rising unemployment as well as declining full-time employment. Furthermore, only about 10% of those working full-time in mid 1990 were sure that they might not lose their job within the next two years. I assume that no individual - having only slim chances of getting rehired once being unemployed - will voluntarily give up employment to get easier access to training funds (which may not even be necessary before 1993, given the official guidelines for obtaining assistance from the labor office). This assumption allows me to consider monthly pre-training information on full-time employment, involuntary short-time work and unemployment as attributes. Additionally, a pre-training change to self-employment is assumed not to be done to obtain training. The risk of self-employment is far too high to be plausible to occur in order to obtain such a comparatively small gain. Therefore, pre-training self-employment, which is measured on a yearly basis, is considered an attribute.

Given the institutional framework outlined above, it is tempting to include the latest pre-training expectation about future job-security in the set of attributes. On the one hand, this variable may very well capture threats to the current employment - important to obtain (AFG-) subsidies from the labor office - and, therefore, it may be considered to be an attribute. On the other hand it may also be considered to be an outcome: For instance, an employer may offer an employee a future training possibility. This will change the expectation of the employee, and the employee will now assume that the current job is safe. Therefore future OFT alters pre-training expectations, which could no longer be

[26] This is even more plausible when one considers the huge prediction errors of West and East German experts with respect to the impact of these changes.

considered an attribute. Hence, they cannot be used as conditioning variables in this framework. Fortunately, Figures 13 and 14 below strongly suggest that this is not a problem in the context considered, because there does not appear to be any significant differences between OFT and the choosen control population.

The groups of variables that are used in the empirical analysis to approximate and describe the above-mentioned four broad categories of determining factors are age, sex, marital status, educational degrees as well as regional indicators. Features of the pre-unification position in the labor market are captured by many indicators including wages, profession, job position, employer characteristics such as firm size or industrial sector, among others. Individual future expectations are described by individual pre-unification predictions about what might happen in the next two years regarding job security, a change in the job position or profession, and a subjective conjecture whether it would be easy to find a new job or not. Details of the particular variables, which are mostly indicators, as well as their means and standard errors in the treatment and control group are contained in Table A.1 of Appendix A. Furthermore, monthly employment status information, as mentioned before, is available from July 1989 to December 1992.

Having discussed potentially important factors and variables available for the empirical analysis, the question is whether some important group of variables might be missing. One such group can be described as motivation, ability and social contacts. I approximate these kind of attributes by the subjective desirability of selected attitudes in society in 1990, such as 'performing own duties', 'achievements at work', and 'increasing own wealth', together with the accomplishment of voluntary services in social organizations and memberships in unions and professional associations before unification, as well as schooling degrees and professional achievements. Additionally, there are variables indicating that the individual is not enjoying the job, that income is very important for the subjective well-being, that the individual is very confused by the new circumstances, and optimistic and pessimistic views of general future developments. Another issue is the discount rate implicitly used to calculate present values of future income streams. I assume that controlling for factors that have already been decided by using the individual discount rate, such as schooling and professional education, will be sufficient. Other issues concern possible restrictions of the maximization problem, such as borrowing constraints, and a limited supply of OFT. Borrowing constraints can be a serious issue, but there seems to be no sure way to find out using this data, because it does not contain information on monetary wealth.[27] Furthermore, it

[27] Property wealth is not informative, because the ownerships of many properties were unclear due to claims of former owners after unification. Therefore, they would be very difficult to sell or to use as collateral for loans.

seems reasonable to assume that OFT supply is concentrated in larger cities. Unfortunately, the only regional information available refers to the six federal states and information on the number of inhabitants of cities and villages, but no the distance to the nearest larger city. However, it is possible that some supply factors as well as information about the availability of OFT could be captured by the indicators for memberships in unions, professional associations and cooperatives. I conclude that, although some doubts could be raised, it seems safe to assume that these missing factors (conditional on all the other observable variables) play only a minor role.

Finally, empirical papers analyzing training programs in the US point to the importance of transitory shocks before training, partly because of individual decision, partly because of the policy of the program administrators. Card and Sullivan (1988) find a decline in employment probabilities before training. Here, the monthly employment status data should take care of that problem. Ashenfelter and Card (1985) observe a decline in earnings prior to training. As will be shown in Section 4.4, there is no evidence of this phenomenon in the sample used here. This could be probably due to the short-time span between the start of OFT and unification.

4.2.1 Econometric considerations

The estimation of the propensity score is not straightforward, because there are potentially important variables - monthly pre-training employment status and yearly pre-training self-employment - which are related to the months or years before the beginning of OFT. Since these dates differ across OFT participants, they are not clearly defined for the control group. An approximation, which might be appealing at first sight, is to choose an arbitrary date for the controls and compute the value of these variables regarding this date. However, having the same date for all controls and different dates for the OFT participants leads to a dependence of this variable on OFT participation, the dependent variable. This dependence is aggravated by the rapidly changing labor market conditions. Therefore, such a variable cannot be considered an attribute or an exogenous variable, so that a probit estimation would lead to inconsistent estimates of the propensity score. Consequently, I have to use a particular form of a balancing score that is different from the propensity score for the conditioning.

Partition the vector of observed attributes in two groups such that $X=(V,M)$, and suppose that $P(S=1|X=x)= P(x)= P[V\beta^0 + f(M,U)>0|V=v,M=m]$. U denotes some attributes not included in X, that are independent of the potential outcomes, but influence OFT participation. V contains pre-unification as well as time invariant attributes. β^0 is a fixed parameter vector. M denotes time variant pre-training variables. If the potential outcomes are independent of S conditional on $P(X) = P(x)$, then it is also true that they are independent of S conditional on $(V\beta^0 = v\beta^0, M=m)$, because $(v\beta^0, m)$ is a balancing score. Note that the use of $v\beta^0$ instead of v can still lead to a dramatic reduction of the dimension of the conditioning set. The rest of this section discusses consistent estimation of $v_n\beta^0, n=1,...,N$, up to scale (and a constant that does not vary in the population). The application of the 'conditioning', along with statistics showing the importance of M, is referred to Section 4.3.

In the following I estimate a conventional binary probit model by maximum likelihood. The basic condition for the consistent estimation of the linear index up to scale is that the conditional expectation of the dependent variable is correctly specified:

$$P(S=1|V\beta^0 = v_n\beta^0) = \Phi(v_n\beta^0), \qquad n=1,...,N. \tag{4}$$

$\Phi(v_n\beta^0)$ denotes the cumulative distribution function of the standard normal distribution evaluated at $v_n\beta^0$. The first of two sufficient conditions for equation (4) to hold is that the propensity score has the additive form $P(x) = P[V\beta^0 + f(M,U) > 0 | V=v, M=m]$. This assumption is not so restrictive, because V may contain flexible functional forms for the attributes, such as polynomials or interaction terms. The crucial assumption is that:

$$[f(M,U)|[V\beta^0 = v_n\beta^0] \sim N(0,1). \tag{5}$$

$N(0,1)$ denotes the normal distribution with mean 0 and variance 1. Neither the assumption of mean zero nor of unit variance is a problem, because required identification is only up to scale and location. The crucial assumptions are normality and independence with respect to $V\beta^0$. Conditional homoscedasticity (implied by independence) and normality is tested using conventional specification tests (similar to Bera, Jarque, and Lee, 1984, Davidson and MacKinnon, 1984, and Orme, 1988, 1990) described and applied in Blundell, Laisney, and Lechner (1993) and in Lechner (1995).[28] A second way used to get indications for the independence of $V\beta^0$ and M is to compute empirical correlations of the estimated index and the observable m_n. Furthermore, the consistency property of the specification tests, in particular of such omnibus tests like the information matrix test will eventually detect any other dependence of $V\beta^0$ and f(M,U).

[28] The use of semiparametric methods, such as SNP estimation suggested by Gabler, Laisney and Lechner (1993) has been considered. However, it is not necessary, because the specification tests indicate no violation of the distributional assumptions necessary for the probit model.

4.2.2 Results

Table 3 presents the results of the probit estimation and the specification tests.[29] All variables that are not contained in Table 3, but described in Table A.1, as well as different functional forms for the (approximately) continuous variables, and interaction terms between *Gender* and variables related to job position and education, are subjected to score tests against omitted variables. None of them appears to be significantly missing at the 4% level. Most results are above the 10% level.

Table 3: Results of the estimation and the specification tests for the participation probit

	estimation		heteroscedasticity test	
Variable	coef.	std.err.	$\chi^2(1)$	p.-val.
Sex: female	0.14	0.14	0.3	62
Federal states (Länder) in 1990				
Berlin	0.34	0.19	2.2	14
Mecklenburg-Vorpommern	-0.39	0.18	1.6	20
Years of schooling (highest degree) in 1990				
12	0.27	0.27	4.9	2.8
10	0.43	0.17	3.9	4.7
Highest professional degree in 1990				
university	0.07	0.35	0.8	38
university and female	0.74	0.29	0.6	44
engineering, technical college	0.31	0.18	2.0	15
master of a trade / craft	0.47	0.20	0.0	97
Job position in 1990: highly qualified, management	0.24	0.20	0.1	71
Job characteristics in 1990				
real wage or salary per month / 1000	-1.95	0.76	0.9	33
ln (real wage or salary per month)	2.95	1.37	1.2	27
temporary job contract	-0.10	0.29	1.2	27
training (unspecified) while full-time employed	0.40	0.16	0.2	67

Table 3 to be continued...

[29] A table for the tests against missing variables is omitted for reasons of space. The results are available on request from the author.

Table 3: Results of the Estimation: continued

	estimation		heteroscedasticity test	
Variable	coef.	std.err.	$\chi^2(1)$	p.-val.
Profession in 1990 (ISCO)				
scientific, technical, medical	-0.25	0.17	1.2	27
production	**-0.74**	0.17	0.0	99
services, incl. trade, office	-0.26	0.16	0.3	60
Employer characteristics in 1990: industrial sector				
agriculture	-0.52	0.29	0.5	50
mining	-0.71	0.44	0.5	48
heavy industry	-0.56	0.31	0.5	49
light industry, consumer goods, electronics, printing	-0.19	0.26	0.0	86
machine building and vehicle construction	0.01	0.28	8.5*)	0.4*)
construction	-0.30	0.31	0.2	65
trade	**-0.66**	0.31	0.2	68
communication, transport	**-1.05**	0.39	0.2	65
other services	-0.42	0.27	0.4	51
education, science	-0.43	0.28	0.1	71
health	**-0.63**	0.29	3.0	8.4
Optimistic about the future in general in 1990	**0.29**	0.13	0.1	82
Expectations for the next 2 years in 1990				
redundancies in firm: certainly not	-0.37	0.28	0.7	39
Other specification tests		$\chi^2(df)$	df	p.-val.
Score test against nonnormality		0.3	2	88
Information matrix test				
All indicators		440	394	5.3
Only main diagonal indicators		43	30	6.1

Note: **Bold** letters: t-value larger than 1.96. N = 1339. (1299 controls)
*) Different covariance estimates lead to very different conclusions.

The t-values and score test results against heteroscedasticty presented in Table 3 are computed using the GMM (or PML) formula given in White (1982).[30] The information matrix tests statistics are computed using the second version suggested in Orme (1988) that appeared to have good small sample properties.[31] Cases when other ways of estimating the covariance matrices of the tests lead to very different results are marked by an asterisk.

The situation in East Berlin - now part of a single federal state with West Berlin - is quite different to the situation in the rest of East Germany. On the one hand, there is easier access to already existing OFT facilities in West Berlin, and on the other hand, the direct and almost immediate exposure of East Berliners to the Western system with the more adequately qualified Western workers may have increased the pressure to obtain additional qualifications. Furthermore, the skill composition of the population differs somewhat from the rest of the country, because East Berlin was the capital and the admistrative center of the former GDR. Therefore, it is not surprising that living in East Berlin is a (weakly) significantly positive factor for OFT participation compared to the other federal states, but in particular compared to northern state of Mecklenburg-Vorpommern.

The differences due to gender and education manifest themselves c.p. basically through a large and significantly higher conditional participation probability for the relatively small group of women with university education (5% of the sample). Other female-education and female-job-position interaction terms are not significant. Furthermore, the reason for the insignificance of *12 years* of schooling, which is the university entrance requirement, may very well be due to its high correlation with professional degree *university* and the respective female interaction term. Taken together, the results in the first part of Table 3 suggest that having a low educational and professional level in the former GDR reduces the probability of OFT participation. This finding is confirmed by the significantly positive effect of a high job position.

---

[30] Five versions are computed: based on the matrix of the outer product of the gradient (OPG) alone, on the empirical hessian alone, on the expected (under the null) hessian alone, and on combining the hessian, respectively the expected hessian (under the null), and the OPG. Previous Monte Carlo studies (e.g. Davidson and MacKinnon, 1984, Lechner, 1991) as well as theoretical papers (e.g. Dagenais and Dufour, 1991) show that tests based on the latter at least avoid some undesirable properties which can occur with other versions (a brief survey of these issues is contained König and Lechner, 1994, see also Davidson and MacKinnon, 1993). Therefore, the results given in Table 3 are computed using these estimates of the covariance matrix.

[31] The first version is almost numerically identical. *Only main-diagonal indicators* refers to a version of the information matrix test using as test indicators only the main diagonal of the difference between OGP matrix and the matrix of the expected hessian.

The estimated effect of gross income in 1990 (in 1993 DM) is nonlinear. It attains its maximum at about 1500, which implies that the income effect is positive for the first third of the income distribution and negative for the remainder part. Individuals who obtained some kind of training while being full-time employed in 1990 have a significantly higher OFT probability. Although there is a correlation with age, the mean of age in this group of 31.7 years is too high to justify the assumption that this variable captures vocational training which - due to a reporting error or an ambiguity in the question - has been described as full-time work by the respondent. The more likely interpretation is that people who were more likely to get some kind of training on the job in the former GDR are also more likely to receive OFT after unification.

The results in Table 3 show also marked differences regarding occupation and sector: production workers and people working in trade and most service sectors are c.p. significantly less likely to be observed in OFT.

It is noteworthy that with one exception none of the subjective expectation variables (in 1990) play any role in the (partial) propensity score. This could either be due to expectations changing so rapidly that those held in mid 1990 had no implication for later OFT decisions, or that OFT participation is much more a reaction to temporary and unexpected shocks, like actual unemployment. The results in Section 4.4 will show that the former explanation is not supported by the data, since when taking into account the latest pre-training expectation about job security, no significant difference appears between OFT and control group (conditional on the partial propensity score and the pre-training employment status). Furthermore, the insignificance of the subjective indicators of the difficulty of finding a new job and the objective 'job-danger' indicators, like having only a temporary contract or being already fired lends support to the claim that expectations concerning the security of the own job did not matter much for the OFT decision. However, the importance of shocks will be demonstrated in the next section by showing that the probability of being unemployed in the month just before OFT is much higher for OFT participants than for the control group in the same month (see Figure 6).

A comparison of Table 3 and Table A.1 reveals that many variables related to marital status, the federal states, motivations and general attitudes, memberships in job related organizations, finer groupings of job positions, occupations and professional degrees, remaining differences between federal states and the sizes of the cities and villages are all superfluous in the estimation of the partial propensity score.

Unfortunately, a comparison with other results concerning training participation is difficult, because some of the studies investigate the training participation in very different environments and/or use very different econometric methods and approaches, etc. (e.g. Fitzenberger and Prey, 1995, Helberger and Pannenberg, 1994, Hübler, 1994, Lynch, 1992, O'Higgins, 1994).

It remains to check some of the stochastic assumptions implied by the mutual independence of the error term $f(M,U)$ and $V\beta^0$, and the normality of $f(M,U)$. First of all, note that the last two columns of Table 3 largely do not seem to contradict the assumption of conditional homoscedasticity. In cases for which a rejection occurs, statistics based on different estimates of the covariance matrix of the test indicators suggest entirely different decisions regarding whether to reject the null of no misspecification or not. This could suggest that in these cases the $\chi(1)$ distribution, which is only valid asymptotically, may be a poor choice in small samples. Resolving this puzzle is left to future work. The normality test as well as the information matrix tests do not reject.

Table 4: Correlation of the estimated propensity score with potentially omitted time variant pre-training variables in %

	unemployment			involuntary short-time work			full-time employed		
monthly	last m.	4 m.	all m.	last m.	4 m.	all m.	last m.	4 m.	all m.
$v\hat{\beta}$	0.4	-0.5	-0.8	-2.1	-1.5	-3.1	-1.7	-3.0	-2.0

yearly	self-employment	full-time employment	unemployment
$v\hat{\beta}$	0.0	0.1	0.2

Note: *last m.*: last month; *4 m.*: four months' average (weighted towards the last month by using the weights: 0.173, 0.217, 0.271, 0.339); *all m.* : average of all months after unification and before OFT. *Yearly* denotes the last yearly observation before the beginning of OFT. The reference month for the control group is Dec. 1991.

Checking the correlation of $v\hat{\beta}$ with the potentially observable part of the error term does also not reveal any particular problem (see Table 4). In conclusion, the results of the various tests can be interpreted as not providing enough evidence to reject the maintained model.

4.3 Nonparametric estimation of causal effects and matching

The considerations in the previous sections suggest to estimate the causal effects by nonparametric methods in order to avoid potentially incorrect functional form restrictions. To ease notation assume that observations in the sample are ordered such that the first N_t observations receive OFT, and the remaining $(N-N_t)$ observations do not. The following two nonparametric regression estimators are obvious choices:

$$\hat{\theta}_N^1 = \hat{E}^1(Y^t - Y^c | S=1) = \frac{1}{N_t} \sum_{n_t=1}^{N_t} y_{n_t} - \hat{g}^c(x_{n_t}), \tag{6}$$

$$\hat{\theta}_N^2 = \hat{E}^2(Y^t - Y^c | S=1) = \frac{1}{N_t} \sum_{n_t=1}^{N_t} \hat{g}^t(x_{n_t}) - \hat{g}^c(x_{n_t}). \tag{7}$$

$\hat{\theta}_N^1$ and $\hat{\theta}_N^2$ denote the estimate of the causal effects that are averaged over the sample of the N_t-treated observations only. $\hat{g}^t(x_{n_t})$ denotes a consistent estimate of $E(Y^t | S=1, X=x_{n_t})$ (estimated in the treated pool with observations close to x_{n_t}) and $\hat{g}^c(x_{n_t})$ denotes a consistent estimate of $E(Y^c | S=1, X=x_{n_t})$ (estimated in the control pool with observations close to x_{n_t}), respectively. Although, generally neither $\hat{g}^t(x_{n_t})$ nor $\hat{g}^c(x_{n_t})$ are square root normal, it can be conjectured that under mild regularity conditions both $\hat{E}^1$ and $\hat{E}^2$ are consistent estimates and that there will be a central limit theorem to ensure that both $\sqrt{N_t}\hat{E}^1$ and $\sqrt{N_t}\hat{E}^2$ converge to a normal distribution with a fixed variance.

Generally, the practical problem of the high dimension of x_{n_t}, which in multivariate nonparametric regressions typically requires a large number of observations close to x_{n_t}, can be overcome be using the propensity score property. Instead of computing the multivariate regression to obtain $\hat{g}^t(x_{n_t})$ and $\hat{g}^c(x_{n_t})$, it is sufficient to compute univariate regressions using the estimated propensity score instead of x_{n_t}. However, in the particular case considered in this paper - some attributes (M) are only defined in relation to a particular treated observation so that the attribute vector hast to be partitionend in V and M - part of the problem of the dimension being too large remains. To capture the employment information appropriately, nine monthly variables (see Table 4) and a yearly variable have to be used additionally to the propensity score. Hence, the dimension of the nonparametric regression is so high that serious small sample problems can be expected for the size of the sample available for this study. Additionally, a separate estimation of $\hat{g}^c(x_{n_t})$ in the control pool is necessary for each different starting date of OFT, which would be a huge computational burden.

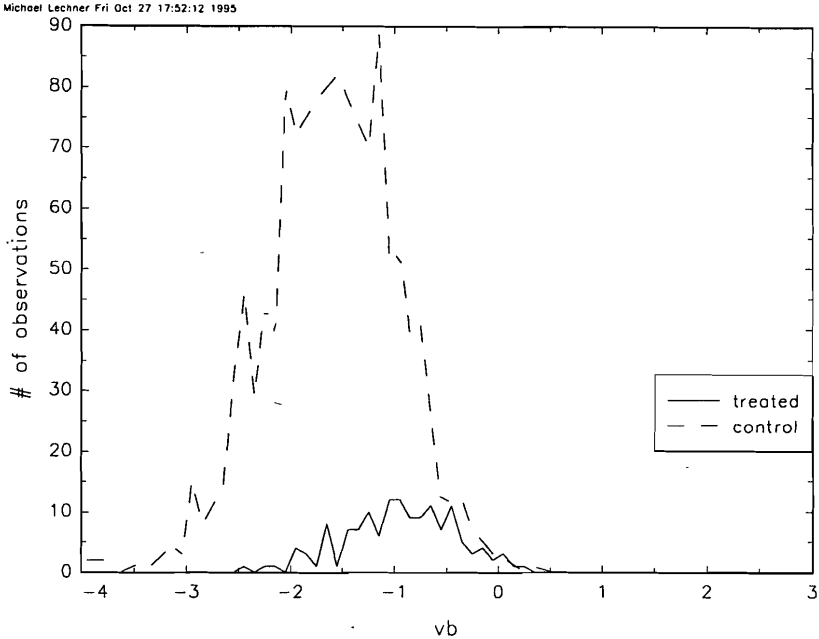
For these reasons I choose to use a simpler nonparametric approach that appeared in the statistic literature (e.g. Rosenbaum and Rubin, 1983, 1985). The basic idea is to find for every treated observation a control observation that is as close to it as possible in terms of a balancing score. When an identical control observation is found, the estimation of the causal effects is unbiased. In cases of 'mismatches', it is often plausible to assume that using local regressions on these differences will remove the bias (see Section 4.4 for details). Note that compared to the nonparametric regression described above, there is an efficiency loss, because observation n_t and its closest neighbor in the control population - instead of possibly many close neighbors - are used.

A basic requirement for a successful (i.e. bias removing) implementation of a matching algorithm is a sufficiently large overlap between the distributions of the conditioning variables in both subsamples. Figure 5 shows the overlap for a very important conditioning variable, $v\hat{\beta}$. Although the mass of the distribution of the controls is to the left of the treated, it seems that there is overlap for most part of the treated distribution. Table 5 contains some descriptive statistics of attributes in the treated and in different control samples. Comparing columns (2) and (3) of that table shows that matching on the propensity score alone makes the distribution in the control sample very similar to the distribution in the treated sample.[32] However, it should be noted that conditioning is on $v_{n_t}\hat{\beta}$ instead of $v_{n_t}\beta^0$. The asymptotic standard error[33] of $v_{n_t}\hat{\beta}$ resulting from the estimation of $\hat{\beta}$ can be considerable and ranges from 0.19 to 0.96 in the OFT sample, and from 0.17 to 1.55 in the control sample. The mean in the OFT (control) sample is 0.31 (0.30), the median 0.29 (0.28), and the empirical standard deviation 0.10 (0.10). Therefore, it can be expected that by matching only approximately on $v_n\hat{\beta}$, but additionally also on some important components of v directly, a better match could be obtained.

[32] The different versions of the matching algorithms are obvious simplifications of the algorithm given in App. C.1.

[33] Computed using the delta method.

Figure 5: Distribution of $v\hat{\beta}$ for OFT and controls



Note: 0.1 grid used. Mean (*std*) in OFT (treated) sample / control sample is -0.95 (*0.53*) / -1.61 (*0.63*).

The details of the matching algorithm used are described in Appendix C.1. It follows Rosenbaum and Rubin (1985) suggestion of "matching within calipers of the propensity score" with the exception that window sizes (caliper widths) depend explicitly on the precision of the estimate $v_n\hat{\beta}$. The more precise $v_n\hat{\beta}$ is estimated, the smaller is the width. The additional variables used (col. 4 in Table 5) are *gender*, *Berlin*, *university*, *10 years of schooling*, the *expectation of no redundancies in the firm for the next two years* (1990), a *highly qualified or management job position* (1990), *monthly wage / salary* (1990) and *training (unspecified) while full-time employed* (1990). Using these variables - a subset of those variables included in v_n - separately is an additional safeguard against any impact due to inconsistent estimation of the partial propensity score. The results that are contained in column (3) of Table 5 appear to resemble the distribution of the OFT sample (6) closely.

Table 5: Descriptive statistics of selected variables of OFT and control sample: Different matching algorithms

	Controls				OFT (122)
	all (1105)	matched on $v\hat{\beta}$ (122)			
			and select. v-variables		
				and m-var	
(1)	(2)	(3)	(4)	(5)	(6)
Variable	mean(std) share in %	mean(std) share in %	mean(std) share in %	mean (std) share in %	mean(std) share in %
$v\hat{\beta}$	-1.61 (.63)	-0.90 (.52)	-0.96(.49)	-0.98(.53)	-0.89(.51)
Age in 1990	35.2(8.1)	34.4 (6.9)	35.2 (7.0)	35.2 (7.2)	35.4 (7.7)
Gender: female	42	66	66	62	64
Federal states (Länder) in 1990					
Berlin	7	13	12	10	13
Mecklenburg-Vorpommern	10	2	2	7	6
Sachsen-Anhalt	15	19	21	18	15
Years of schooling (high. deg.) 1990					
12	17	29	33	27	31
10	60	66	61	66	63
8 or no degree	22	6	6	7	6
Highest professional degree in 1990					
university	11	23	25	23	25
engineering, technical college	16	27	33	25	33
skilled worker	65	43	36	43	34
Job position in 1990					
highly qualified, management	19	42	43	40	43
skilled blue and white collar	57	36	39	46	40

Table 5 to be continued...

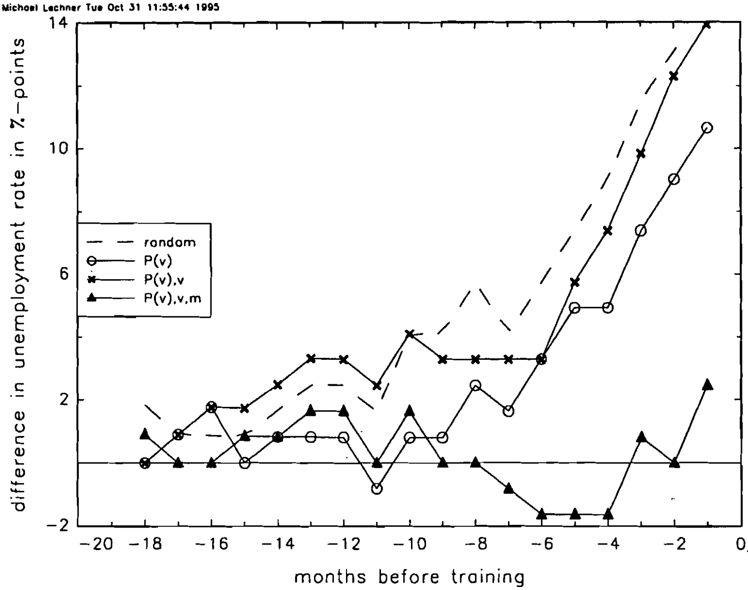
Table 5: Descriptive statistics: continued

(1)	Controls (2) all (1105)	(3) matched on $\boxed{v\hat{\beta}}$ (122)	(4) and select v-variables	(5) and m-var	OFT (122) (6)
Variable	mean(std) share in %	mean(std) share in %	mean(std) share in %	mean(std) share in %	mean(std) share in %
$v\hat{\beta}$	-1.61 (.63)	-0.90 (.52)	-0.96(.49)	[illegible]	-0.89(.51)
wage / salary per month(defl.)	1714(526)	1734(399)	1787(388)	[illegible]	1736(398)
training (unspecified) while full-time employed	7	17	16	18	16
Profession in 1990 (ISCO)					
scientific, technical, medical	19	41	45	43	39
production	43	11	12	16	13
services, incl. trade, administ.	23	26	21	20	21

Note: (2) no matching; (3) matched on $v\hat{\beta}$ (122); (4) matched on $v\hat{\beta}$ (122) and selected v-variables;
(5) matched on $v\hat{\beta}$ (122), selected v-variables and m (monthly, yearly)-variables; 1990 relates to the date of interview which for almost all cases was completed before July 1990 (EMSU). Ratio of variance of $v\hat{\beta}$ in OFT sample over variance in control sample is 0.71. Average width of a caliper is 0.98. v-variables used for the additional conditioning are: *gender, Berlin, university, 10 years of schooling, expectation of no redundancies in firm for the next two years* (1990), *highly qualified or management job position* (1990), *monthly wage / salary* (1990), *training (unspecified) while full-time employed* (1990); see also note to Tables A.1.

As mentioned in Section 4.1 conditioning on monthly employment information to capture the impact of temporary shocks could be important. Figure 6 shows indeed that including only $v_n\hat{\beta}$ in the balancing score is insufficient. The figure displays the difference in the unemployment rate between OFT and different control samples relative to the number of months before OFT. The three lines that are highest in the right hand part of the plot are based on the matching methods mentioned so far plus a random draw in the control pool (col. 2 in Table 5). They are very similar and reveal unemployment rates that are up to 14%-points lower than for the OFT sample.[34] Conditioning additionally on the yearly and monthly pre-training employment information (see Table 4 for details on the variables used) reduces the bias significantly. Although there is still a small upward bias, figures in the next subsections will show that it is not significantly different from zero. Therefore, all the following evaluations are based on this matched sample.

Figure 6: Difference of registered unemployment between OFT and matched control groups: a comparison of different matching algorithms



Note: See note to Tables 4 and 5.

It is noteworthy that in the first part of their paper Card and Sullivan (1988) choose a very similar approach. They match treated and controls regarding their pre-training employment history. Unfortunately, they are in a worse position, because their data is subject to potentially considerable measurement error concerning these variables. Additionally, the variables are only measured on a yearly basis, so that the employment status just prior to training is unknown. Furthermore, they completely ignore the kind of variables that enter the partial propensity score in this analysis. Therefore, it is not surprising that they decide

[34] The level of unemployment in the month just prior to OFT is 20% for those receiving OFT (involuntary short time work: 11%, full-time work: 67%).

that this kind of conditioning is insufficient to yield unbiased estimates and switch over to a model-based-approach.

4.4 Evaluation

4.4.1 Outcomes

This paper is particularly interested in the effects of OFT on post-training changes in actual and anticipated labor market status and prospects. It is due to the nature of the data and circumstances (German unification in **1990**) that at the time this paper is written no long run effects of OFT can possibly be discovered.

The following actual outcomes are measured on a monthly basis by way of the retrospective employment calendar: involuntary short-time work, registered as being unemployed, and full-time employment. In addition, the latter two variables are also available for the date of the yearly interview. Another variable capturing characteristics of the actual labor market status - measured once a year - is gross monthly income. For those being employed, it is defined as the gross monthly income in the month before the interview. For those not being employed, imputed unemployment benefits or social assistance - whichever is higher - are used instead (see Appendix A for details). Labor market prospects are measured once a year as individual expectations or worries. They include expectations whether one might lose one's job in the next two years, and whether one is very worried about the security of the current job.[35] Additionally, there is information whether individuals expect an improvement or a worsening of the current job (career) position.[36] It is important to note for the discussion in the following subsection that, except for the income variable, all outcome variables are coded as binary indicators.

Finally, there is the issue of comparing outcomes for individuals participating in courses with different end dates. Here, two concepts of comparison are applied. They consist either in specifying a date (early 1993 or 1994 for yearly information, or a specific month before Jan. 1993 or 1994 for monthly information) or a specific time span (months or intervals of 0-1, 1-2, 2-3 years for yearly information) after the completion of the course. Note that the number of observations available for the evaluations decreases with the length of the time span considered.

[35] For non-employed individuals these variables are coded as being very worried and as "expecting unemployment".

[36] For non-employed individuals these variables are coded as "expecting no improvement and no worsening".

4.4.2 Econometric issues

Define the differences in matched pairs in the sample, which consists of independently drawn observations, as $\Delta y_{n_t} = y^t_{n_t} - y^c_{n_t}$, $\Delta x_{n_t} = x^t_{n_t} - x^c_{n_t}$, $n_t = 1,\ldots,N_t$, where $y^c_{n_t}$ and $x^c_{n_t}$ denote values of an observation from the pool of individuals not participating in OFT (controls) that is matched to the treated (OFT) observation n_t. When the outcomes are approximately continuous variables, e.g. income, then Δy_{n_t} is approximately continuous. Otherwise, the outcomes are measured with indicators (0, 1) and Δy_{n_t} takes on the discrete values -1, 0 and 1. The estimate of the average causal effect and the respective standard error are computed as:

$$\hat{\theta}_{N_t} = \frac{1}{N_t}\sum_{n_t=1}^{N_t} \Delta y_{n_t}, \qquad \mathrm{Var}(\hat{\theta}_{N_t}) = \frac{1}{N_t}(S^2_{y^t} + S^2_{y^c}). \tag{8}$$

$S^2_{y^t}$ and $S^2_{y^c}$ denote the square of the empirical deviation of y_t in the OFT sample and in the sample matched to the OFT-sample, respectively.[37] As mentioned in the previous section, when a perfect match is achieved, implying that $\Delta x_{n_t} = 0$, $n_t = 1,\ldots,N_t$, these estimates are unbiased (cf. Rosenbaum and Rubin, 1983). When the sample is large enough the normal distribution can be used to perform tests and compute confidence intervals. Equation (8) denotes the baseline nonparametric estimate of the causal effect to be discussed in the following subsection. Those are also computed for subpopulations defined by attributes or training characteristics. Note that no assumption is necessary regarding whether or not the treatment effects may differ across the population.

Now, let us consider the case when there is local mismatch, in the sense that although Δx_{n_t} is close to zero - and would be closer if the sample would have been larger ($N_t < N / 2$)- it is actually different from zero. There may be two reasons for local mismatches: on the one hand the coefficients of the propensity score are estimated, and, therefore, matching on $v_n \hat{\beta}$ could be different from matching on $v_n \beta^0$ in finite samples. On the other hand, the pool of available control observations may be too small to contain exact matches. Again, this problem is less severe with large (control) samples that have a sufficient overlap of attributes with the OFT participants. To correct for biases that could arise from these problems, some modeling is used.

[37] Note the variance estimate exploits the fact that the matching algorithm proposed in App. C.1 never chooses an observation twice.

Define for the treated subpopulation the variables $\Delta Y_i = Y_i^t - Y_i^c$ and $\Delta X_i = X_i^t - X_i^c$ as the population difference for pairs that would have been matched, for example if the complete population were available to the researcher. As before, Y_i^c and X_i^c denote the attributes of a control observation matched to the treated observation *i*. Assume for the purpose of illustration, that these matches remain imperfect, so that ΔX_i may be small, but different from 0. In the case of continuous variables it seems reasonable to assume that the conditional expectation of the dependent variable is linear in ΔX_i, because matching has already removed almost (if N is finite) all differences in the X variables, so that in fact the ΔX_i or Δx_{n_t} are local deviations. Local smoothing using a linear conditional expectation is not very restrictive and standard linear regression methods can be used to estimate the average treatment effect θ^0 (cf. Rubin, 1979) by regressing the differences in the attributes and a constant on the differences in outcomes.[38] Appendix C.2 shows what conditions are necessary for the estimated constant term of that regression to be a consistent estimate of θ^0, when the treatment effect actually varies over the population. Suppose now that the outcome consists of only two values, say 0 and 1. Clearly, using a linear approximation for these differences of probabilities is not so attractive as before, except when ΔX_i is very small. Therefore, I use an ordered probit model instead of a linear model to estimate the coefficients of these probabilities based on an underlying latent normal model. Given consistent estimates of the coefficients, the difference of the probabilities is estimated. The standard errors are computed using the delta method (for details see Appendix C.2).

The same approaches are chosen to check whether the treatment effects vary either with characteristics of the courses, such as its duration, or with characteristics of the individuals participating in OFT. Note that this procedure is not nested in the previous one, because now the assumption that either the treatment effect is stable or varies in a particularly specified way is indispensable (Appendix C.2). Therefore, splitting the samples in subpopulations and performing estimations in these subpopulations that do not require such an assumption is an attractive alternative for discrete attributes and characteristics. However, when the attributes and characteristics have too many different values some modeling is required given the size of the sample used in this study.

It should also be remarked that whenever regression-type adjustments are used for different dates (time spans) for the same outcome variable, no cross-period-coefficient restrictions are assumed to hold, but the estimations are performed for each date or time span separately. Finally, for the yearly variables all means,

[38] Standard errors are computed using a heteroscedasticity robust estimator. The particular variant is labeled as HC_2 by Davidson and MacKinnon (1993, p.554) and has good small sample properties.

variances and regressions are also computed using the appropriate panel weights. Since there are only minor differences among weighted and unweighted estimates, the former are not computed for the monthly data.

4.4.3 Results

The first set of results is given in Figures 7 to 14. It shows the differences between the control and the OFT group for specific time spans before and after the training for a selected group of outcome variables (multiplied by 100 for outcomes that are indicators).[39] For variables measured by the calendar (see Figure 3) the distance is expressed in months, for those measured only for the month of the yearly interview, the distance is expressed in years.[40] The figures cover up to 18 months or up to 3 'years' before the training and up to 41 months or 4 'years' after OFT. They display the mean effect (solid line; + for mismatch corrected estimate) and its 95% confidence interval based on the normal approximation (dashed line; ∇, Δ for the mismatch corrected estimates). The estimates that are not corrected for mismatch are shown as lines. The mismatch corrected estimates (post-treatment only) are displayed as unconnected symbols.

The number of observations available to compute the respective statistics decrease the longer the distance to the incidence of OFT is. The implications of this are threefold: First of all, the variance increases. Although this is reflected in the widening of the confidence interval, the accuracy of the estimated interval itself may deteriorate, because the normal distribution may be not a very good approximation of the sample distribution when the sample gets too small. Finally, a mismatch correction may be impossible or very imprecise, because there may be too few observations to identify and estimate the parameters of the ordered probit model.[41]

---

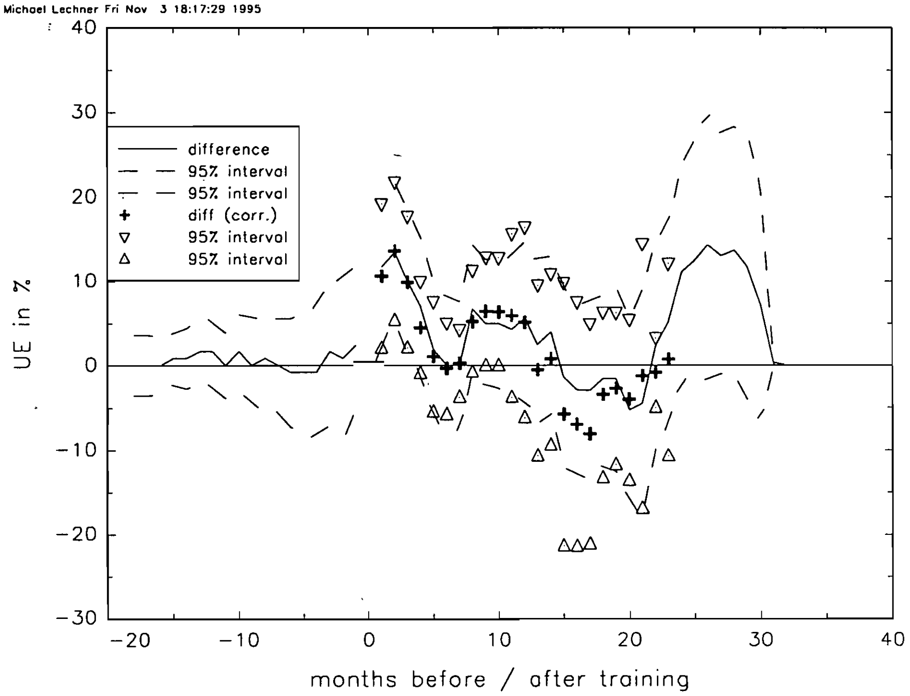
[39] The results for those outcomes that are mentioned in Section 4.4.1, but do not appear here, are not qualitatively different from the ones presented. Therefore, they are omitted for the sake of brevity.

[40] The time span denoted as the first year is actually the time after the end of OFT and the next interview. Therefore, this time span may vary among individuals. Currently, the monthly data available starts in July 1989 and ends in December 1993, whereas the yearly data ranges from mid 1990 to early 1994.

[41] All computations based on less than 5 observations are suppressed. Furthermore, for the plots based on the monthly data do not display any effect or confidence bounds above +40 or below -40.

Figures 7 and 8 present the monthly unemployment status for the complete sample and a subsample of individuals not employed during OFT. The part left to the 0 vertical mark allows a judgment about the quality of the matches concerning the particular variable.[42] As already noted in the discussion of Figure 6, there is small excess unemployment just prior to the beginning of the course, which is however not at all significantly different from zero.

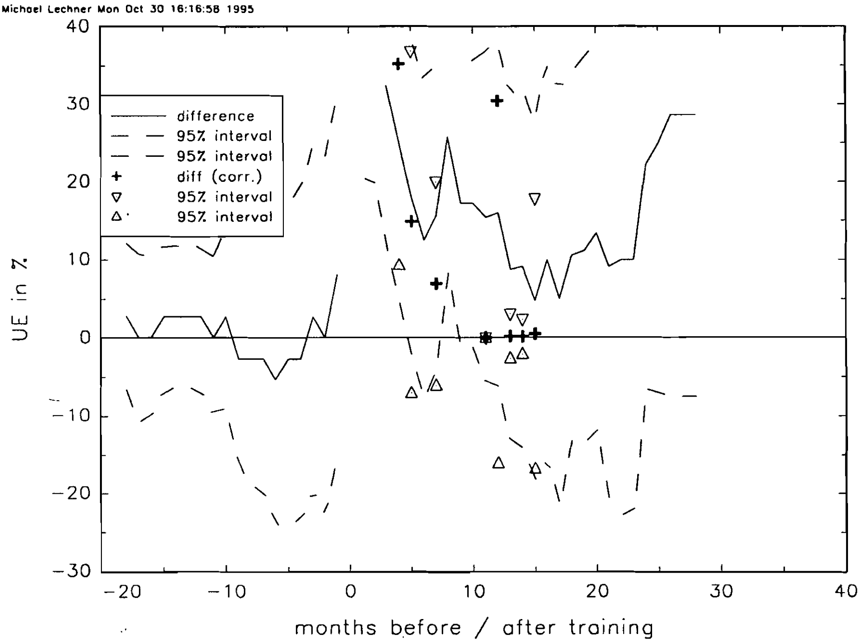
Figure 7: Registered unemployment



Note: $N_t = 122$.

[42] Testing whether these lines deviate significantly from zero is in the same spirit as the tests suggested by Rosenbaum (1984) to use overidentifying restrictions to try to invalidate CIA. The pre-OFT outcomes here are denoted as unaffected outcomes in his terminology.

Figure 8: Registered unemployment: only OFT participants not employed during OFT



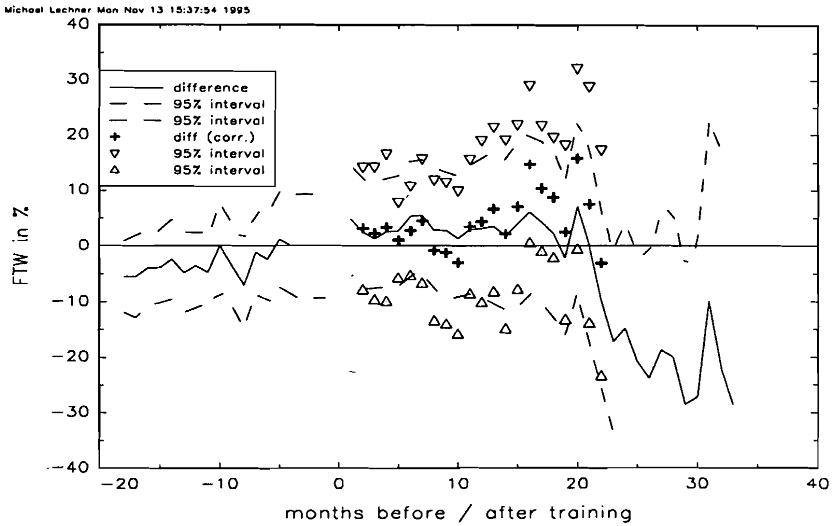
Note: N_t =37.

The effect of training appears to be higher unemployment in the months directly following the end of it.[43] This is a plausible effect when one takes into account that for those unable to keep their previous occupation job search is required. Since this is time consuming, it may not be performed with full intensity until OFT ends. Meanwhile, more members of the control group already found a new employment. This point is particularly obvious when considering only the subsample of individuals not employed during OFT (Figure 8). These effects disappear entirely after about 6 to 12 months. These conclusions are confirmed by the inverted shape (Figure 10) of the mean of full-time employment of those OFT participants who are not employed during OFT. However, Figure 9 suggests that for the remaining sample, which is employed during OFT, these considerations are - for obvious reasons - not important. Here, OFT does not appear to have any impact whatsoever.

---

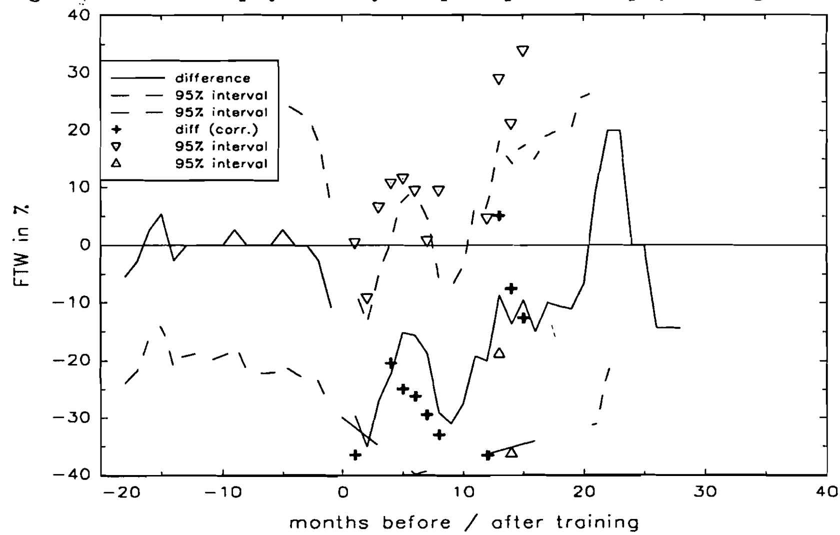
[43] The reader is reminded that the end date is measured with error. Here, it is coded to be never earlier than the true end date. However, there may be a few cases with longer durations for which it could be several months too late (details in App. A).

Figure 9: Full-time employment: only OFT participants employed during OFT



Note: $N_t = 85$.

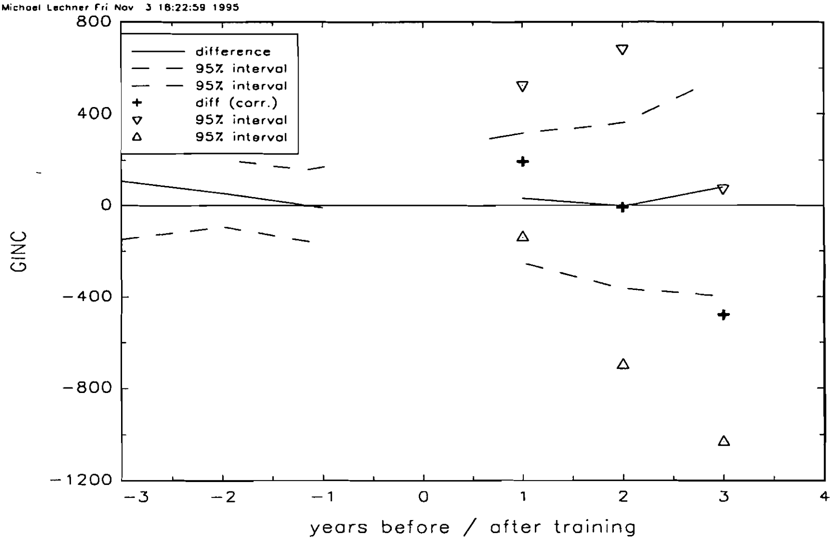
Figure 10: Full-time employment: only OFT participants not employed during OFT



Note: $N_t = 37$.

Figures 11 to 14 feature outcome variables that are only measured once a year, such as gross monthly income, being very worried about keeping one's job, and expected improvement or decline in the professional career in the next two years. On the one hand, there are no significant differences for the pre-training outcomes. On the other hand, the same is true for the post-treatment period. This general result is valid for all yearly variables. It is also robust concerning other functional forms (such as logs) of the income variable, for instance.

Figure 11: Gross income (in 1993 DM)



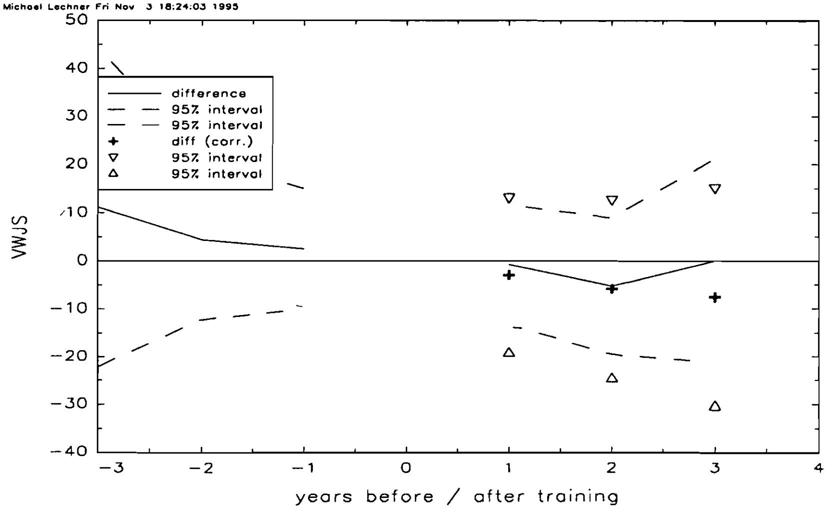
Note: N_t = 122. Income when not employed coded as unemployment benefit or social assistance, whichever is higher. See Appendix A for details.

To check whether there might be differences of the average treatment effects in specific subgroups the sample is divided according to gender, job position, professional degree, and as already mentioned, whether the individual was employed during OFT.[44] No significant differences appear. Finally, to check the results for sensitivity with respect to the definition of OFT, the courses used in the estimation are split in several subsamples according to whether (i) they began not earlier than January 1991 (N_t = 108), (ii) they have a minimum duration of one week (N_t = 95), (iii) the objective is qualification for promotion (N_t = 45) or (iv) the adjustment of skills (N_t = 84), and whether (v) a certificate has been obtained by the participant that could be helpful for future job applications (N_t = 101). As a final sensitivity check I also considered a control and treatment group that did not participate in any other form of training (N_t = 108). None of the

[44] See table A.1 for the value N_t corresponding to the particular partition.

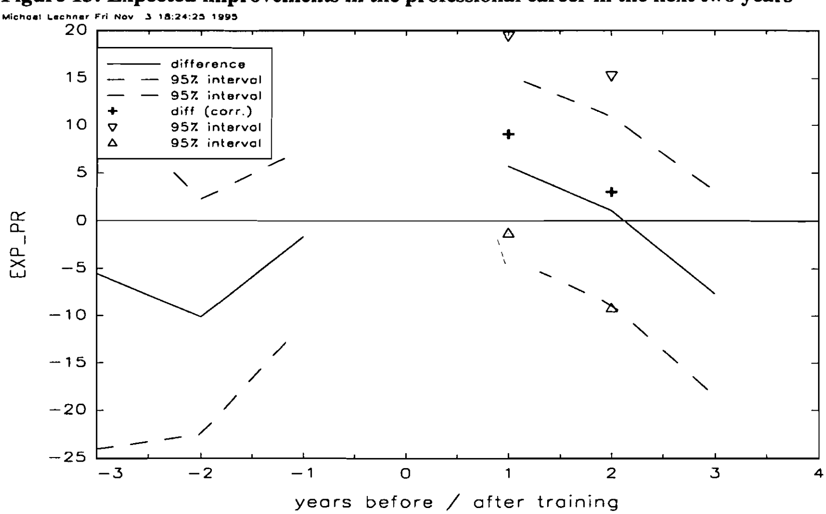
subsamples reveals a substantial difference compared to the results presented above.

Figure 12: Very worried about possibility of future job loss (or unemployed)



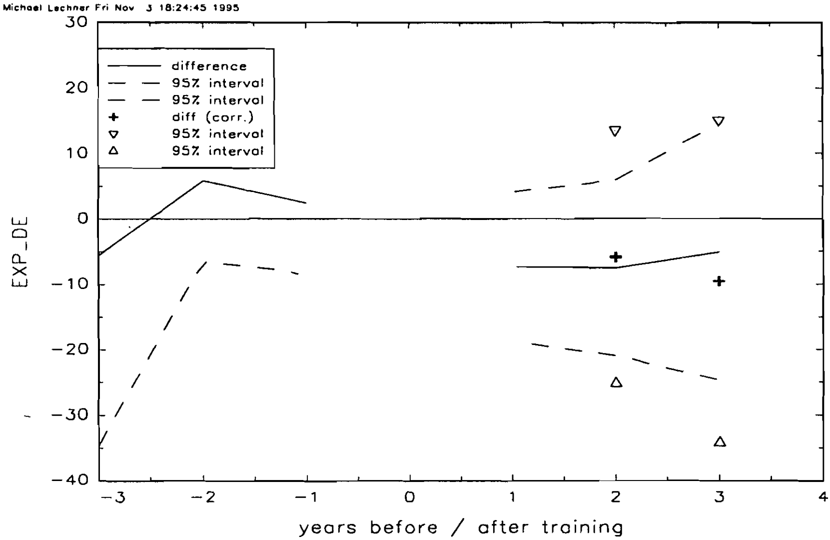
Note: N_t = 122. Nonemployment coded as being very worried.

Figure 13: Expected improvements in the professional career in the next two years



Note: N_t = 122. Nonemployment coded as not expecting improvement.

Figure 14: Expected decline in the professional career in the next two years



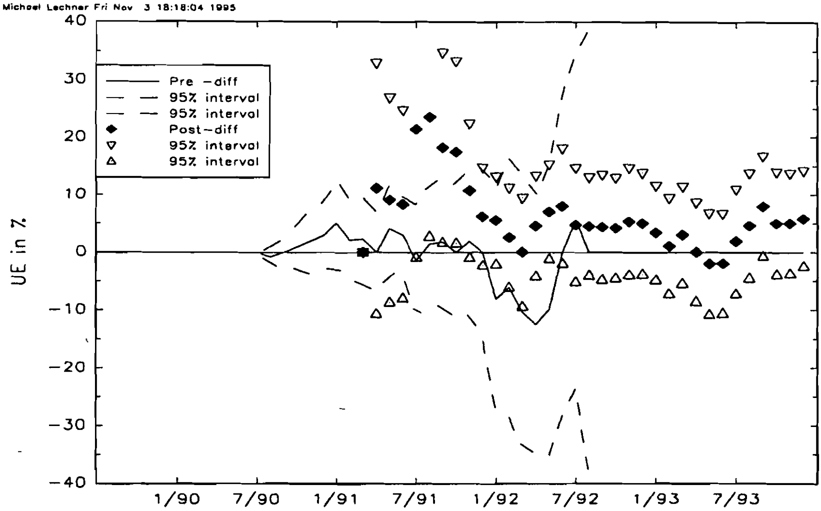
Note: N_t = 122. Nonemployment coded as expecting decline.

A technical note is in order: the closeness of the mismatch-adjusted and not-mismatch-adjusted results, as well as the insignificance of the differences of pre-OFT outcomes, suggest that matching already removes almost all of the bias due to different distributions of the attributes in the OFT and the control sample. Although matching on pre-OFT training may be invalid for reasons already discussed, I checked pre-OFT training history in the same way as done for the other monthly variables (e.g. the left hand part of Figures 7-10). No significant differences appeared between the control and OFT group.

Having discussed results concerning the distance in time to the beginning and ending of a training course, I now turn to the second perspective and consider results for specific dates. Figures 15 and 16 show the development of pre-training (lines) and post-training outcomes (unconnected symbols) over time.[45] Note that when moving from left to right the number of observations is decreasing for pre-training outcomes and increasing for post-training outcomes. However, the conclusions drawn above regarding matching quality and nonexisting OFT effects are adequate for this perspective as well. Since the perspective used above is more informative concerning training outcomes, and because there are no qualitative differences, the results for the other variables are omitted.

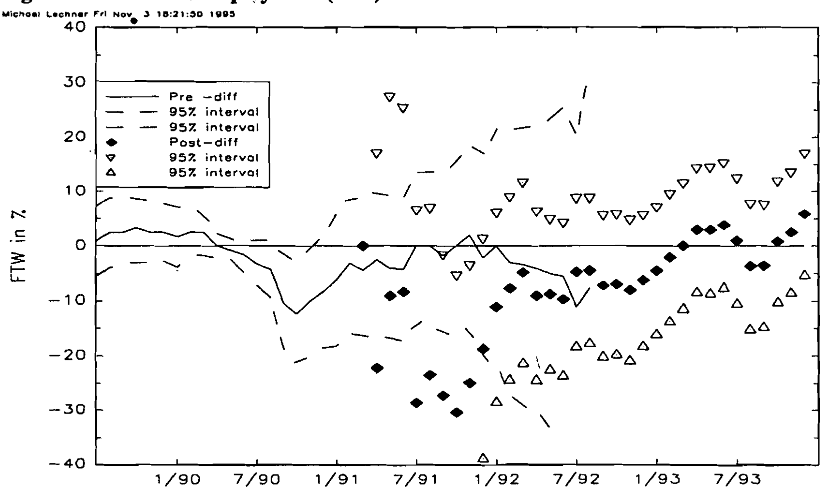
[45] Mismatch adjustment is not performed for these two figures.

Figure 15: Registered unemployment (date)



Note: $N_t = 122$.

Figure 16: Full-time employment (date)



Note: $N_t = 122$.

Finally, several yearly outcome measures are evaluated for early 1993 (Table 6) and early 1994, which is the latest date available (Table 7). These estimates are based on the full matched sample using the modelling methods described in Appendix C.2. The upper part of Table 6 shows various estimates of the average effects of OFT. None of them is significant, which is in accordance with the findings above. The same results for 1994 indicate only a small positive effect with respect to ***expected improvements in the own professional career*** in the next two years (1995 and 1996). This could hint at possible future positive findings with respect to actual labor market outcome, if the individuals do foresee the future correctly (which was apparently not true in the past).

Table 6: Evaluation results for several outcomes measured in early 1993

	monthly income in DM		full-time employment		very worried about possible future job loss		expecting improving profes.career		expecting declining profess.career	
	$\hat{\theta}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.
mean ($\hat{\theta}_{N_1}$)	24	153	-4	6	5	7	4	5	5	6
weighted mean	30	185	-0	6	6	7	2	5	3	7
mismatch adjust.	65	136	-3	7	9	9	1	5	6	7
Average effect for mean course duration and marginal effect of one month of duration [2) 3)]										
average effect	64	146	-3	6	9	9	1	5	6	7
marginal effect	-2	58	0	2	1	2	0	1	-1	2
Separate marginal effects of attributes										
Other course characteristics; reference group: no certificate received and other objectives of course										
certificate receiv.	**1242**	395	r)	-	-10	19	r)	-	r)	-
aim: adjustment	203	306	-4	13	7	16	10	11	-17	15
aim: promotion	-30	351	1	15	3	21	19	12	-11	16
Not employed	-494	424	-25	15	27	19	1	9	18	15
Months in respective labor market status after 6/1990 (/ by number of months prior to OFT)										
reg. unemployed	-1089	1456	-72	60	**163**	31	18	26	**117**	53
inv. short-time w.	326	847	-34	37	7	50	-38	26	21	37
Gender: female	-197	342	7	13	1	16	-12	10	-16	14
Age	-71	209	1	8	11	11	**13**	5	-2	9
Years of schooling (highest degree); reference group: less than 10 years										
12 y. of schooling	394	680	37	28	-46	35	-11	17	-43	28
10 y. of schooling	17	559	31	27	-51	34	3	15	38	27

Table 6 to be continued...

Table 6 Evaluation Results: continued

	monthly income in DM		full-time employment		very worried about possible future job loss		expecting improving profes.career		expecting declining profess.career	
	$\hat{\theta}$	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.
Federal states (Länder) in 1990; reference state: Sachsen										
Berlin	-514	469	-20	24	36	26	-4	14	**49**	24
Thüringen	-165	385	-13	16	26	20	10	11	**37**	18
Mecklenburg-VP	-315	608	r)	-	33	33	16	18	r)	-
Brandenburg	311	463	13	20	16	26	r)	-	-8	21
Sachsen-Anhalt	-365	427	-27	17	**66**	22	6	12	**52**	18
Highest professional degree in 1990; reference group: skilled and unskilled worker										
university	485	459	23	14	-18	19	-16	13	-23	15
engineer., tec.col.	-200	359	5	15	1	20	-5	11	-5	17
master trade/craft	450	582	-5	30	-24	49	r)	-	10	31
Job position in 1990; reference group: lower job positions										
highly qual., man.	607	457	**51**	15	-9	26	-0	13	**-48**	19
master trade/craft	247	631	r)	-	-19	43	r)	-	11	23
skilled blue, wh.c.	460	461	**48**	16	-15	27	5	12	**-46**	20
Propensity score	-689	1147	4	4	-5	5	-2	3	-5	5

Note: **Bold** letters: t-value > 1.96. [1] %-points; [2] Linear models: Regression includes constant (average effect), duration in months normalized to mean 0 (marginal effect), mismatch adjusting covariates. Nonlinear models: see App. C.2. [3] Evaluated at mean duration; r) Included in reference group, because of small cells and resulting insufficient variation within groups of dependent variable for ordered probit;
Marginal effects for binary outcome variables with binary attributes are computed by changing status of members of respective attribute group (e.g. 12 years of schooling) to status of reference group (e.g. 8 y.). Entries in the table are differences of these effects separately for every attribute. They are bounded by -200 and +200. The details of the computations which include also changes in continuous characteristics (age: ± 0.5 years, duration: ± 0.5 months; duration of unemployment, IST: ± 0.05 months) are given in App. C.2. The computations are based on the subsample of those who have already completed their OFT participation by early 1993.

Table 7: Evaluation results for several outcomes measured in early 1994

	monthly income in DM		full-time employment		very worried about possible future job loss		expecting improving profess. career		expecting declining profess.career	
	$\hat{\theta}$	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.	$\hat{\theta}$ 1)	std.er.
mean ($\hat{\theta}_{N_1}$)	22	158	0	4	-5	6	7	4	-10	6
weighted mean	100	174	1	5	-6	7	6	4	-12	7
mismatch adjust.	12	176	1	6	-9	8	**8**	4	-14	8
Average effect for mean course duration and marginal effect of one month of duration 2) 3)										
average effect	71	184	0	6	-13	8	7	4	-14	9
marginal effect	62	47	1	1	-4	2	1	1	0	2
Separate marginal effects of attributes										
Other course characteristics; reference group: no certificate received and other objectives of course										
certificate receiv.	427	420	-7	19	-19	18	5	11	3	17
aim: adjustment	621	361	**26**	13	-23	14	5	6	**-31**	13
aim: promotion	397	389	**33**	14	-31	17	**16**	8	**-30**	15
Not employed	-119	322	**-24**	13	3	14	-1	6	**31**	12
Months in respective labor market status after 6/1990 (/ by number of months prior to OFT)										
reg. unemployed	1032	1174	-66	46	**72**	36	27	21	**93**	42
inv. short-time work	-127	904	-5	32	-46	34	-8	11	-5	29
Gender: female	-515	301	6	11	14	12	-8	7	10	12
Age	-1658	2216	-3	7	10	8	**-8**	4	3	7
Years of schooling (highest degree); reference group: less than 10 years										
12 y. of schooling	84	746	0	28	-18	28	8	11	-30	26
10 y. of schooling	195	702	14	28	-40	33	-1	9	-50	34
Federal states (Länder) in 1990; reference state: Sachsen										
Berlin	-315	562	9	16	-17	18	-26	15	0	16
Thüringen	-145	455	5	18	r)	-	r)	-	-8	17
Mecklenburg-VP	-320	673	**-57**	27	57	33	r)	-	32	30
Brandenburg	362	488	26	18	-3	18	-16	11	-15	18
Sachsen-Anhalt	-23	387	-2	15	**38**	15	-9	7	3	17

Table 7 to be continued...

Table 7: Evaluation Results: continued

	monthly income in DM		full-time employment		very worried about possible future job loss		expecting improving profess. career		expecting declining profess.career	
	$\hat{\theta}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.	$\hat{\theta}^{1)}$	std.er.
Highest professional degree in 1990; reference group: skilled and unskilled worker										
university	-286	221	-3	14	21	17	-4	7	-4	14
engineering, t.col.	391	783	8	16	-5	25	-9	8	-11	17
master trade/craft	-598	389	**61**	28	18	18	r)	-	**-62**	27
Job position in 1990; reference group: lower job positions										
highly qual., man.	-170	367	14	15	-6	22	-7	8	-19	16
master of trade cr.	517	599	r)	-	-57	31	r)	-	r)	-
skilled blue, wh.c.	204	361	21	14	-28	21	-11	9	-18	15
Propensity score	110	1152	8	4	-1	5	**-5**	2	-6	4

Note: **Bold** letters: t-value > 1.96. 1) %-points; 2) Linear models: Regression includes constant (average effect), duration in months normalized to mean 0 (marginal effect), mismatch adjusting covariates. Nonlinear models: see App. C.2. 3) Evaluated at mean duration; r) Included in reference group, because of small cells and resulting insufficient variation within groups of dependent variable for ordered probit;
Marginal effects for binary outcome variables with binary attributes are computed by changing status of members of respective attribute group (e.g. 12 years of schooling) to status of reference group (e.g. 8 y.). Entries in the table are differences of these effects separately for every attribute. They are bounded by -200 and +200. The details of the computations which include also changes in continuous characteristics (age: ± 0.5 years, duration: ± 0.5 months; duration of unemployment, IST: ± 0.05 months) are given in App. B.2.

Using the techniques discussed in the previous subsection to detect differences among different types of training courses or different attributes of the participants also reveals little.[46] One of the few exceptions for 1993 is a significant positive income effect for participants who obtained a certificate compared to those who did not get one. However, this does not imply that there is a positive income effect for the first group, but it merely means that the second group faces a large and significant negative effect.[47] This effect does not reappear in 1994. Similar negative effects concerning being full-time employed (or unemployed) or the future professional career can be observed in 1993 for those who had very low job positions in the GDR. Training effects regarding worries about job security and future declining career prospects differ according to the occurrence and duration of unemployment prior to training (DUPT). The higher DUPT is, the worse are the subjective expectations concerning the career and employment prospectives. This may very well be related to the fact, noted above, that the immediate impact of OFT is in many cases some months of unemployment. Regional variations can be found for some outcome variables. However, comparing them to differences in unemployment rates across the federal states (cf. Statistisches Bundesamt, 1994, Table 6.13) reveals only little correlation. Finally, differences according to the age of the participants do only appear for the expected improvement in career prospects. However, its sign is only for the 1994 estimates as expected whereas the fact that it is larger for older individuals in 1993 seems hard to explain. In 1994 the group of other course characteristics and unemployment during OFT appears to make significant differences. The effect of the latter is as could be expected from the previous figures, whereas the latter indicate that those not on OFT with the explicit aim of adjusting the skills and future promotion do badly indeed.

Summarizing the results presented in the figures and tables in this subsection, it should be stressed that no robust positive effects of OFT can be found, and even some temporary negative effects surfaced. There are three possible general reasons for this finding: First of all, the effects can be so small that it is impossible to determine them with the available sample size. Secondly, there could be positive effects in the longer run that cannot yet be seen. Finally, it could well be that there are no positive effects at all. This conclusion, if also confirmed for more recent courses, would have very serious implications for public policy. It is worth noting that this study has found no evidence whatsoever to rule out that possibility.

These results are in contrast to more positive results obtained in a recent study by Fitzenberger and Prey (1995, FP). FP use the first six waves of the

[46] Marginal effects are also computed controlling for course duration linearly, and using the panel sample weights, but the qualitative results do not change.

[47] The same effect is significant for ln(income).

Arbeitsmarktmonitor which is a panel study with interviews every 4 or 6 months. This data set is not as informative as the GSOEP (for example there is no monthly employment information), but contains considerably more observations. To correct for observed and unobserved selectivity due to panel attrition and due to OFT participation they model both processes using joint normality and a particular random effects specification for the joint error covariance matrix together with the process determining employment outcomes. From a methodological point of view this study certainly presents a huge improvement compared to the other evaluation studies done for East German labor markets after unification known to me, because it tries to correct explicitly and not in a purely mechanical manner for possible selection effects. However, it is the opinion of the author of this paper that it shares the problem of all model based evaluation precedures by identifing the estimation problem with a combination of (latent) linearity of conditional expections and distributional assumptions (joint normality and covariance restrictions) of the error terms. It is difficult to discuss the validity of these kind of identifying assumptions (partly necessary because of the not so informative data) in terms of the economic problem under investigation. However, their results depend on the premisses that an approach suggested by Heckman and Hotz (1989) to check the specification by comparing whether there is an effect of future training participation on current pre-training labour market outcomes, can be used to redefine training effects. However, results presented in Appendix B of this paper suggest that this may not always be appropriate.

5 Conclusion

The major empirical result of this paper is that no robust positive effects of OFT are found. There are three possible reasons for this: First of all, the true effects can be so small that they are impossible to determine with the available sample size. Secondly, there could be positive effects in the longer run that cannot yet be seen. Finally, it could be that there are no positive effects at all. This conclusion, if also confirmed for more recent courses, would have very serious implications for public policy. However, although the study raises serious doubts, one should be cautious to conclude that the training part of the active labor market policy (as defined in the Work Support Act, AFG) in East Germany has no positive impact even in the shorter run. The definition of off-the-job training used in this paper includes several courses that are not subsidized by the AFG. Immediate future research will be devoted to AFG subsidized courses only, and should provide more direct information about this issue.

The results are obtained by using the potential outcome approach to causality - first explicitly suggested by Rubin (1974) - as a general framework to define causal effects of off-the-job training on individual actual and future post-training labor market outcomes. The paper argues that due to the specific situation in East

Germany after unification and the rich data available, the assumption that the outcomes and the assignment mechanisms are independent conditional on observed attributes, including monthly pre-training employment status, is very plausible. Hence, the identification problem inherent in causal analysis is solved that way. Estimation is performed using a suitably adapted nonparametric matching approach which incorporates the (partial) propensity score as well as other attributes that could not possibly be captured by the partial propensity score, because they depend on the particular date of the beginning of the training. In conclusion, this nonparametric approach appears to be well suited for such an analysis.

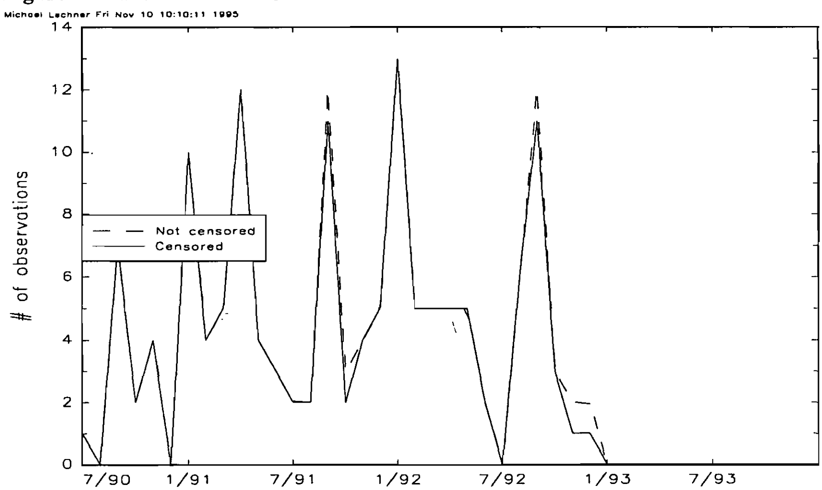
Interesting future research should investigate jointly the effects of different types of training, such as on-the-job training versus off-the-job training versus no training at all. Likewise, it could be an issue whether the quality of the publicly funded training did really improve during the transformation process, as claimed by official sources.

Appendix A: Data

This appendix briefly explains the coding of the start, duration, and end date of OFT courses. It also contains a histogram for the distribution of start dates and the ending dates in Figures A.1 and A.2. Furthermore, the exact definition of income variables used in the evaluations are given. Finally, Table A.1 shows descriptive statistics for all variables used in the estimation.

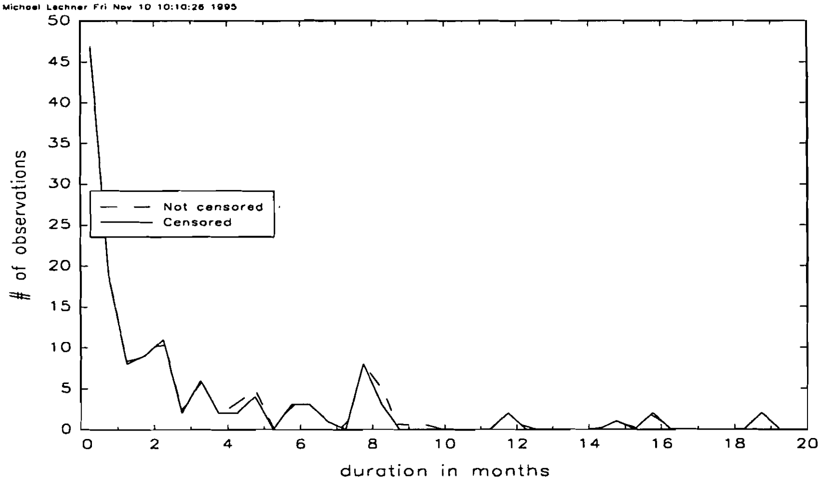
The first month of the course is directly indicated by the individual. When there are several courses classified as OFT, the start date is coded as the earliest one. The duration of each course is computed using the midpoint of the indicated duration interval (see footnote in Section 3) multiplied by the weekly hours. In cases of several OFT courses the single durations are added. The last month of each course is computed using the endpoint of the duration intervals to make sure that post-training outcomes are really *post*-training. Note that this is only important for courses with a duration of more than one month. As explained in the main text, the resulting measurement error for these courses (and some of the durations) is reduced by using additionally monthly calendar information on training. In cases of several courses the end date is coded as the end date of the last course.

Figure A.1: Distribution of OFT start dates



Note: Monthly information.

Figure A.2: Distribution of OFT durations



Note: A 0.5 month interval is used.

Table A.1: Descriptive statistics

	No OFT (1105)		OFT (122)	
Variable	mean/share in %	std	mean/share in %	std
Age in 1990	35.2	8.1	35.4	7.5
Gender: female	42		64	
Marital status in 1990				
married	78		78	
single	16		13	
divorced, separated	7		9	
Very desirable behavior / attitudes in society in 1990				
performing own duties	72		63	
achievements at work	72		72	
increasing own wealth	29		20	
Voluntary services in social organizations in 1990:	38		47	
Federal states (Länder) in 1990				
Berlin	7		13	
Brandenburg	15		18	
Mecklenburg-Vorpommern	10		6	
Sachsen	31		32	
Sachsen-Anhalt	20		15	
Thüringen	17		16	
Size of city / village				
< 2000	25		21	
2000 - 20000	28		34	
20000 - 10000	25		24	
> 100000	22		20	
Years of schooling (highest degree) in 1990				
12	17		31	
10	60		63	
8 or no degree	22		6	

Table A.1 to be continued ...

Table A.1: Descriptive Statistics: continued

	No OFT (1205)		OFT (122)	
Variable	mean/share in %	std	mean/share in %	std
Highest professional degree in 1990				
university [1]	11		25	
engineering, technical college[2]	16		33	
master of a trade / craft	6		6	
skilled worker[3]	64		34	
no degree	2		2	
Job position in 1990				
highly qualified, management	19		43	
master of a trade / craft[4]	8		7	
skilled blue and white collar[5]	57		40	
Job characteristics in 1990				
wage / salary per month	1240	381	1256	288
tenure in years	10.5		9.6	
temporary job contract	4		4	
professional degree in other than current profess.	36		31	
already fired	4		7	
training (unspecified) while full-time employed	7		16	
Profession in 1990 (ISCO)				
scientific, technical, medical	19		39	
production	43		13	
managerial	3		5	
administrative	9		11	
trade	5		2	
agriculture	3		2	
services	8		5	
services, incl. trade, administrative	23		21	
Memberships in 1990				
union	75		80	
professional association	7		8	
cooperative (LPG / PGH)	8		4	

Table A.1 to be continued ...

Table A.1: Descriptive Statistics: continued

	No OFT (1205)		OFT (122)	
Variable	mean/share in %	std	mean/share in %	std
Employer characteristics in 1990				
firm size (number of employees)				
0-19	10		10	
20-199	27		25	
200-1999	37		39	
2000 and more	26		26	
industrial sector				
agriculture	11		7	
energy and water	3		4	
mining	3		2	
heavy industry	10		4	
light ind., consumer goods, electronics, printing.	16		18	
machine building and vehicle construction	5		10	
construction	7		4	
trade	7		5	
communication, transport	8		1	
other services	11		13	
education, science	10		20	
health	7		9	
redundencies announced	46		52	
Finding a similar new job is (in 1990)				
impossible	11		16	
difficult	69		70	
easy	20		13	
Very worried about job security in 1990	37		39	
Optimistic about the future in general in 1990	17		18	
Not at all optimistic about the future in general in 1990	7		9	
Not enjoying work	5		6	

Table A.1 to be continued ...

Table A.1: Descriptive Statistics: continued

	No OFT (1205)		OFT (110)	
Variable	mean/share in %	std	mean/share in %	std
Very confused by new circumstances	5		4	
Income very important for subjective well-being	65		54	
Expectations for the next 2 years in 1990				
redundancies in firm: certainly	32		40	
redundancies in firm: certainly not	7		3	
losing the job: certainly	5		7	
losing the job: possibly	35		38	
losing the job: certainly not	12		7	
improvements in professional career: certainly	1		1	
improvements in professional career: certainly not	43		38	
decline in professional career: certainly	3		3	
decline in professional career: certainly not	49		42	
new profession: certainly	4		7	
new profession: certainly not	48		40	

Note: 1) University and 'Fachhochschule'; 2) 'Ingenieur- und Fachschule', not 1); 3) 'Berufsausbildung', 'Facharbeiter', 'sonstige Ausbildung', not 1), 2) or master of a trade / craft; 4) Includes 'Brigadier', 'Meister im Angestelltenverhältnis'; 5) 'Facharbeiter', 'Angestellte mit qualifizierter Tätigkeit'.
1990 relates to the date of interview which for almost is earlier than July 1990 (EMSU).

Gross monthly income is only measured for those employed. Due to the selection criteria that creates a sample of full-time employees in mid 1990 it is not a problem for 1990, but for the following years. For those unemployed unemployment benefits are computed using 67% of the last *gross* income, which should be a conservative estimate of the value of the gross equivalent for the actual net payment. However, it is assumed that all those unemployed remain eligible for unemployment benefits as opposed to unemployment assistance until 1993. This assumption is plausible, because of the special regulations for East Germans after unification (ratios of people receiving unemployment assistance relative to those receiving assistance or benefits: 1991: 3%, 1992: 8%, 1993: 14%; Statistisches Bundesamt, 1994, Table 6.15.4). It is assumed that these benefits increase yearly in line with the price index for private consumption.[48] This should again be a conservative estimate. After performing these imputations, it is ensured that income levels are not below average social assistance levels (Bundesministerium für Arbeit und Sozialordnung, 1994, Table 8.16A). Finally, all income variables are converted to 1993 DM by using the private consumption price index for East Germany (Bundesministerium für Arbeit und Sozialordnung, 1994, Table 6.9, and Institut der Deutschen Wirtschaft, 1994, Table 8).

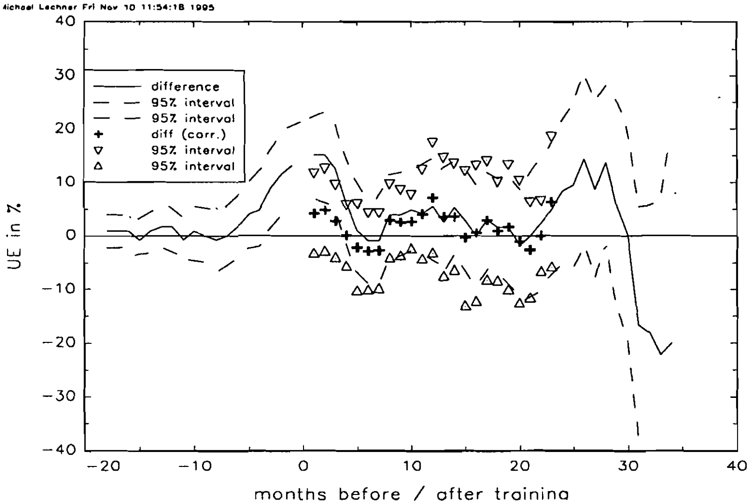
Appendix B: The danger of misinterpreting results based on mismatched pairs

I will use a particular example to show a fallacy that could appear when making before / after comparisons to evaluate the effect of training participation. Note that these kind of comparisons are implicitly or explicitly underlying many model based approaches suggested for example by Heckman and Hotz (1989) as well as standard fixed effects approaches.

Matching is performed exactly used for the figures in the main part of paper, with the only difference that the monthly pre-OFT information is ignored for the matching. This is similar to the situation with samples where no exact monthly information is available. Although the yearly pre-OFT information is used, Figures B.1 and B.2 show that the result is a substantial and significant mismatch for the unemployment as well as for the full-time employment variables. Somewhat surprising the conclusions from these evaluations are similar to the ones using the matching with the monthly information. However, if the OFT-effects are 'adjusted' by redefining effects as difference to the respective pre-training outcomes, suddenly OFT significantly decreases the probability of unemployment and increases the probability of full-time employment. However, from the figures presented in the main body of the paper we know that this not true. I should make it clear that this appendix is not intended to suggest that checking pre-OFT employment history is not important. Quite to the contrary, it is very important to do this, because this comparison reveals the success of making OFT and controls explicitly (as in this paper) or implicitly (as in model based approaches) comparable which is a prerequisite of a successful evaluation. Instead this appendix is intended to warn against the use of this useful specification test for redefining estimators. Clearly, the better option appears to resolve the problem of incomparability of the control and the OFT group. In this paper this is done by using the monthly information explicitly.

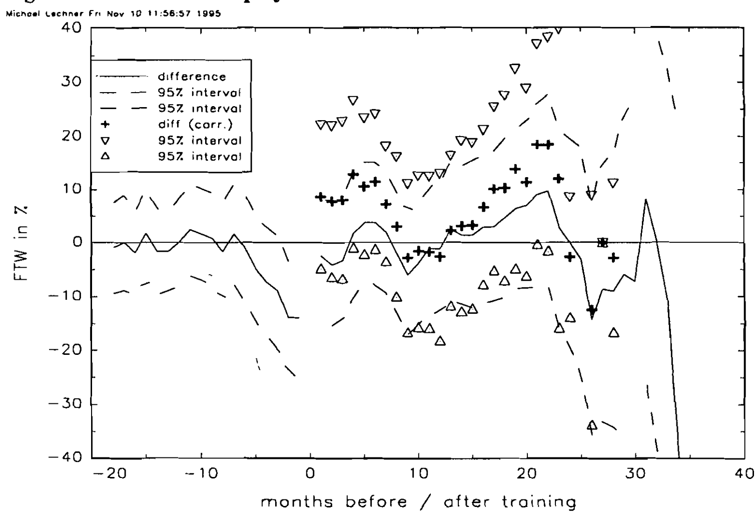
[48] It would be preferable to use the wage deflator, but the time series are not complete.

Figure B.1: Registered unemployment



Note: $N_t = 122$.

Figure B.2: Full-time employment



Note: $N_t = 122$.

Appendix C: Econometrics

C.1 Matching protocol

This section gives the details of the matching protocol used for the final evaluations.

Step 1: Split observations in two exclusive pools according to whether they participated in OFT (T-pool) or not (C-pool).

Step 2: Draw randomly an observation in T-pool (denoted by n_t) and remove from T-pool.

Step 3: Define caliper of partial propensity score for observation n_t in terms of the predicted index $v_i\hat{\beta}$ and its conditional variance $\mathrm{Var}(V\hat{\beta} \mid V = v_{n_t})$. The latter is derived from $\mathrm{Var}(\hat{\beta})$ by the delta method.

Step 4: Find observations in C-pool (denoted by *j*) obeying $v_j\hat{\beta} \in [v_{n_t}\hat{\beta} \pm c\sqrt{\mathrm{Var}(v_{n_t}\hat{\beta})}]$. The constant *c* is chosen such that the interval is identical to a 90% confidence interval around $v_{n_t}\hat{\beta}$.

Step 5: (a) If there is only one or no observation in this interval: find observation *j* in C-pool that is closest to observation *i*, such that it minimizes $(v_j\hat{\beta} - v_{n_t}\hat{\beta})^2$.

(b) If there are two or more observations in this set generated by Step 4: take these controls and compute the variables m in relation to the start date of observation n_t. Denote these and perhaps other variables already included in V as $\tilde{m}_j$ and $\tilde{m}_{n_t}$, respectively. Define a distance between each control *j* and *i* as $d(j,n_t) = (v_j\hat{\beta}, \tilde{m}_j)' - (v_{n_t}\hat{\beta}, \tilde{m}_{n_t})'$. Choose control *j* such that it has the smallest Mahalanobis distance $m(j,n_t) = d(j,n_t)'Wd(j,n_t)$ within the caliper. W denotes the inverse of the estimated variance of $(v\hat{\beta}, \tilde{m})'$ in the C-pool.

Step 6: Remove *j* from C-pool.

Step 7: If there are any observations in the T-pool left, start again with step 2.

This matching protocol is very close to the one proposed by Rosenbaum and Rubin (1985) and Rubin (1991). They find that this kind of protocol produces the best results in terms of 'match quality' (reduction of bias). The difference is that instead of using a fixed caliper-width (based on considerations about the true propensity score) for all observations, I allow the widths to vary individually with the precision of a monotone function of the partial propensity score (step 4). The (unbounded) linear index $v_{n_t}\hat{\beta}$ is used instead of the (bounded) partial propensity score $\Phi(v_{n_t}\hat{\beta})$. Matching on the latter with this kind of symmetric metric leads to an asymmetry when $\Phi(v_{n_t}\hat{\beta})$ is close to 0 and 1, depending on which side of the control j is. This is undesirable. Furthermore, defining the balancing score in terms of $(v_j\hat{\beta}, \tilde{m}_j)$ has also the advantage of making it easier to state under what conditions this type of condition has similar properties as conditioning on the (unknown and not estimable) propensity score itself.

C.2 Correction for mismatches and the modeling of conditional expectations

C.2.1 The linear case: homogenous effects

The question here is whether the price to pay for the use of the suggested regression methods to adjust for differences in attributes and course characteristics is the assumption of a homogenous treatment effect. This can be seen by considering whether such a regression can identify the mean causal effect θ^0, even when the individual causal effect is not constant in the population. Assume that the following linearity condition holds (given matching has already be performed in an unspecified way):

$$E(\Delta Y|S=1, \Delta X=\Delta X_i; \theta_i^0, \lambda^0) = \theta_i^0 + \Delta X_i \lambda^0 \tag{C.1}$$

Assume that the matches remain imperfect, so that ΔX_i may be different from 0 (for example when some components are continuous so that the probability of a perfect match is zero, even in very large samples).[49] Define the following population means: $\overline{\Delta Y} = E\Delta Y|S=1$, $\theta^0 = \overline{\theta} = E\theta|S=1$, $\overline{\Delta X} = E\Delta X|S=1$, $\overline{\Delta XX} = E(\Delta X'\Delta X)|S=1$ and $\overline{\Delta X\theta} = E[\Delta X'(\theta - \overline{\theta})]|S=1$. $\hat{\theta}_\infty$ denotes the population (probability) limit for the constant term of an OLS regression of a constant and the difference in attributes on the difference of an outcome. It can be computed by using the Frisch-Waugh-Lovell theorem (cf. Davidson and MacKinnon, 1993) or by applying the rules for the partial inversion of matrices directly. The result is:

$$\hat{\theta}_\infty = \theta^0 + [1 - \overline{\Delta X}(\overline{\Delta XX})^{-1}\overline{\Delta X}']^{-1}\overline{\Delta X}(\overline{\Delta XX})^{-1}\overline{\Delta X\theta}' \tag{C.2}$$

Therefore, generally the estimated OLS coefficient of the constant will not converge towards the population mean, unless $\overline{\Delta X\theta}'$ is zero. This is true when the

---

[49] Note that, for simplicity, this is again an argument about identification in the population only, so that the respective 'population' notation is used.

difference regressors and the causal effects are uncorrelated. A very important case is when θ is the same for all members of the population, another important case is the case of perfect matches.[50] However, note that the bias is reduced when the match quality increases and when effects are more homogenous in the (sub-) population. Similar arguments apply to the nonlinear case. Note that in practice this problem jointly with the linearity of the conditional expectation may be relevant even in large samples, when 'nonvanishing' regressors are considered instead of ΔX, like the characteristics of the courses or attributes of individuals participating in OFT.

C.2.2 The nonlinear case

As mentioned in Section 4 of this paper, several of the outcome variables are indicators so that $\Delta Y_i \in \{-1,0,1\}$. The elements of the set either denote a positive effect, no measurable effect, or a negative effect, respectively. The average causal effect is of the form given in (B.3):

$$\theta^0 = E(\Delta Y | \Delta X = 0, S = 1) = P(\Delta Y = 1 | \Delta X = 0, S = 1) - P(\Delta Y = -1 | \Delta X = 0, S = 1) \tag{C.3}$$

A consistent estimate of the average treatment effect can be obtained by substituting sample analogs for the population probabilities (B.4):

$$\hat{\theta}^M_{N_t} = \frac{1}{N_t} \sum_{n_t=1}^{N_t} [P(\Delta y_{n_t} = 1 | \Delta x_{n_t} = 0) - P(\Delta y_{n_t} = -1 | \Delta x_{n_t} = 0)] \tag{C.4}$$

Approximating differences of nonlinear probabilities (expectations) by a linear function as in the previous case may not be - for various reasons (e.g. Maddala, 1983) - a palatable option. Therefore, I choose a more parsimonious specification. In a first step a three-group-ordered probit model is estimated with Δy_{n_t} as dependent variable and Δx_{n_t} plus a constant as independent variables.[51] The asymptotic covariance matrix for the estimated coefficients of the ordered probit model are computed using the combination of OPG and expected hessian which has already been discussed in the context of the estimation of the propensity score. In the second step the above probabilities are directly derived from this model and computed for the individual observations using the estimated

[50] In this case the notation has somewhat to be changed to allow for noninvertible matrices.

[51] One bound and the variance of the underlying linear models are normalized (see Maddala, 1983, for details on the ordered probit model).

coefficients of the ordered probit model. Finally, the variance of $\hat{\theta}_{N_t}^{M}$ is derived from the variance of the estimated coefficients of the ordered probit model by the use of the delta method. Note that the functional form assumption for the conditional mean of ΔY_i is asymptotically unimportant as long as the differences in attribute (Δx_{n_t}) disappear.

The same approach is chosen to check whether the conditional expectations vary with either characteristics of the courses or with characteristics of the individuals having decided to participate in OFT. In the nonlinear case the average marginal effect of a continuous variable W can be defined and estimated as:

$$\gamma_{N_t}^{M} = \frac{1}{N_t} \sum_{n_t=1}^{N_t} [P(\Delta y_{n_t} = 1 | \Delta x_{n_t} = 0, w_{n_t} + a/2) - P(\Delta y_{n_t} = -1 | \Delta x_{n_t} = 0, w_{n_t} + a/2) - P(\Delta y_{n_t} = 1 | \Delta x_{n_t} = 0, w_{n_t} - a/2) + P(\Delta y_{n_t} = -1 | \Delta x_{n_t} = 0, w_{n_t} - a/2)] \tag{C.5}$$

where a is an appropriately chosen constant. The particular values of it used in the empirical study are given in the note to Table 6.

Equation (B.6) gives a similar expression for an average effect of an indicator variable D:

$$\gamma_{N_t}^{M} = \frac{1}{N_t} \sum_{n_t=1}^{N_t} [P(\Delta y_{n_t} = 1 | \Delta x_{n_t} = 0, d_{n_t} = 1) - P(\Delta y_{n_t} = -1 | \Delta x_{n_t} = 0, d_{n_t} = 1) - P(\Delta y_{n_t} = 1 | \Delta x_{n_t} = 0, d_{n_t} = 0) + P(\Delta y_{n_t} = -1 | \Delta x_{n_t} = 0, d_{n_t} = 0)] \tag{C.6}$$

Estimation of these marginal effects defined in (C.5) and (C.6) is accomplished as for the case of mismatch correction, but W or D are used as additional independent variables in the ordered probit estimation.

References

Angrist, J.D. and G.W. Imbens (1991): "Sources of Indentifying Information in Evaluation Models", *NBER Technical Working Papers*, 117.

Ashenfelder, O. and D. Card (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs", *The Review of Economics and Statistics*, 67, 648-660.

Bera, A., C. Jarques, and C. F. Lee (1984): "Testing the Normality Assumption in Limited Dependent Variable Models", *International Economic Review*, 24, 21-35.

Burtless and Orr (1986): "Are Classical Experiments Needed for Manpower Policy", *Journal of Human Ressources,* 21, 606-639.

Blundell, R.W., F. Laisney and M. Lechner (1993): "Alternative Interpretations of Hours Information in an Econometric Model of Labour Supply", *Empirical Economics*, 18, 393-415.

Bundesanstalt für Arbeit (BA, 1994a), *Geschäftsbericht 1993*, Nürnberg.

Bundesanstalt für Arbeit (BA, 1994b): *Berufliche Weiterbildung: Förderung beruflicher Weiterbildung, Umschulung und Einarbeitung im Jahr 1993*, Nürnberg.

Bundesminister für Arbeit und Sozialordnung (1991), *Übersicht über die Soziale Sicherheit, Textergänzung Kapitel 26:Übergangsregelungen für die neuen Bundesländer*, Bonn.

Bundesministerium für Arbeit und Sozialordnung (1994), *Statistisches Taschenbuch 1994: Arbeits- und Sozialstatistik*, Bonn.

Bundesministerium für Bildung und Wissenschaft (1994), *Berufsbildungsbericht 1994*, Bad Honnef: Bock.

Card, D. and D. Sullivan (1988): "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment", *Econometrica*, 56, 497-530.

Conrad, C., M. Lechner, and W. Werner (1995): "The Fall of the East German Birth Rate After Unification: Crisis or Means of Adaptation", *Center for European Studies Working Paper Series* #5.6, Harvard University.

Dagenais, M. G. and J.M. Dufour (1991): "Invariance, Nonlinear Models, and Asymptotic Tests", *Econometrica*, 59, 1601-1615.

Davidson, R. and J.G. MacKinnon (1984): "Convenient Specification Tests for Logit and Probit Models", *Journal of Econometrics*, 25, 241-262.

Davidson, R. and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Dehejia, R. and S. Wahba (1995a): "A Matching Approach for Estimating Causal Effects in Non-Experimental Studies", Harvard University, *mimeo.*

Dehejia, R. and S. Wahba (1995b): "Causal Effects in Non-Experimental Studies", Harvard University, *mimeo.*

Deutsches Institut für Wirtschaftsforschung (DIW, 1994), *Wochenbericht, 31/94*, Berlin.

Ehrenberg, R.G. and R.S. Smith (1994): *Modern Labor Economics: Theory and Public Policy*, 5th ed., New York: HarperCollins.

Gabler, S., F. Laisney and M. Lechner (1993): "Seminonparametric Estimation of Binary-Choice Models With an Application to Labor-Force Participation", *Journal of Business and Economic Statistics*, 11, 61-80.

Garfinkel, I., C.F. Manski and C. Michalopoulos (1992): "Micro Experiments and Macro Effects", in: C.F. Manski and I. Garfunkel (eds.), *Evaluating Welfare and Training Programs*, Cambridge: Harvard University Press, 253-273.

Fitzenberger, B. and H. Prey (1995): "Assessing the Impact of Training on Employment: The Case of East Germany", *Unpublished manuscript*, University of Konstanz..

Heckman, J.J. and V.J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, 84, 862-880 (includes comments by Holland and Moffitt and a rejoinder by Heckman and Hotz).

Heckman, J.J. and R. Robb (1985): "Alternative Methods of Evaluating the Impact of Interventions", in: J.J: Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, New York: Cambridge University Press.

Holland, P.W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Society*, 81, 945-970 (includes comments by Cox, Granger, Glymour, Rubin and a rejoinder by Holland).

Hübler, O. (1994): "Weiterbildung, Arbeitsplatzsuche und individueller Beschäftigungsumfang - eine ökonometrische Untersuchung für Ostdeutschland", *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 114, 419-447.

Institut der Deutschen Wirtschaft (1994): *Zahlen zur wirtschaftlichen Entwicklung der Bundesrepublik Deutschland 1994*, Köln.

Infratest Sozialforschung (1990, 1991, 1992, 1993): *Das sozio-ökonomische Panel - Ost, Welle 1, Welle 2, Welle 3, Welle 4*, Anlagenbände zum Methodenbericht, München.

Imbens, G.W. and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 446-475.

König, H. and M. Lechner (1994): "Some Recent Developments in Microeconometrics - A Survey", *Swiss Journal of Economics and Statistic,* 130, 299-331.

LaLonde, R.J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 76, 604-620.

Lechner, M. (1991): "Testing Logit Models in Practice", *Empirical Economics*, 16, 177-198.

Lechner, M. (1995): "Some Specification Tests for the Panel Probit Model", forthcoming in: *Journal of Business and Economic Statistics*, 13.

Lynch, L.M. (1992): "Private Sector Training and the Earnings of Young Workers", *The American Economic Review*, 82, 299-312.

Manski, C.F. and I. Garfinkel (1992): *Evaluating Welfare and Training Programs*, Cambridge: Harvard University Press.

O'Higgins, N. (1994): "YTS, Employment, and Sample Selection Bias", *Oxford Economic Papers*, 46, 605-628.

Orme, C. (1988): "The Calculation of the Information Matrix Test for Binary Data Models", *The Manchester School*, 56, 370-376.

Orme, C. (1990): "The Small Sample Performance of the Information Matrix Test for Binary Data Models", *Journal of Econometrics*, 4, 529-559.

Pannenberg, M. and C. Helberger (1994): "Kurzfristige Auswirkungen staatlicher Qualifizierungsmaßnahmen in Ostdeutschland: Das Beispiel Fortildung und Umschulung", to appear in: *Schriftenreihe des Vereins für Sozialpolitik.*

Pischke, J.-S. (1994): Continous Training in Germany, *mimeo.*

Rosenbaum, P.R. (1984): "From Association to Causation on Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignement", *Journal of the American Statistical Association*, 79, 41-48.

Rosenbaum, P.R. and D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrica*, 70, 41-50.

Rosenbaum, P.R. and D.B. Rubin (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, 39, 33-38.

Rosenbaum and Rubin (1985): "The Bias due to Incomplete Matching", *Biometrics*, 41, 103-116.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association*, 74, 318-328.

Rubin, D.B. (1991): "Practical Implications of Models of Statistical Inference for Causal Effects and the Critical Role of the Assignement Mechanism", *Biometrics*, 47, 1213-1234.

Sobel, M.E. (1994): "Causal Inference in the Social and Behavioral Sciences", in: G. Arminger, C.C. Clogg and M.E. Sobel (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press.

Statistisches Bundesamt (1994), *Statistisches Jahrbuch für die Bundesrepublik Deutschland, 1994*, Stuttgart: Metzler-Pöschel.

Wagner, G.G., R.V. Burkhauser and F. Behringer (1993): "The English Language Public Use File of the German Socio Economic Panel", *Journal of Human Ressources*, 28, 429-433.