Institut für
Volkswirtschaftslehre
und Statistik

No. 539-96

An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany

Michael Lechner

Beiträge zur
angewandten
Wirtschaftsforschung

Universität Mannheim
A5, 6
D-68131 Mannheim

An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany

Michael Lechner*

Universität Mannheim

This version: September 1996

DISCUSSION PAPER (BEITRÄGE ZUR ANGEWANDTEN WIRTSCHAFTSFORSCHUNG, # 539-96), UNIVERSITY OF MANNHEIM

Comments welcome

Address for correspondence:

Michael Lechner

Fakultät für Volkswirtschaftslehre

Universität Mannheim

D-68131 Mannheim

Germany

Email: lechner@haavelmo.vwl.uni-mannheim.de

WWW: http://www.vwl.uni-mannheim.de/lehrst/ls_oek/NBL_1

* Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. I thank the DIW for supplying the data of the German Socio-economic Panel. Furthermore, I thank Klaus Kornmesser for competent help with the data and Martin Eichler, as well as participants of seminars at the Humboldt Universität zu Berlin, the Europa-Universität in Frankfurt / Oder, and the University of Magdeburg, as well as of the Sixth Conference on Panel Data in Amsterdam, for helpful comments and suggestions. Discussions with Bernd Fitzenberger and Hedwig Prey were also very valuable for this work. All remaining errors are my own.

Abstract

This study analyses the effect of public sector sponsored continuous vocational training and retraining in East Germany after unification with West Germany in 1990. It presents estimates of the average gains from training participation in terms of earnings, employment probabilities and career prospects after the completion of training. The data is from the German Socio-Economic Panel (GSOEP, 1990-1994). The GSOEP allows to observe individual behaviour on a monthly, respectively yearly, basis. The results suggest that despite public expenditures of more than DM 25 bn (1991 to 1993), there are no positive effects in the first year after training, but that participants expect positive effects over a longer time horizon. The latter however is beyond the sampling period.

Keywords

Evaluation of training programs, causal analysis, panel data, matching on the propensity score, East German labour markets.

JEL classification: J24, J31, J60, C33

1 Introduction

Unification of the East and West German economies in July 1990 - Economic, Social and Monetary Union - came as a shock to the formerly centrally planned East German economy. The almost immediate imposition of the West German type of market economy with all its distinctive institutional features and its relative prices led to dramatic imbalances on all markets. For example the official unemployment rose from about 2% in the GDR to more than 15% in 1992. It remained on that high level for the following years. To avoid higher unemployment as well as to adjust the stock of human capital to the new demand structure the state conducted an active labour market policy. The evaluation of the continuous vocational training and retraining part of that policy is the focus of this paper. Since more than DM 25 bn are devoted to this purpose until the end of 1993, the need for an evaluation is obvious. This paper presents estimates of the average individual gains to the workers of the former GDR participating in such training beginning between July 1990 and December 1992. The targets of the evaluations are labour market outcomes after the completion of the training, such as earnings, labour market status, and career prospects.

In typical microeconometric evaluations of training programs, outcomes measured for the sample undergoing the training are compared to outcome measures for a *comparable* group, 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 randomisation, for instance. Hence, such a control group should not systematically differ from the trainees. This simplifies the evaluation dramatically, because the difference of simple sample means in the trainee and the control population is an unbiased and consistent estimator for the average effect of training for the trainees. However, the huge time lag between the beginning of such an experiment and the results of the evaluations is one reason why conducting an experiment was never an option in East Germany.[1] In 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 to avoid potentially serious *sample selection biases*. Holland (1986) and Heckman and Hotz (1989) provide extensive and excellent discussions on these issues.

Various model-based procedures are suggested in the econometrics' literature to avoid such biases (see for example Heckman and Hotz, 1989, or Heckman and Robb, 1985).[2] However, Ashenfelter and Card (1985) and LaLonde (1986) - among others - conclude 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

[1] The state of the discussion about whether it is advantageous or not to base evaluations on social experiments can be found in Burtless (1995) and Heckman and Smith (1995a).

[2] Chapter 1 in Bell, Orr, Blomquist, and Cain (1995) provides a more complete account of the development of the econometric evaluation literature.

that social experiments are necessary to evaluate training programs. 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 analysed. Recently, Dehejia and Wahba (1995a, 1995b) - using an approach very similar to the one chosen here - re-evaluate the LaLonde (1986) data. By using nonparametric techniques - partly to be discussed later - they come to far more positive conclusions about the potential quality of inferences based on observational data than LaLonde (1986) himself. This issue is not yet settled.

Many problems with the statistical modelling procedures stem from the fact that the data does not provide sufficient information on all important factors that influence program participation as well as labour market outcomes. Then it is necessary to introduce unobserved ´error terms´ and to model their joint distribution with the variables of the model. It is one of the major advantages of the data used in the empirical analysis in this paper that it is a highly informative panel data set. This data set is a random sample from the East German population and hence contains both trainees and non-trainees. It contains many socio-economic variables and allows for example to track down employment histories on a monthly basis beginning twelve months prior to the economic union. Therefore, I do not need to introduce error terms and I can concentrate on controlling for observable difference of trainees and controls. Since this is done nonparametricly by extending the methods proposed by Rubin (1979) and Rosenbaum and Rubin (1983, 1985a), the results should be reasonably immune to the above criticisms.

Although, there is a large number of evaluation results for US-training programs available (e.g. LaLonde, 1995), there are only very few econometric evaluations of training in East Germany. The results in this paper do not confirm previous positive findings of the effectiveness of training in East Germany (e.g. Fitzenberger and Prey, 1995, 1996, Pannenberg and Helberger, 1994, Pannenberg, 1995). Although there are only few studies conducted so far, they differ in many respects ranging from the database to the implementation of the evaluation, treatment of the selection problems, and the definition of the training itself. 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 modelling the distributions of the outcome variables or error terms given certain covariates. In a very similar way as the companion paper Lechner (1995a), this paper explicitly avoids imposing such restrictions and puts emphasis of the particular notion of causality behind the results. However, Lechner (1995a) investigates a different sort of training based on a different kind of data. That paper focuses on off-the-job training excluding retraining but including many short spells of training not publicly subsidised. In contrast to the mentioned papers, this study here focuses exclusively on public sector sponsored training. Hence it can be used as one piece of evidence to answer the question whether the huge amount of public money has been spent wisely.

The paper contributes to the ongoing discussion of the effectiveness of the training in East Germany by analysing 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 labour market. The findings suggest that in the short run public sector sponsored training has a negative impact, because it reduces the job search efforts for the trainees during training as compared to a comparable spell of unemployment. However, several months after the end of the training there are no statistically significant differences between the controls and the trainees. Hence, training has no positive effect on the trainees' labour market outcomes. However, there is some evidence that trainees expect positive returns over a longer time horizon that however is beyond the sampling period available for this analysis. If these expectations materialise, then future evaluations will find a positive effect of training.

The paper is organised as follows: The next section outlines basic features of the East German labour markets after unification. It includes a brief discussion of the training part of the active labour market policy. Section 3 introduces the longitudinal data used in this study and presents several characteristics of the sample chosen. Issues related to the econometric methodology and the empirical implementation are discussed in the four subsections of Section 4. The first subsection details the causality framework used and discusses general conditions for the identification of average causal effects. The following two subsections identify factors influencing (potential) labour market outcomes as well as training participation and show that shocks, such as the occurrence of unemployment, play a very important role for the participation probability. An adaptation of the matching approach is suggested. It allows for these 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 consists of several more technical parts concerning the econometric methods. Appendix C presents the detailed results of the estimation as well as specification tests for the estimation of the (partial) participation probabilities. Appendix D shows some details about the differences between the training population and several potential control populations. Finally, Appendix E contains additional evaluation results.

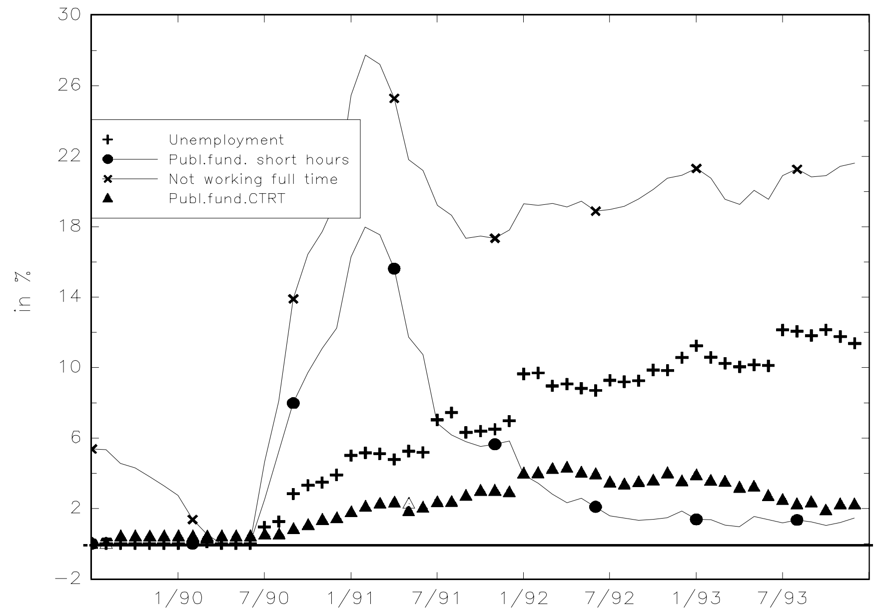
2 East German labour markets in transition

The Economic, Monetary, and Social Union in July 1990 (EMSU) came as a shock to the formerly centrally planned East German economy. The almost immediate imposition of the West German type of market economy with all its distinctive institutional features and its relative prices led to a large drop of the GDP in 1990. After that drop the GDP grew by about

6-8% p.a. in the period 1991 to 1994.[3] In the same period average earnings per worker increased from about 48% of the West German level in 1991 to about 73% of that level. However, labour productivity increased only from about 31% to about 51%. As a result there were severe disequilibria in the labour market. The labour force dropped by about 2 million people from 8.3 million in the second half of 1990 to 6.3 million in 1992. It remained approximately stable afterwards. Similarly, (official) unemployment rose from about 2% in the GDR to more than 15% in 1992. It remained on that high level for the following two years. To avoid higher unemployment, among other objectives, the state conducted an active labour market policy. The evaluation of the continuous vocational training and retraining part (CTRT) of that policy is the focus of this paper.

Figure 1 shows monthly pre- and post- unification developments for various indicators, such as unemployment, publicly funded short-time work (STW, "Kurzarbeit") and the proportion of those individuals not in full-time employment. The figure describes the population that I am most interested in: individuals not younger than 20 and not older than 50 in 1990. They worked full-time just before unification, lived in East Germany at least until 1993, and are not in bad health conditions. These individuals constitute the active labour force of the late GDR, and they are too young to consider (regular) retirement in the next years after unification.

Figure 1: Labour market states and public sector sponsored training and retraining (CTRT)



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

[3] The data used in 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), Bundesministerium für Wirtschaft (1995), and Bundesminister für Arbeit und Sozialordnung (1991).

Figure 1 shows that for this population full-time employment (100 minus share not full-time-employed; denoted by a line with ✖´s) declines from 100% in mid 1990 to about 72% in early 1991 and than stabilises at around 80%.[4] A very significant proportion of the early fall is absorbed into STW (•), which means a reduction of working hours in the firm accompanied by a subsidy from the labour office to compensate employees for the otherwise occurring earnings loss. In particular in the first year after unification this reduction of working hours could be substantial.[5] However, STW was only temporarily an important tool of the active labour market policy. It became unimportant after 1991. As a result of the decline of STW after early 1991 as well as of the worsening general labour market conditions, the unemployment rate (+) increased steadily up to about 12 % in the end of 1993.[6] Finally, the number of people taking part in CTRT (Δ**)** increased steadily after unification and reached its peak in early 1992 with about 4% of those full-time employed in 1990 and fell thereafter due to policy changes.

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 labour offices, conducted an active labour 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 STW (14 bn DM) and Public Work Programs (ABM, 26 bn DM). However, a discussion of the latter two policies is beyond the scope of this paper.

Continuous vocational training and vocational retraining for a new profession are subsidised by the labour office under provisions of the Work Support Act ("Arbeitsförderungsgesetz", AFG). They form the largest 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 (RT), and (iii) employers are subsidised for a limited period to provide on-the-job training for individuals facing difficult labour market conditions to allow them to familiarise themselves with the new job (FJ). The focus of this paper is on the first and second group, that account for more than 90% of all entries in these subsidised courses. The share of the third group is small and declining (1991: 14.9%, 1992: 12.8%, 1993: 10.6%, 1994: 6.7%). The major difference between groups (i), (ii) versus (iii) is that (i) and (ii) are typically classroom (i.e. off-the-job) training (99%), whereas (iii) is always on-the-job training.

---

[4] The definition of full-time work used here includes Public-Work-Programs (ABM) which account for about 5-10% of full-time employment. After the decline of STW, it could partly be seen as a substitute for it.

[5] 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.

[6] Unemployment, STW, and CTRT numbers are lower than the official rates for the total population, because of the age restriction and because different definitions of the relevant populations appearing in the denominator of the ratios. Furthermore, CTRT includes only individuals receiving compensation for potential earnings losses ("Unterhaltsgeld"). For a precise definition see below.

When certain conditions are met, the labour office pays for the provision for the training as well as for the foregone earnings. These conditions are related to the employment history, the approval of the course by the labour office, and the potential termination of unemployment or the avoidance of a possibility to become unemployed soon. Until 1993 the last principle has been applied using a broad interpretation in East Germany, so that it includes more groups of the labour force than in the West. In most cases the payments cover almost all the costs for the provision of the course as well as 65% or 73% of the previous net earnings ("Unterhaltsgeld", UHG). For comparison, this is about 10% higher than unemployment benefits (60% or 67%). Additionally, until the end of 1991, workers participating in CT during STW obtained a slightly increased STW compensation. During 1993 and in the beginning of 1994 the rules have been significantly tightened to make sure that the now reduced budget is more precisely targeted to those being unemployed. Also from 1994 on, the amount of UHG individuals can receive is not larger than unemployment benefits. The current analysis is based on recipients of UHG (including STW with training) who began their training not later than 1992. In the following, this group is appreviated as CTRT.

Table 1 gives the (official) numbers of entrants into different parts of CTRT and shows the ratios of previously unemployed participants and the average shares of participants obtaining UHG from 1991 to 1993. In 1990 they were not important at all. CT is divided in two subgroups. The second subgroup covers training with very short duration (some days), which is no longer supported by the labour office after 1992. In 1991 and 1992 the number of entrants is very large and close to about 10% of total employment each year. In 1993 the policy changes led to a significant drop of entrants. Interestingly, the share of rejected applications for any sort of CTRT subsidy is very low (1991: 1.8%, 1992: 5.5%, 1993: 7.7%). However, particularly in 1993 many potential applicants may not have applied, because they were informed before a formal application that they will not qualify.

Table 1: Entries into continuous vocational training (CT) and retraining (RT) 1991 to 1993

	1991			1992			1993		
	entries	UE	UHG	entries	UE	UHG	entries	UE	UHG
	x 1000	share in %		x 1000	share in %		x 1000	share in %	
CT, no §41a AFG	442	35+)	n.a.	462	70	85*)	182	74	82+)
CT, §41a AFG only	187	100	n.a.	129	100	85*)	0	-	-
RT	130	35+)	n.a.	183	81	85*)	81	91	82+)
Total employment	7219	10	-	6344	16	-	6128	16	-

Note: *UE*: Unemployed before entering; *UHG*: Recipient of UHG ("Großes UHG"); *§41a AFG* denotes very short term courses for the unemployed to improve their job search skills, +) aggregated over 2 categories; *) aggregated over all categories; *n.a.*: not available; Sources: BA (1992), BA (1993a), BA (1993b), BA (1994), BA (1995), own computations.

The share of participants who were unemployed before CTRT increases over time due to the worsening situation of the labour market as well as the tightening of the admission rules set by the BA. The share of UHG recipients is above 80% for 1992 and 1993.

The BA is the most important source of finance for CTRT. Table 2 shows the expenditure of the BA for CTRT from 1991 to 1993 (they were very small in 1990). In 1992 and 1993 more than 60% of the total expenditure of about DM 10 bn was allocated to direct earnings support for participants. Most of the remaining share covers direct costs of CTRT, and a small proportion goes as direct support to the providers of the training. Taking together the number of entries in Table 1 and the expenditures in Table 2 suggests an average cost per entry of about 14.200 DM. However, this is only a lower bound for the true average cost per head, because several of the CTRT entries will still be in CTRT after 1993.

Table 2: Expenditure of the labour office (BA) for CTRT 1991 to 1993

	1991	1992	1993
UHG in bn DM	1.6	6.0	6.6
Other expenditure in bn DM	2.7	4.7	3.7
Institutional support in bn DM	0.2	0.1	0.1
Total expenditure for CTRT in bn DM	4.4	10.8	10.4
Share of CTRT expenditures of total expenditure of BA in East Germany	15 %	23 %	21 %

Note: BA (1995), Table 34. Own calculations. Spending for FJ is excluded.

For more details about training providers and their financing the reader is referred to Müller (1994). Buttler (1994) presents more information about the different sources of financing the active labour market policy in East Germany.

3 Data

The sample used for the empirical analysis is drawn from the German Socio-Economic Panel (GSOEP), which is very similar to the US Panel Study of Income Dynamics. 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).

A very useful characteristic of this panel survey is the availability of monthly information between yearly interviews. This covers different employment states and income categories. The information is obtained by retrospective questions about what happened in particular months of the previous year. Figure 2 shows a sketch of one type of 'calendar': the income calendar.

Figure 2: Selected items of the retrospective questions about income in the 1993 questionnaire (income calendar)

"Please indicate for each month of the previous year (1992) whether you had some income of the type or the source given on left hand side of the following calendar:"

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
employment as employee												
self-employment												
...												
unemployment benefits												
UHG (labour office)												
...												
no such income												

Note: For the complete questionnaires see Infratest Sozialforschung (1990, 1991, 1992, 1993, 1994). Own translation (summarised).

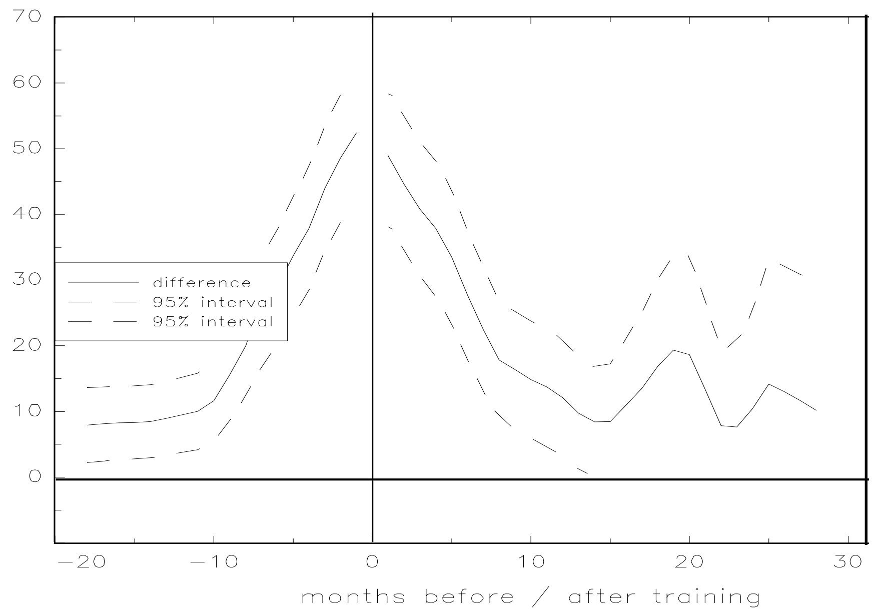
The related employment calendar contains information on the employment status, such as full time employment, part time employment, STW, vocational training, schooling, and so on. These calendars allow a precise observation of the individual employment states and income sources before and after CTRT. This kind of information will figure prominently in the empirical analysis.

A balanced sample of all individuals born between 1940 and 1970 who responded in in the first four waves is selected. The upper age limit is set to avoid the need of addressing early retirement issues. The population of interest is the one that formed the labour force of the GDR, therefore it is required that all selected individuals work full-time just before unification. Furthermore, the self-employed in the former GDR (2% of non-CTRT sample), those individuals working in the GDR in the industrial sectors energy and water (3%), mining (3%), and health (8%), and persons stating in 1990 that expect certainly improvements in their professional career in the next two years (2%) are not observed taking part in CTRT, so they are deleted from the sample (see Table A.1 in Appendix A for details). 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. In order to be able to control for the entire labour market history before CTRT (beginning in mid 1989) - necessary to control for the selection issues - it is required that all individuals answer the relevant survey questions in all four yearly surveys. Since the fifth survey (1994) is only used to measure post-CTRT labour market outcome, it is not necessary to impose such a requirement.

The income and employment calendars are used to define the training measure CTRT. Individuals are considered to participate in CTRT if they receive UHG or obtain continuous vocational training during STW. It is required that the training periods starts after unification but not later than March 1993 to ensure that all CTRT is obtained under the regime valid up to 1992.

The mean (std.) of the durations is about 12 (7). 10% of the CTRT spells have a duration of no more than 3 months, 25% of no more than 6 months, 65% of no more than 12 months, and 95% of no more than 24 months. Comparing these number with the durations of CT, RT and FJ (FJ durations are 6 to 12 months) spells as given by the BA, it is found that a substantial part of short spells is missing from the sample. However, note that not only the comparison is not really valid because of the inclusion of FJ in the official numbers, but also that there are other issues related to the questionnaire (calendar): Firstly, the fact that it is retrospective information about last year may result in participants forgetting very short training spells. Secondly, it may be that respondents do not bother to tick boxes for a particular month in case of very short spells of some days. Thirdly, multiple spells are added (10%) which increases duration per spell. However, by omitting these very short spells that may be related to AFG §41a (no longer supported by BA after 1992!) the following empirical analysis is more focused on longer spells that obviously absorb a much larger amount of resources. It is these longer CTRT spells that are a priori considered to be most effective. More details on the CTRT spells can be found in Appendix A.

Figure 3: Share of registered unemployed before and after CTRT for CTRT participants in %



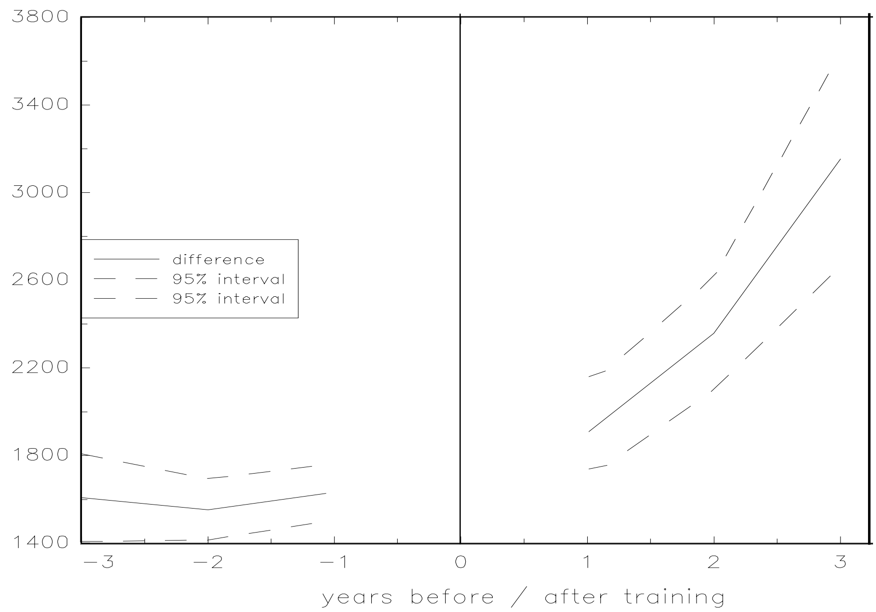
Note: Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure 3 shows CTRT participants that are unemployed in any given number of months before or after CTRT. There is a dramatic surge in unemployment in 10 months prior to CTRT culminating in an unemployment ratio of about 51% in the month just prior to training. The respective rates for full-time employment are 24% (Figure A.5) and 73% for the combined rate of unemployment or STW (Figure A.6). From these figures it is already clear that CTRT participants are not a random sample from the population. This is of course at least with respect to unemployment history intentional, because CTRT is - at least in principle - targeted to the unemployment and to those under a general threat of unemployment.

Considering the post-CTRT period, it appears that many CTRT participants find jobs fairly quickly. Whether they do this fast enough to make up for the time *lost* for search during CTRT participation, which is on average twelve months, will be seen below. The BA also publishes the share of unemployed and for particular points in time the share of unemployed six months after the end of CTRT. Although an exact comparisons is difficult, because of the different concepts of time used, their numbers are within the ranges shown in Figure 3 (see Buttler and Emmerich, 1994, Blaschke and Nagel, 1995, IAB ,1995, p.134).

Figure 4 shows a similar plot for the real earnings variable. Note however that earnings are measured only at the yearly interview. Due to data availability, the deflator used is the cost-of-living price index. Hence, the sharp increase of average earnings after CTRT may merely reflect the divergence of wage growth and cost of living together with the increasing proportion of people working after CTRT. The relative flat behaviour of the curve before CTRT should reflect the increasing unemployment rates together with rapid wage growth. In this figure earnings for non-workers are coded as unemployment benefits (see Appendix A for details), but the same shape of the curve emerges when earnings for non-workers are coded as zeros.

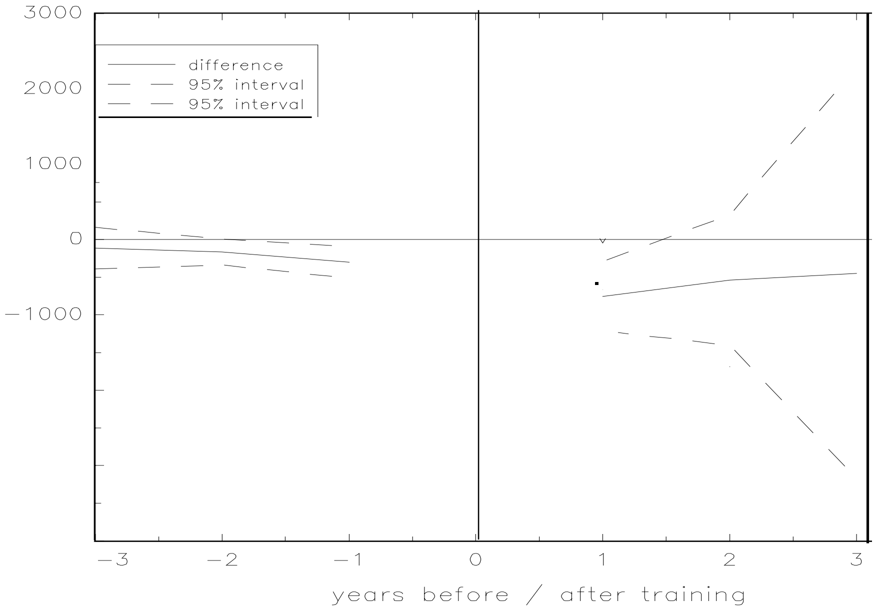
Figure 4: Gross earnings (in 1993 DM) before and after CTRT for CTRT participants



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

When comparing pre- and post CTRT earnings of the trainees to a randomly chosen group of individuals not participating in CTRT, Ashenfelter´s (1978) dip in earnings prior to a training program appears. This dip is clearly due to the increased proportion of unemployed persons prior to training (see Figure 3). Heckman and Smith (1995b) noted that when earnings dynamics are largely driven by unemployment dynamics, controlling for lagged earnings is not sufficient to evaluate the impact of CTRT.

Figure 5: Difference of gross earnings (in 1993 DM) before and after CTRT between CTRT participants and random control group



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

Considering other socio-economic variables, there does not appear to be a large age difference, but there are far more women in CTRT 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 CTRT. More details on the socio-economic variables can be found in Table A.1 of Appendix A.

4 Econometric methodology and empirical implementation

The previous section showed that before / after comparison are insufficient to control for the selectivity problem which is clearly visible in the data. In this section I introduce a notation which allows to address this problem directly. Section 4.1 begins with a brief discussion of causal modelling and the restrictions that are used to identify the training effects. Subsection 4.2 shows that this identifying assumption is reasonable for the problem analysed 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 CTRT 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 B and additional results can be found in Appendices C and D.

4.1 Causality, potential outcomes, identification and balancing scores

The empirical analysis tries to answer questions like "What is the average gain for CTRT participants compared to the hypothetical state of nonparticipation?" 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 CTRT effects the respective outcomes, such as earnings or employment status. This is very different from asking whether there is an empirical association, typically related to some kind of correlation, between CTRT and the outcome. Therefore, I do not try to answer the question whether CTRT is associated with higher earnings for example, but whether the effect of CTRT is higher earnings.[7]

The framework that will serve as guideline for the empirical analysis is the potential-outcome approach to causality suggested by Rubin (1974). This idea of causality is inspired by the set-up of experiments in science. Its main building blocks for the notation are *units* (here: individuals), for which I will assume that they belong to the large population defined in the previous section, *treatment* (participating in CTRT or not participating in CTRT) and potential *outcomes*, which are also called *responses* (earnings, labour market states, either at a particular time, or at a particular span of time after having completed CTRT). Y^t and Y^c denote the outcomes (t denotes treatment, c denotes control, i.e. no treatment).[8] Additionally, denote variables that are unaffected by treatments - called *attributes* by Holland (1986) - by X. Attributes are exogenous in the sense that their potential values for the different treatment states coincide (X^t=X^c). It remains to define a binary *assignment* indicator S, which determines whether unit n gets the treatment ($S = 1$) or not ($S = 0$). If the unit participates in CTRT the actual (observable) outcome (Y) is Y^t, and Y^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* $(y_n^t, s_n = 0)$ or $(y_n^c, s_n = 1)$ to the observable outcome (y_n) is never observed.

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. Furthermore, it is a helpful device to design 'informative' social experiments, or - if this is not possible or not de-

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

8 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.

sirable - 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. 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 emphasises 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 the treatment group. This means that they should be identical in all aspects effecting the training decision as well as the potential outcomes. If it is not possible to find such individuals, additional assumptions have to be invoked to - in some sense - adjust for their dissimilarity.

Using the previous notation, the estimand of interest, which is the average causal effect of CTRT, is denoted by θ^0 and defined in equation (1):

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

The short hand notation *E(· |S=1)* denotes the mean in the population of all units 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 is how θ^0 can be identified from a large random sample of the population. The problem is the term $E(Y^c | S = 1)$, because the pair $(y_n^c, s_n = 1)$ is not observable. 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 using the observable nontreatment outcomes of the untreated population $(y_n^c, s_n = 0)$ in different ways.

If there is random assignment as in a suitably designed experiment, then the potential outcomes would be independent from the assignment mechanism and $E(Y^c | S = 1) = E(Y^c | S = 0)$. Then the untreated population could be used as the 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, as has been shown in the previous section the assumption of random assignment is hardly satisfied. There appear to be several variables which influence assignment as well as outcomes.

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

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

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 non-treatment outcome conditional on the value of a covariate or attribute (CIA).[9] 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 the population of participants ($S=1$). The following sections show that this 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 nontreatment outcomes. This implies that there is no important variable left out which influences nontreatment outcomes as well as assignment given a fixed value of the relevant attributes.[10] There are different restrictions (e.g. Angrist and Imbens, 1991, Imbens and Angrist, 1994, Heckman and Hotz, 1989, Heckman and Robb, 1985) available to solve the identification problem, but this one appears to be the most fundamental in its close resemblance of the experimental context, and, given the data available and the nature of the objective pursued here, it is best suited for this context.

A brief comparison with standard econometric approaches is in order. When a prototypical econometric approach is used, based on modelling particular moments of the potential outcomes (e.g. Heckman and Hotz, 1989, Heckman and Robb, 1985, Maddala, 1983, and many others), the same issues as mentioned above need to be addressed to make causal 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 approach is more likely to hide important issues related to the causal or noncausal nature of the intended inference, because it bases identifying assumptions on unobservables of the assumed models. But only in rare cases has the researcher a precise idea what the error term really embodies. Substantive relationships between important components of the analysis, such as assignment mechanisms are easier to analyse, to communicate to non-econometricians, and to discuss. This paper goes the latter route.[11] Finally, it is also worth pointing out that no assumptions about any constancy of the treatment effects for different individuals are made here.

Rosenbaum and Rubin (1983) show 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 non-trivial probability $(0 < P(x) < 1)$ of being assigned to the treatment conditional on the possibly

---

9 It is important to note that this does **not** exclude the case of the treatment outcome being correlated with the selection mechanism. In practice this may be quite important because having a higher treatment outcome as some somebody who has the same nontreatment outcome may an (partially) efficient and often used selection rule.

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

11 See also Angrist (1995) for a discussion of model-based versus nonparametric evaluation in econometrics.

high dimensional vector of characteristics *x*. Furthermore, let *b(x)* be 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*, then they are also independent of the assignment mechanism conditional on *b(X)=b(x)*. Obviously, this result applies also to the case with only the nontreatment outcome being independent of S, 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 has been introduced by Lechner (1995) and is especially useful for the specific problems encountered in this evaluation studies.

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 CTRT 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 the potential outcomes. Typical outcomes considered are gross monthly earnings, employment status, expected unemployment and expected changes in job positions in the next two years. Two concepts of timing are used for these outcomes, specifying 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 will identify reasons for participation in CTRT by supposing that individuals maximise future utility, or more precisely, the difference between the present values of future earnings streams for both states. It seems plausible that at least factors influencing both earnings and participation in CTRT can be identified in this fashion. It is not necessary to develop any formal behavioural model in any detail. Considering the broad building blocks of such a model is sufficient to identify potentially important attributes.[12] In principle one would

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

like to condition directly on these expected earnings (utility) streams, but since they are unobserved, they have to be decomposed into the cost of CTRT and the additional returns of CTRT. These factors have to be uncovered, because they are potentially important determinants of the training decision.[13] Although approval of the labour office is always necessary for CTRT participation, it is important to note that the decision for CTRT participation has nevertheless two dimensions: (i) the individual may push the labour office to allow him to participate in the subsidised CTRT (getting this approval was easy until 1993), or (ii) the labour office may push unemployed or STW individuals to participate in CTRT by threatening to reduce unemployment or STW benefits. Therefore, I will start with discussing reason for (i) from the point of view of the individual, and then for (ii) from the point of view of the labour office.

There are at least two hypotheses why earnings with CTRT 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, CTRT can act as a signalling device for an employer who has incomplete information on the worker's productivity. Participation in CTRT might signal in particular higher motivation, and the successful completion of longer CTRT courses may also signal higher ability (or reverse, if there is stigma associated with CTRT!), 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. Therefore, <u>age</u> should not increase the participation probability, but should most likely decrease it. The magnitude of the effect of age under the signalling hypotheses depends crucially on the ability of the employer to learn quickly the true productivity of the worker, because 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 earnings as well as employment prospects until retirement. This implies again a negative impact of age on CTRT participation (in case of a negative expected signal CTRT participation will not occur). Another factor is how the individual subjectively estimates the own future earnings 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 <u>subjective expectations about the own labour market prospects</u> the individual holds. This information is available on a yearly basis on the GSOEP.

It is useful to divide the potential costs of CTRT for the individual in two broad groups: direct costs and indirect or opportunity costs. Almost all the potential direct costs are beared by the labour office. The labour office tends to give subsidies to individuals with low nontraining

[13] Note that for these considerations, it does not matter how the labour market really works, but how the individual (and/or the labour office) believes it to work when deciding to participate in CTRT. 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.

labour market prospects, as estimated by the labour office and high CTRT prospects. Opportunity costs basically consist of lost earnings and / or leisure. Since the marginal utility of leisure should be lower during non-full-time work (a larger amount is available), the actual labour market status can be an important factor of 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 above analysis has identified age, expected labour market prospects, actual employment 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 that 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.

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 any long term plans that connect past employment behaviour to CTRT participation. It was generally impossible for East German workers to predict the impact and timing of the system change. Even when it was partly correctly foreseen, it was generally impossible to adjust behaviour 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 that is related to the condition of the labour market in the rapidly contracting East German post-unification economy. Figure 1 shows that the labour market is characterised 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 (this may not even be necessary before 1993, given the official guidelines for obtaining assistance from the labour office). This assumption allows me to consider monthly pre-training information on full-time employment, involuntary short-time work and unemployment, etc. as attributes.

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 labour 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, 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 1993.

Having discussed potentially important factors and variables available for the empirical analysis, the question is whether some important groups 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 organisations 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 high 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 earnings 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 maximisation problem such as a limited supply of CTRT. Supply information is available, however it is aggregated either within states (6) or in four groups defined by the number of inhabitants of cities and villages. 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 analysing 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 (1978) and Ashenfelter and Card (1985) observe a decline in earnings prior to training, but it has been shown in the previous section that, there is no evidence of this phenomenon in the sample used here.

4.2.2 Econometric considerations and first results for CTRT participation

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 for example - that are related to the distance (months or years) to the beginning of CTRT. Since these dates differ across CTRT participants, they are not clearly defined for the control group. An approximation, which might look 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 CTRT participants leads to a dependence of this variable on CTRT participation, the dependent variable. This dependence is aggravated by the rapidly changing labour market conditions. Therefore, such a variable cannot be considered exogenous. 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 CTRT 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).

In the following I estimate a 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:

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

$N(0,1)$ 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 mutual independence of *f(M,U)* and $V\beta^0$. These assumptions are tested with specification tests. See Appendix C for details.

Although all detailed results of the probit estimation and the specification tests are presented in Appendix C, let me now briefly sketch the results. Living in East Berlin is a significantly positive factor for CTRT participation compared to the other federal states. 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, because of the closeness of the already existing CTRT supply and the functioning labour office bureaucracy in West Berlin. Furthermore, the skill composition of the population differs from the rest of the country, because East Berlin was the capital and the administrative centre of the former GDR, although this effect should be largely - but perhaps not totally - captured by the schooling variables. The result for these variables suggest that higher schooling is associated with higher CTRT participation.

Women are more likely to participate in CTRT, which is not surprising because women experience far more unemployment than men during the post-unification period. However, this effect is lower for women in high job positions and women with high tenure.

Individuals who already know in 1990 that they will lose there their job, as well as those who expect redundancies in the firm are also more likely to be observed participating in CTRT. Additionally, there appears to be significant heterogeneity across different professions / occupations and industrial sectors. Finally, the negative coefficient of the variable that measures that the individual expects a decline in the professional career appears to be counter-intuitive, thereas the negative coefficient for individuals for whose income is very important for their subjective well-being may be due to a lower unemployment probability for these well motivated individuals.

Many other variables have been considered as well, but none of them appeared to be missing in the partial propensity score (see Tables A.1 and C.1 for details). The specification test do not provide any evidence against the chosen specification.

4.3 Nonparametric estimation of causal effects and matching

This section summarises the nonparametric methods used to estimate the causal effects of CTRT as discussed by Lechner (1995). The reader is referred to that paper for more details on the estimation methods.

The considerations in the previous sections suggest to estimate the causal effects by nonparametric methods in order to avoid potential inconsistencies due to misspecification. To ease notation assume that observations in the sample are ordered such that the first N^t observations receive CTRT, and the remaining (N-N^t) observations do not. The following nonparametric regression estimator is an obvious choice:

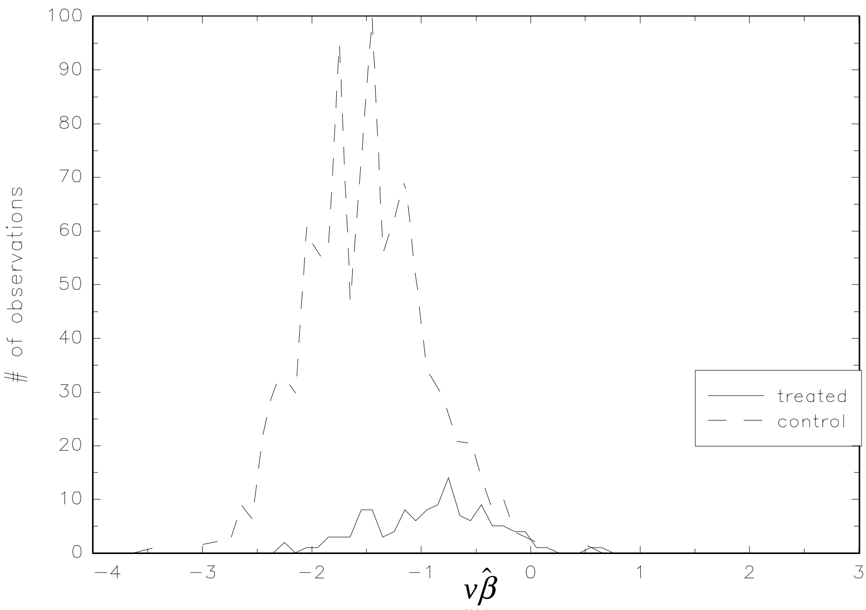
$$\hat{\theta}_N = \hat{E}(Y^t - Y^c | S=1) = \frac{1}{N^t}\sum_{n=1}^{N^t} y_n - \frac{1}{N^t}\sum_{n=1}^{N^t} \hat{g}^c[b(x_n)], \tag{6}$$

$\hat{\theta}_N$ denotes the estimate of the causal effects that is averaged over the sample of the N^t-treated observations only. $\frac{1}{N^t}\sum_{n=1}^{N^t} \hat{g}^c[b(x_n)]$ denotes a consistent estimate (N^t increasing) of $E(Y^c | S=1)$. This will be satisfied under standard conditions, if $\hat{g}^c[b(x_n)]$ is asymptotically unbiased (N-N^t increasing) for $E[Y^c | S=0, b(X)=b(x_n)]$. Nonparametric regression could be used to provide such an estimate. However, it will be subsequently shown that the balancing score most useful in this particular evaluation study has necessarily a high dimension. Given the size of the available sample, nonparametric regressions are subject to the typical *curse* of dimensionallity.

For these reasons I use a simpler nonparametric approach that appeared in the statistic literature (e.g. Rosenbaum and Rubin, 1983, 1985a). The idea is to find for every treated observation a single 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 avergae causal effects is unbiased.[14] In cases of 'mismatches', it is often plausible to assume that using local regressions on these differences will remove the bias. Appendix B gives the exact matching protocol as well as the mismatch adjustment procedures. Appendix D contains information about the exact empirical implementation of the algorithm.

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 6 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, there is overlap for all of the treated distribution.

Figure 6: Distribution of $v\hat{\beta}$ for CTRT and controls



Note: 0.1 grid used. Mean (*std*) in CTRT (treated) sample / control sample is -0.90 (0.55) / -1.50 (0.55).

Table 3 presents the marginal means in various control groups and the CTRT group for several important variables. Additional descriptive statistics are given in Table D.1 and Figure

[14] Compared to the nonparametric regression described above, there is an asymptotic efficiency loss, because observation n ($n \leq N^t$) and its closest neighbour in the control population - instead of possibly many close neighbours - are used to compute $\hat{g}^c[b(x_n)]$.

D.1 in Appendix D. A comparison of column (2), based on randomly matching controls to CTRT observations, with column (5) exhibits the expected fact that the increased pre-training unemployment rates are not observed in the control population as a whole. Even when the matching is on the partial propensity score (col. 3), unemployment and full-time work rates differ dramatically. This shows clearly that matching on the partial propensity score is insufficient and that the monthly information has to be taken into account (col. 4). The entries in Table 4 show that the proposed matching algorithm removes almost all of the differences in the employment status variables in the month / year before CTRT. Indeed, the figures in the following section will show that over the whole pre-CTRT period the CTRT observations and controls based on the matching algorithm underlying column (4) do not differ significantly. However, the exception to the rule is the variable *high job position* and variables correlated with it, such as schooling (see Table D.1). This points to a control pool that is not quite rich enough to allow a perfect match, because there appears to be a insufficient number of highly qualified unemployed individuals (given the other characteristics). Therefore, in the following section, an econometric adjustment mechanism is proposed to control for these kinds of mismatches.

Table 3: Descriptive statistics of selected variables of CTRT and control sample: different matching algorithms (103 observations)

	Controls			CTRT
	random	matched on $v\hat{\beta}$ and selected v-variables	and m-var.	
(1)	(2)	(3)	(4)	(5)
Variable	mean (std), share in %	mean (std), share in %	mean (std), share in %	mean (std), share in %
Unemployment: month before CTRT	6	7	44	52
STW: month before CTRT	6	5	20	21
Full time employment: month before CTRT	88	88	32	24
Unemployed (yearly before CTRT)	5	5	15	18
Full time employed (yearly)	85	83	65	65
Gross earnings in 1993 DM (yearly, 0 if not working)	1873	1747	1363 (715)	1368 (784)
Job position: highly qualified, management (yearly)	12	23	13	23
Expectations for the next 2 years or unemployed (yearly)				
losing the job: certainly, possibly	47	55	72	75
decline in professional career: certainly, possibly	23	29	44	43

Note: (2) random controls; (3) matched on $v\hat{\beta}$ and selected v-variables; (4) matched on $v\hat{\beta}$ (103), selected v-variables and m (monthly, yearly)-variables; v-variables used for the additional conditioning are: *gender*, *university*, *12 and 8 years of schooling*, *women in highly qualified or management job positions* (1990), m-variables are: *expectation of losing own job in the next two years (yearly), expectation of a declining career in the next two years (yearly), monthly wage / salary (yearly), training (unspecified, yearly), self-employment (yearly), highly qualified or management job positions (yearly), unemployment (monthly), STW (monthly), full-time work (monthly);* see also note to Table A.1 for the exact definition of variables; see Appendices B.1 and D for details on the matching algorithms used.

A comparison of the differences in yearly and monthly measured variables such as unemployment is worth a remark. The difference between these variables and the drastic increase of unemployment for example in the months before CTRT (see Figure 3) emphasizes the point that the monthly calendar constitutes a very valuable information.

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 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 CTRT on post-training changes in actual and anticipated labour 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 CTRT can possibly be discovered.

The following 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 labour market status - measured once a year - is gross monthly earnings. For those being employed, it is defined as the gross monthly earnings in the month before the interview. For those not being employed, either imputed unemployment benefits or social assistance - whichever is higher - or zeros are used instead (see Appendix A for details). Labour 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.[15] Additionally, there is information whether individuals expect an improvement or a worsening of the current job (career) position.[16] It is important to note for the discussion in the following subsection that, except for the earnings variable, all outcome variables are coded as binary indicators.

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

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

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 or a specific time span (months or intervals of 0-1, 1-2, 2-3 years for yearly information) after the completion of CTRT. Note that the number of observations available for the evaluations decreases with the length of the time span considered.

4.4.2 Econometric issues

Define the differences in matched pairs in the sample, which consists of independently drawn observations, as $\Delta y_n = y_n^t - y_j^c$, $\Delta b(x_n) = b(x_n^t) - b(x_j^c)$, $n = 1, \ldots, N^t$, where y_j^c and x_j^c denote values of an observation from the pool of individuals not participating in CTRT (controls) that is matched to the treated (CTRT) observation *n*. The estimate of the average causal effect and the respective standard error are computed as:

$$\hat{\theta}_{N^t} = \frac{1}{N^t} \sum_{n=1}^{N^t} \Delta y_n, \qquad Var(\hat{\theta}_{N^t}) = \frac{1}{N^t} (S_{y^t}^2 + S_{y^c}^2). \tag{7}$$

$S_{y^t}^2$ and $S_{y^c}^2$ denote the square of the empirical deviation of *Y* in the CTRT sample and in the sample matched to the CTRT-sample, respectively.[17] As mentioned in the previous section, when a perfect match is achieved, implying that $\Delta b(x_n) = 0$, $n = 1, \ldots, N^t$, 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 (7) gives the principle nonparametric estimate of the causal effect to be refined in the following in order to take account of time before and after CTRT. Denote by N_τ^t, $\tau \in \{\ldots, -3, -2, -1, 1, 2, 3, \ldots\}$ the number of pairs observed at any distance to CTRT. Let $\iota_\tau(n)$ = 1 if observation n is observed at distance τ. The observability of an observation in a particular distance depends only on the ending dates of CTRT (see Appendix A). In the following I will assume that they are independent random variables.[18] Therefore, the refined estimators based on the distance as opposed to the date concept of time are defined as:

$$\hat{\theta}_{N_\tau^t} = \frac{1}{N_\tau^t} \sum_{n=1}^{N^t} \iota_\tau(n) \Delta y_{n,\tau}, \qquad \tau \in \{\ldots, -3, -2, -1, 1, 2, 3, \ldots\}; \tag{8}$$

---

17 The variance estimate exploits the fact that the matching algorithm given in Appendix C.1 never chooses an observation twice.

18 Two checks are performed with respect to this assumption. First of all, the end dates (months) are regressed on (1,p(v),p(v)^2,p(v)^3,m). None of the variables, except the constant, is significant, and the adjusted R^2 is 0.1 (N_{-1}^t=103). Secondly, the sample is split according to different end dates, but the qualitative results do not change at all. Therefore, there is no evidence from the data that the independence assumption (typically used in unbalanced panels) is suspect.

$$\hat{\theta}^{T}_{N^t_T} = \frac{1}{N^t_T}\sum_{n=1}^{N^t}\sum_{\tau=1}^{T} \iota_\tau(n)\Delta y_{n,\tau}\,, \qquad\qquad T \in \{1,2,3,\ldots\}\,. \tag{9}$$

The variances are computed appropriately. When τ is negative, then $\hat{\theta}_{N^t_\tau}$ denotes the mismatch in period τ before CTRT, otherwise it denotes the effect of training in period τ after CTRT. $\hat{\theta}^{T}_{N^t_T}$ indicates the accumulated effect T periods after CTRT. These effects are also computed for subpopulations defined by attributes or training characteristics. No assumption is necessary regarding whether or not the treatment effects may differ across the population. It should also be remarked that whenever regression-type adjustments are used for different dates (time spans) for the same outcome variable (see Appendix B for details), 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, 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 results of the evaluations are given in following Figures 7 to 10, in Table 4 as well as in Appendix E. Using eq. (8) for the figures and eq. (9) for Table 4 to estimate the causal effects of CTRT, they show the differences between the control and the CTRT 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).[19] For variables measured by the monthly calendar the distance is expressed in months, for those measured only for the particular month of the yearly interview, the distance is expressed in years.[20] The figures cover up to 18 months or up to 3 'years' before the training and up to 27 months or 3 'years' after CTRT. They display the mean effect (solid line; + for the mismatch corrected estimate) and its 95% pointwise confidence interval based on the normal approximation (dashed line; ∇, Δ for the mismatch corrected estimates). The number of observations available to compute the respective statistics decrease the longer the distance to the incidence of CTRT is (see Table 4 for the remaining number of observations). The implications of this are that the variance increases. This is reflected in the widening of the confidence intervals. However, the accuracy of the estimated intervals itself may deteriorate, because the normal distribution may be not a very good approximation of the sample distribution when the sample gets too small. Additionally, a mismatch correction may be impossible or very imprecise, because there may be too few observations to identify and

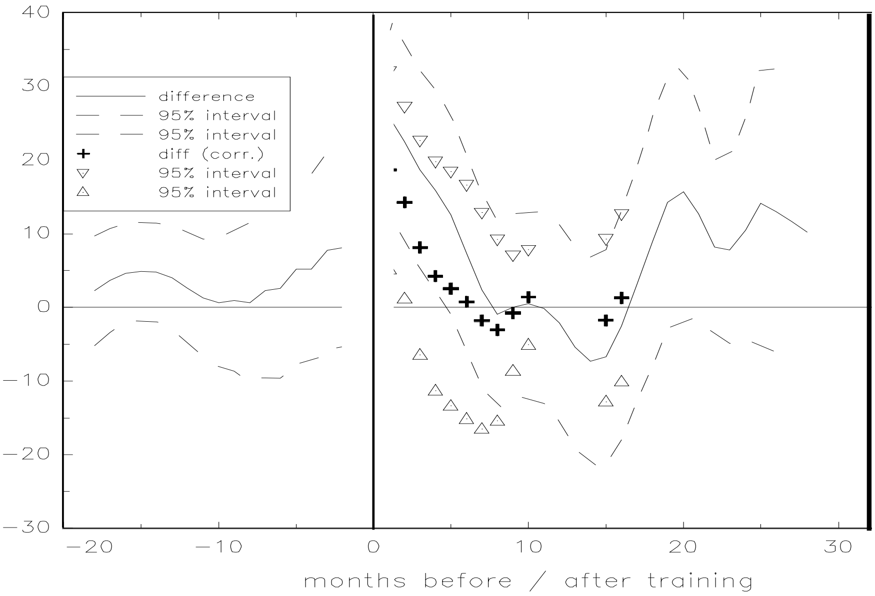
[19] The results for those outcomes that are mentioned in Section 4.4.1, but do not appear here, are not qualitatively different from the ones presented. Therefore, they are either relegated to Appendix E or omitted for the sake of brevity.

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

estimate the parameters of the ordered probit model.[21] Hence, the results on the very right side of tables have be interpreted with care.

Figures 7 and 8 present the result of the evaluations for the monthly outcome variables unemployment and full-time employment.[22] The part left to the 0 vertical mark allows a judgement about the quality of the matches concerning the particular variable.[23]

Figure 7: Difference of unemployment rates in %-points



Note: N^{t}_{-1} = 103. Smoothed using 3 month moving averages for $|\tau| > 1$.

As already noted in the discussion of match quality, there is small excess unemployment just prior to the beginning of the course, that is however not at all significantly different from zero. Figure 7 shows that the immediate effect of CTRT is additional unemployment in the months following the end of CTRT. After a few months these effects disappear. Indeed Table 4 shows the total effects after 9 months cannot be distinguished from zero. At first sight this seems surprising, because Table 2 shows that the unemployment rate of CTRT participants is indeed falling rapidly during the first 12 months after CTRT.

However, there seems to be a simple explanation for this effect. Remember that more than 50% of CTRT participants are unemployed before CTRT. For an unemployed person the immediate effect of (full-time) CTRT is that during CTRT his or her search efforts will be reduced (mean duration is 12 months!) compared to the controls. The results suggest that if

21 All computations based on less than 5 observations are suppressed.

22 *Unemployment* here indicates that the individual has registered for unemployment. There is another monthly indicating the receipt of unemployment benefits ("Geld" or "Hilfe"). The results are almost exactly the same when using this second measurement of unemployment.

23 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-CTRT outcomes here are denoted as unaffected outcomes in his terminology.

there is a positive effect of CTRT it is not large enough to compensate for this initial negative outcome.

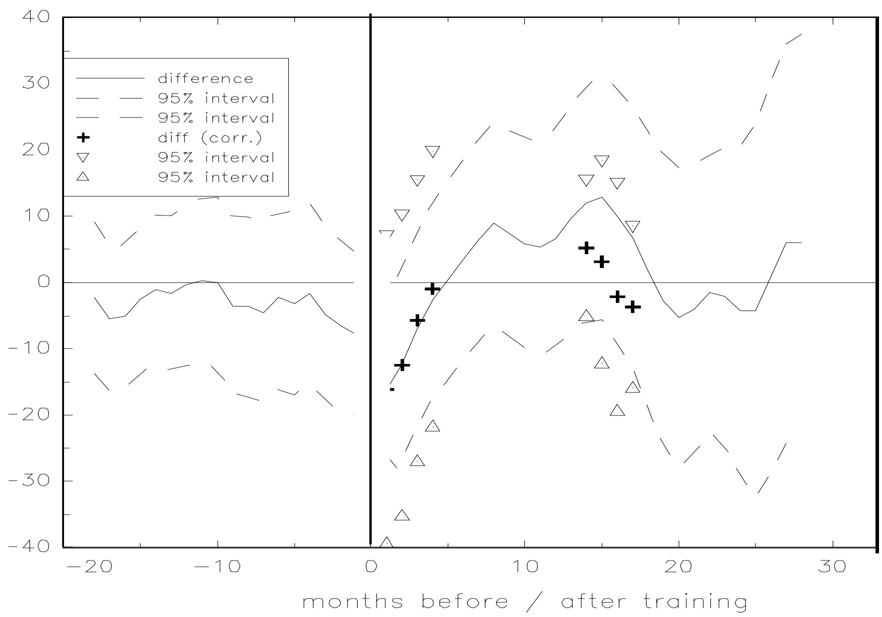
Table 4: Average effects in terms of additional months in particular employment state

Months after CTRT (τ)	unemployment		unemployment or STW		full-time employment		N_τ^t
	mean (std.)	corrected	mean (std.)	corrected	mean (std.)	corrected	
3	**0.7** (0.2)	0.4 (0.2)	**0.8** (0.2)	0.4 (0.3)	-0.4 (0.2)	-0.3 (0.3)	90
6	**0.9** (0.4)	0.5 (0.5)	**0.8** (0.4)	0.4 (0.6)	-0.3 (0.4)	-0.2 (0.6)	80
9	0.8 (0.5)	-0.6 (0.5)	-0.5 (0.6)	-0.3 (0.8)	0.1 (0.7)	0.6 (0.8)	67
12	0.6 (0.7)	0.3 (0.8)	-0.1 (0.8)	0.2 (1.2)	0.3 (0.9)	1.0 (1.3)	51
18	0.3 (1.0)	0.0 (1.5)	-0.8 (2.1)	-0.3 (1.6)	2.4 (1.5)	1.6 (2.0)	32
24	0.6 (1.2)	n.a.	1.6 (1.8)	n.a.	0.6 (2.4)	n.a.	16

Note: *corrected*: estimates are corrected for mismatch by using pairwise regression (see Appendix B); n.a.: not available; Standard errors in brackets; **Bold** letters: |t-value| larger than 1.96. The entries should be read as "CTRT caused on average a total of XX months of unemployment (for example) YY months after the end of CTRT".

These general findings are confirmed by considering either STW and unemployment together or by considering full-time employment the respective labour market outcome. Although for the later the result is less precise in a statistical sense; the significance of it at the 95% level depends on the method used. However, the mean shows exactly the same shape that is of course inverted compared to Figure 7. Considering only a sample of individuals who are either unemployed or on STW before CTRT sharpens these results (see Figures E.1 and E.2 of Appendix E).

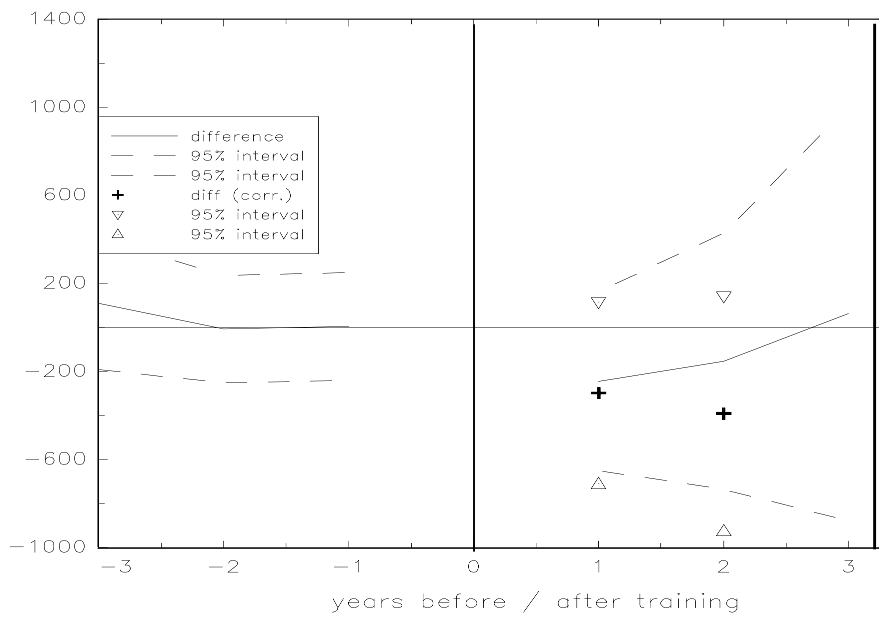
Figure 8: Difference of full time employment rates in %-points



Note: N_{-1}^t = 103. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figures 9 and 10 feature outcome variables that are only measured once a year, such as gross monthly earnings and expected improvement of the professional career in the next two years. There are no significant differences for the pre-training outcomes in both cases. For the earnings variable presented in Figure 9 there does not appear to be an effect of training either. This remains true when nonemployment earnings is coded differently (see Figure E.3 in Appendix E) or when other functional forms such as log´s are used. Note however, that the estimated earnings effects are mainly driven by the estimated full-employment probabilities. Estimation of causal effects conditional on full-employment is complicated, because post-training full-employed individuals are not a random sample from the treated / control population. Addressing the additional selection issues is beyond the scope of this paper. Another issue with respect to problems of estimating earnings effects has already been noted by Ashenfelter (1978): The returns-to-schooling literature typically suggests that one year of schooling has an effect of less than 10% of additional earnings capacity. Since the duration of training courses is typically much less than a year, and since earnings variable typically exhibit a large variance, it will be very difficult to detect the comparatively small earnings effects expected a priori. This view is confirmed by the large confidence intervals appearing in Figure 9.

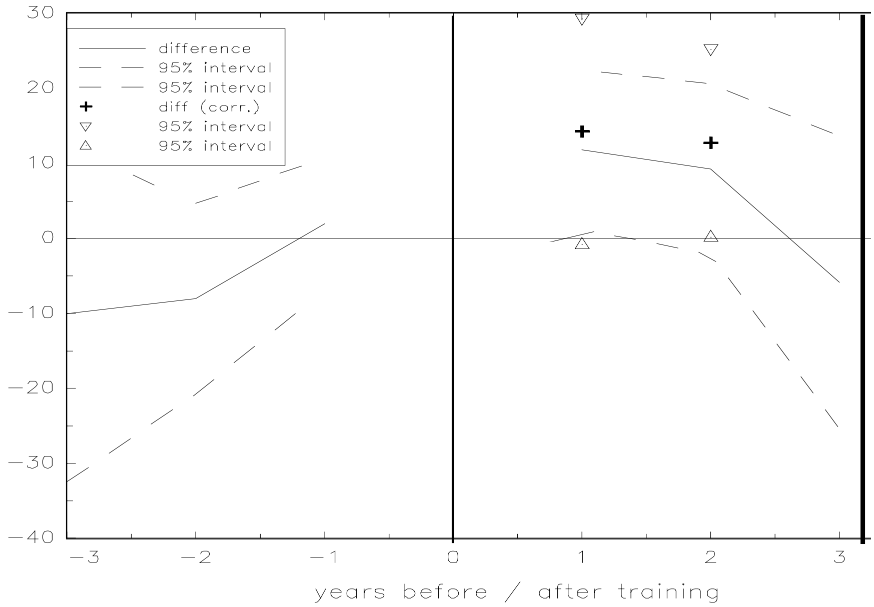
Figure 9: Difference of gross earnings (in 1993 DM)



Note: N_{-1}^{t} = 103. See Appendix A for details. 0 when not employed.

Figure 10 reveals a potential shortcoming of this study: Individuals do think that CTRT will improve their career prospective in the next two years. Since they CTRT participants expect to improve their situation even in the two years after year two, and since they have already made up the initial loss during CTRT, it might be that they will overtake the controls outside the sample period. Unfortunately, from the data at hand it is impossible to decide whether this variable really contains information about future realisation of labour market outcome, or whether this is just *wishful thinking* of CTRT participants.

Figure 10: Differences in expected improvements in the professional career in the next two years in %-points



Note: N^{t}_{-1} = 103. Nonemployment coded as not expecting improvement.

These results are in contrast to more positive results obtained in a recent study by Fitzenberger and Prey (1995). However, they use a different data set and model the joint stochastic processes of selection, panel attrition and outcomes using joint normality. With a similar data and a similar estimation and identifying strategy than this paper, Lechner (1995a) arrives at comparable results in finding no clear-cut positive effects of off-the job training. However, both studies analyse different kinds of training, for example by including training spells not funded by the BA. Furthermore, the descriptive statistics for the kind of training analysed in these papers clearly indicate that the results are very difficult to compare to the results of this paper that exclusively focuses on longer training spells funded by the BA.

4.4.4 Sensitivity

Additionally to the already mentioned use of different functional forms as well as a different way of computing non-employment earnings, the sensitivities of the results are checked in several other directions.

The perspective of time is changed: instead of considering the distance from a point in calendar time to the beginning or end of CTRT, the pre- and post CTRT outcome are compared and averaged for the same months / years in calendar time. Two examples for these sorts of results that do not lead to different conclusions are given in Figures E.6 and E.7 of Appendix E.

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, age and pre-training employment status. No significant differences appear.

To check the results for sensitivity with respect to the definition of CTRT, the courses used in the estimation are split in several subsamples according to whether: (i) they began not earlier than January 1991 (N_{-1}^{t} = 88), (ii) they began not later than December 1991 (N_{-1}^{t} =57), (iii) they ended not later than June 1993 (N_{-1}^{t} = 85), (iv) they ended not later than December 1992 (N_{-1}^{t} = 59), (v) they have a minimum duration of two months (N_{-1}^{t} = 98), (vi) they have a minimum duration of six months (N_{-1}^{t} = 76), (vii) they have a minimum duration of two months and a maximum duration of 18 months (N_{-1}^{t} = 79), and (viii) there are no multiple spells (N_{-1}^{t} = 92). As a final sensitivity check I also considered a control and treatment group that did not participate in any other form of training (N_{-1}^{t} = 89). None of the subsamples reveals a substantial difference compared to the results presented above.

In conclusion the sensitivity analysis reveals a remarkable stability of the results.

5 Conclusion

The general findings of the paper suggest that there are no positive earnings and employment effects of public sector sponsored continuous vocational training and retraining (CTRT) in the short-run. Regarding the risk of unemployment there are indeed negative effects of CTRT directly after training ends. However, these negative effects are compensated over the first year after training. It is open question whether the lack of a positive effect is due to a bad signal (i.e. lack of skills) participants send to prospective employers, or whether it is due a lack of quality in a narrow sense. Nevertheless, the results in this paper suggest that CTRT after unification was very much a waste of resources, providing quantity without sufficient quality (or a sufficiently positive signal). The quality problem has been realised by the labour office, which subsequently tried to improve quality and changed the selection process to include a higher share of individuals previously unemployed in CTRT.

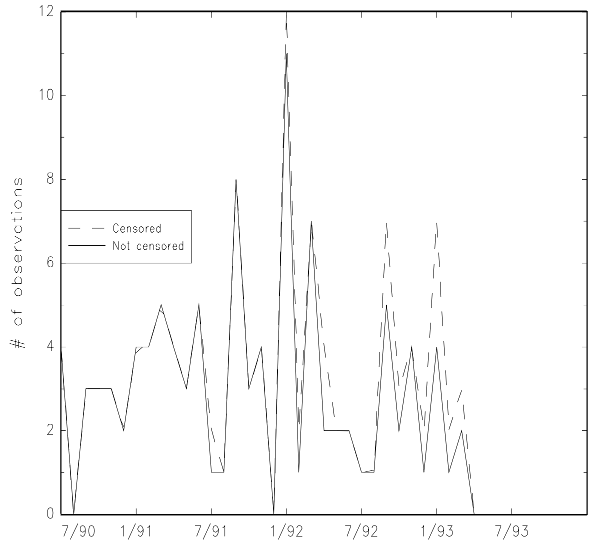
This negative picture may be an exaggeration of the situation for several reasons: Firstly, money spent for CTRT in the first two to three years may be seen as investments in the East German training infrastructure, that had to be build from scratch. In this sense future CTRT might still yield some returns on these early investments. Secondly, the massive use of CTRT achieved a significant reduction of the official unemployment rate. This was politically desired, and hence it might be seen as an achievement per se. Finally, there is some evidence that trainees expect positive returns over a longer time horizon. Since this horizon is beyond the sampling period available for this study, it is impossible to learn from the data whether these expectations are correct or not. If these expectations materialise, then future evaluations will find positive effects of training.

Although the data and the suggested nonparametric estimation strategy appeared to be well suited for the problem at hand, the small sample remains a problem. It is mainly reflected in comparatively large standard errors when considering medium or long term effects. Therefore, interesting future research should investigate these effects over a longer period. Additionally, one might investigate jointly the effects of different types of training, such as on-the-job training versus off-the-job training, or publicly funded versus privately funded training. Likewise, it will be an issue whether the quality of the publicly funded training did really improve after 1992, as claimed by official sources.

Appendix A: Data

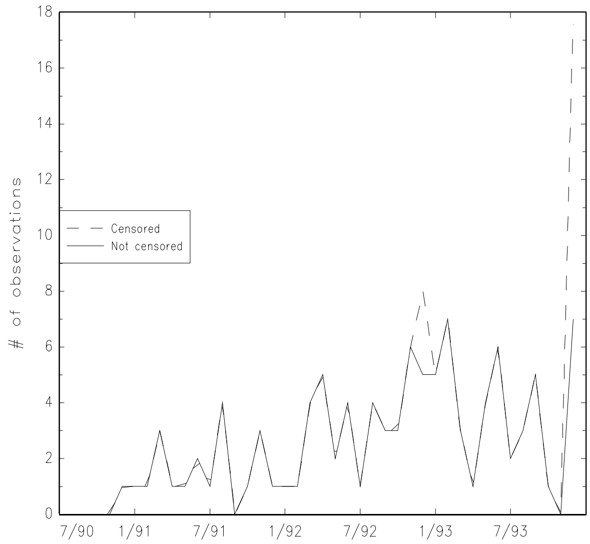
This appendix briefly explains the coding of the start, duration, and end date of CTRT courses. It also contains a histogram for the distribution of start dates and the ending dates in Figures A.1 and A.2 as well as the pdf and cdf of durations in Figure A.3 and A.4. Furthermore, the exact definition of the earnings variables used in the evaluations are given. Finally, Table A.1 and Figures A.5 and A.6 show descriptive statistics for all variables used in the estimation.

Figure A.1: Distribution of CTRT start dates



Note: Monthly information. *Censored* includes right censored spells.

Figure A.2: Distribution of CTRT end dates

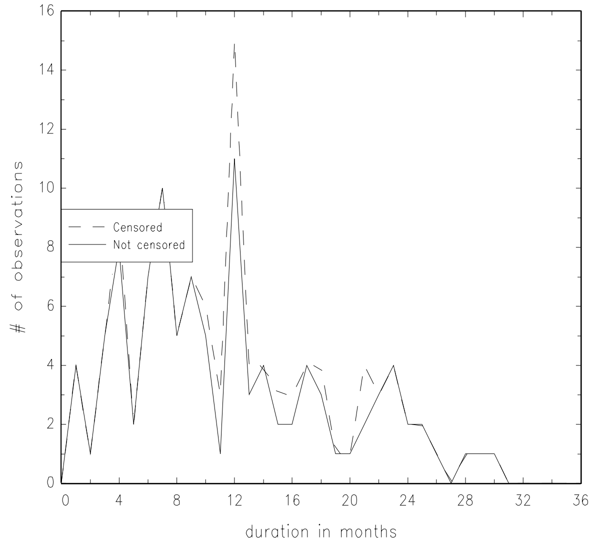


Note: Monthly information. *Censored* includes right censored spells. Last observed date is coded as end date for censored spells.

There are several cases with multiple spells of training (10%). In these cases single durations of the spells are added, and Figure A.1 shows the beginning date of the first spell, and Figure A.2 shows the end date of the last spell. *Censored* refers to the sample that includes spells not completed by Dec. 1993, respectively Dec. 1992 in cases when the data for 1993 was not available. Furthermore, a few spells are included that begin in early 1993 (Jan. 6 cases, Feb. 1 case, March 2 cases). It must be that the cases beginning in January 1993 have been approved in 1992 (and this seems to be a reasonable assumption for the 3 other cases as well). Hence, they are subject to the same rules as the other CTRT observations. Additionally, the first major tightening of rules in 1993 happened in early May (see BA 1993b).

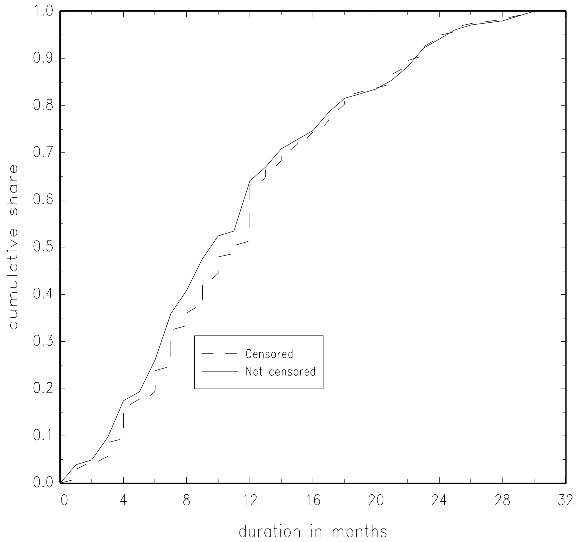
The sample given in the following table is the sample used for the estimation of the partial propensity score. For the CTRT sample this means that observations with uncomplete training spells are *not* deleted. In the control sample several observations have been deleted related to specific attributes for which no CTRT observation is observed. For those variables the sample mean in the control sample before deleting them is given in brackets.

Figure A.3: Distribution of CTRT durations



Note: Monthly information. *Censored* includes right censored spells. The mean (std.) of the durations is 11.7 (7.2) for the censored sample, and 11.9 (6.9) for the uncensored sample.

Figure A.4: Empirical distribution functions for durations of training



Note: Monthly information. *Censored* includes right censored spells.

Gross monthly earnings 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 only for the following years. For those unemployed unemployment benefits are computed using 67% of the last *gross* earnings, 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.[24] After performing these imputations, it is ensured that earnings levels are not below average social assistance levels (Bundesministerium für Arbeit und Sozialordnung, 1994, Table 8.16A). Finally, all earnings 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). A second measure of earnings is computed by stipulating zero earnings for those unemployed. This second measure of earnings is motivated by the idea that earnings are in some sense a measure of individual productivity.

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

Table A.1: Descriptive statistics

	CTRT (125 obs.)		No CTRT (1038 obs.)	
Variable	mean / share in %	std.	mean / share in %	std.
Age in 1990	33.4	7.5	35.1	8.1
Gender: female	56.8		39.7	
Marital status in 1990				
married	71.2		76.7	
single	22.4		16.8	
divorced, separated	6.4		7.2	
Federal states (Länder) in 1990				
Berlin	10.4		6.9	
Brandenburg	16.0		14.1	
Mecklenburg-Vorpommern	8.8		10.8	
Sachsen	35.2		31.4	
Sachsen-Anhalt	14.4		19.7	
Thüringen	15.2		17.8	
Size of city / village				
< 2000	22.4		27.0	
2000 - 20000	36.0		28.4	
20000 - 10000	23.2		22.9	
> 100000	18.4		21.8	
Supply of CTRT				
hours per 100 inhabitants of community	0.76	.12	0.77	.13
Länder subsidies to providers per employed person	127	24	122	24
places per 100 employed persons (state)	0.93	.29	0.89	.24
state expend. for active labour market pol. per inh.	0.95	.31	0.97	.34
Years of schooling (highest degree) in 1990				
12	29.6		18.3	
10	60.8		60.2	
8 or no degree	9.6		21.5	
Highest professional degree in 1990				
university [1)]	19.2		12.7	
engineering, technical college[2)]	16.8		15.1	
master of a trade / craft	4.8		6.0	
skilled worker[3)]	58.4		65.1	
no degree	0.8		1.0	
Job position in 1990				
highly qualified, management	24.0		21.8	
master of a trade / craft[4)]	4.8		6.9	
skilled blue and white collar[5)]	56.8		55.2	
Job characteristics in 1990				
wage / salary per month in 1993 DM	1625	487	1682	494
tenure in years	8.7	6.9	10.2	8.1
temporary job contract	4.0		3.7	
professional degree in other than current profess.	34.4		35.3	
already fired	11.2		4.1	
training (unspecified) while full-time employed	8.0		7.1	
Memberships in 1990				
union	79.2		74.4	
professional association	9.6		6.4	
cooperative (LPG / PGH)	4.8		8.8	

Table A.1 to be continued ...

Table A.1: Descriptive statistics: continued

Variable	CTRT (125 obs.) mean / share in %	No CTRT (1038) mean / share in %
Profession in 1990 (ISCO)		
scientific, technical, medical	24.8	19.1
production	26.4	40.8
managerial	2.4	2.6
administrative	8.0	10.2
trade	6.4	3.5
agriculture	5.6	5.4
services	4.8	7.1
services, incl. trade, administrative	21.6	23.5
Employer characteristics in 1990		
self-employed	-	(2.0)
firm size (number of employees)		
0-19	5.6	10.2
20-199	26.4	28.5
200-1999	40.0	36.2
2000 and more	27.2	24.0
industrial sector		
agriculture	14.4	13.3
energy and water	-	(2.6)
mining	-	(3.1)
heavy industry [i)]	11.2	10.1
light ind., consumer goods, electronics, printing	29.6	18.5
machine building and vehicle construction	8.8	6.5
construction	5.6	8.2
trade	8.8	8.1
communication, transport	4.0	8.9
education, science	10.4	12.4
health	-	(7.8)
other services [ii)]	7.2	13.4
redundencies announced	65.6	48.6
Finding a similar new job is (in 1990)		
impossible	16.8	10.8
difficult	72.0	70.1
easy	11.2	18.2
Very worried about job security in 1990	49.6	38.1
Optimistic about the future in general in 1990	12.8	17.0
Not at all optimistic about the future in general in 1990	12.8	7.5
Very confused by new circumstances	4.8	4.3
Voluntary services in social organizations in 1990:	34.4	40.3
Not enjoying work	6.4	5.4
Very desirable behavior / attitudes in society in 1990		
performing own duties	59.2	67.8
achievements at work	70.4	73.2
increasing own wealth	14.4	28.0
Income very important for subjective well-being	56.0	63.7
Work very important for subjective well-being	48.0	54.5

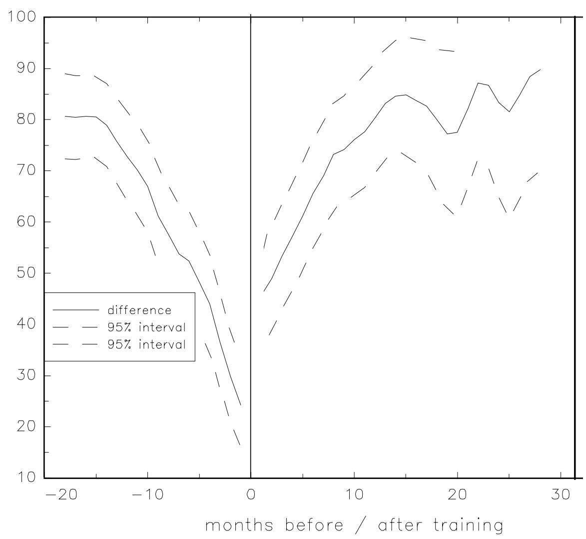
Table A.1 to be continued ...

Table A.1: Descriptive statistics: continued

Variable	CTRT (125 obs.) mean / share in %	No CTRT (1038) mean / share in %
Expectations for the next 2 years in 1990		
redundancies in firm: certainly	50.4	33.4
redundancies in firm: certainly not	3.2	6.5
losing the job: certainly	12.8	5.8
losing the job: possibly	47.2	36.5
losing the job: certainly not	6.4	10.1
improvements in professional career: certainly	-	(2.3)
improvements in professional career: certainly not	45.6	42.8
decline in professional career: certainly	2.4	3.0
decline in professional career: certainly not	43.2	47.5
new profession: certainly	6.4	4.5
new profession: certainly not	32.0	44.7

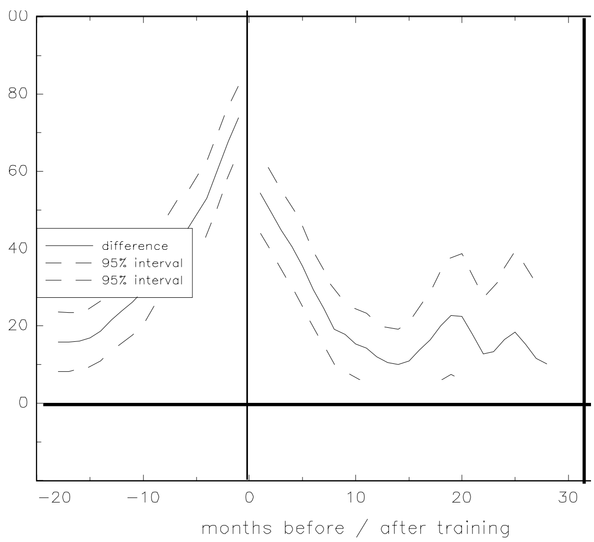
Note: *1)* University and 'Fachhochschule'; *2)* 'Ingenier- und Fachschule', not 1); *3)* 'Berufsausbildung', 'Facharbeiter', 'sonstige Ausbildung', not 1), 2) or master of a trade / craft; *4)* Includes 'Brigadier', 'Meister im Angestelltenverhältnis'; *5)* 'Facharbeiter', 'Angestellte mit qualifizierter Tätigkeit'; i) plastics, chemicals, stone, clay, glass, steel, ii) incl. nonprofit, banks, insurance, goverment, legal, personal services, cleaning, waste disposal, hotels, restaurants; *a)* Müller (1994), Figure 3; *b)* Müller (1994), Tables 2, 3; *c)* Müller (1994), Table 7; *d)* Buttler (1994), Table 4; *1990* relates to the date of interview (earlier than July 1990 for almost everyone).

Figure A.5: Share of full-time employed CTRT participants in %



Note: Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure A.6: Share of unemployed or STW CTRT participants in %



Note: Smoothed using 3 month moving averages for $|\tau| > 1$.

Appendix B: Econometrics

B.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 CTRT (T-pool) or not (C-pool).

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

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

Step 4: Find observations in C-pool (denoted by *j*) obeying $v_j\hat{\beta} \in [v_n\hat{\beta} \pm c\sqrt{Var(v_n\hat{\beta})}]$. The constant *c* is chosen such that the interval is identical to a 95% confidence interval around $v_n\hat{\beta}$.

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 *n*, such that it minimises $(v_j\hat{\beta} - v_n\hat{\beta})^2$.

(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*. Denote these and perhaps other variables already included in *V* as $\tilde{m}_j$ and $\tilde{m}_n$, respectively. Define a distance between each control *j* and *n* as $d(j,n) = (v_j\hat{\beta}, \tilde{m}_j)' - (v_n\hat{\beta}, \tilde{m}_n)'$. Choose control *j* such that it has the smallest Mahalanobis distance $m(j,n) = d(j,n)'Wd(j,n)$ within the calliper. *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 close to the one proposed by Rosenbaum and Rubin (1985a) 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 calliper-width for all observations, the widths are allowed to vary individually with the precision of the estimates of a monotone function of the partial propensity score (step 4). The widths are large, because matching is not only on the partial propensity score and its components, but also on additional variables. The unbounded linear index $v_n\hat{\beta}$ is used instead of the bounded partial propensity score $\Phi(v_n\hat{\beta})$. Matching on the latter with this kind of symmetric metric leads to an undesirable asymmetry when $\Phi(v_n\hat{\beta})$ is close to 0 and 1, depending on which side the control *j* is.

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 the same properties as conditioning on the (unknown and not estimable) propensity score itself.

B.2 Correction for mismatches

This appendix briefly gives the method used to correct for any mismatch remaining after using the algorithm described in section B.1 of this appendix. Denote the difference of the potential outcomes by $\Delta Y = (Y^t - Y^c)$. The realisations of the sample and the matching process gives us pairs $(\Delta y_n, \Delta x_n)$, $n \leq N^t$. Define the difference in terms balancing scores as $\Delta b(x_n) = b(x_n) - b(x_j)$. x_j denotes the value of *x* for observation *j* that is matched to observation *n*. Note that in general, equation (B.1) holds:

$$E[\Delta y_n | \Delta b(x_n) = 0] = E\{E[\Delta Y | b(X) = b(x_n)] | S = 1\} = \theta^0 . \qquad \text{(B.1)}$$

However, $\Delta b(x_n)$ may not be exactly zero. In this case the exact type of the suggested correction will depend on whether the outcome variables are continuous or discrete. In the case of continuous variables it is reasonable to assume that the conditional expectation of the dependent variable is linear in $\Delta b(x_n)$, because matching has already removed almost (if N is finite) all differences in the balancing scores, so that the $\Delta b(x_n)$ are local deviations:

$$E[\Delta y_n | \Delta b(x_n) = \eta_n] = \theta^0 + \eta_n \lambda, \qquad n \leq N^t. \qquad \text{(B.2)}$$

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 by regressing the differences in the balancing score and a constant on the differences in outcomes (cf. Rubin, 1979).[25]

Suppose now that the outcome consists of only two values, say 0 and 1, hence the support of ΔY is the set {-1,0,1}. In this case, the treatment effect can be written as:

$$\theta^0 = E(\Delta Y | S = 1) = P(\Delta Y = 1 | S = 1) - P(\Delta Y = -1 | S = 1) \qquad \text{(B.3)}$$

A consistent estimate of the average treatment effect can be obtained by substituting sample analogues for the population probabilities:

$$\hat{\theta}_{N^t} = \frac{1}{N^t} \sum_{n=1}^{N^t} \{P[\Delta y_n = 1 | \Delta b(x_n) = 0] - P[\Delta y_n = -1 | \Delta b(x_n) = 0]\} \qquad \text{(B.4)}$$

[25] In the empirical evaluations 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.

Using a linear approximation for these differences of probabilities is not so attractive as before, except when $\Delta b(x_n)$ is very small. Therefore, I choose a more parsimonious specification. In a first step a three-group-ordered probit model is estimated with Δy_n as dependent variable and $\Delta b(x_n)$ plus a constant as independent variables.[26] The asymptotic covariance matrix for the estimated coefficients of the ordered probit model are computed using the combination of OPG and expected hessian. In the second step, the above probabilities are directly derived from this model and computed for the individual observations using the estimated coefficients of the ordered probit model. Finally, the variance of $\hat{\theta}_{N^t}$ 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 is asymptotically unimportant as long as the differences in attributes Δx_n disappear.

The same approaches as for the mismatch corrections 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 CTRT. 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. 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 continuous characteristics. However, when the attributes and characteristics have too many different values some modelling is required given the size of the sample used in this study. For more details see Lechner (1995).

Appendix C: Estimation results for the partial propensity score

Table C.1 in this appendix presents the results of the probit estimation of the partial propensity score as well as the results of the specification tests.[27] All variables that are not contained in Table C.1, but described in Table A.1, as well as different functional forms for the 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 5% level. Most results are above the 10% level.

Conditional homoscedasticity (implied by independence) and normality are 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

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

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

Lechner (1993) and in Lechner (1995b).[28] 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)*.

Table C.1: Results of the estimation and the specification tests for the participation probit

	estimation		heteroscedasticity test	
Variable	coef.	std.err.	$\chi^2(1)$	p.-val.
Gender: female	**0.83**	0.18	0.2	69
Federal states (Länder) in 1990: Berlin	**0.43**	0.21	0.0	90
Years of schooling (highest degree) in 1990				
12	**0.57**	0.25	1.4	24
10	**0.36**	0.17	1.5	23
Highest professional degree in 1990: university	0.24	0.25	0.0	90
Job position in 1990				
highly qualified, management	-0.02	0.24	0.0	98
highly qualified, management and female	**-0.54**	0.27	0.1	82
Job characteristics in 1990				
women´s tenure (in years)	**-0.023**	0.011	0.1	83
already fired	**0.60**	0.23	0.1	83
Profession in 1990 (ISCO)				
production	-0.23	0.17	0.0	95
services, incl. trade, office	**-0.38**	0.16	0.4	54
trade	**0.84**	0.32	0.0	86
Employer characteristics in 1990: industrial sector				
agriculture	-0.13	0.17	3.8	5.1
construction	-0.30	0.25	2.2	14
trade	**-0.57**	0.25	0.0	91
communication, transport	**-0.56**	0.27	0.6	43
other services	**-0.75**	0.21	0.0	93
education, science	**-0.56**	0.20	1.8	18
Optimistic about the future in general in 1990	-0.07	0.15	3.4	6.4
Income very important for subjective well-being	**-0.30**	0.15	2.3	13
Expectations for the next 2 years in 1990				
redundancies in firm: certainly	**0.27**	0.12	0.1	78
losing the job: certainly	0.21	0.21	0.8	37
decline in professional career: certainly	**-0.82**	0.29	3.9	4.9
Other specification tests		$\chi^2(df)$	d.f.	p.-val.
Score test against nonnormality		0.95	2	62
Information matrix test				
all indicators		250.7	243	35
only main diagonal indicators		20.7	23	60

Note: **Bold** letters: t-value larger than 1.96. N = 1163 (1038 controls).

[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.

The t-values and score test results against heteroscedasticity presented in Table C.1 are computed using the GMM (or PML) formula given in White (1982).[29] The information matrix tests statistics are computed using the second version suggested in Orme (1988) that appeared to have good small sample properties.[30]

The results of the specification does do not provide any evidence against the chosen specification. First of all, the last two columns of Table C.1 do not contradict the assumption of conditional homoscedasticity. Furthermore, the normality test as well as the information matrix tests do not reject.

Appendix D: Implementation of the matching methods and descriptive statistics for the match quality

In this appendix the implementation of the different matching algorithms are explained further, and some descriptive statistics for the match quality are given.

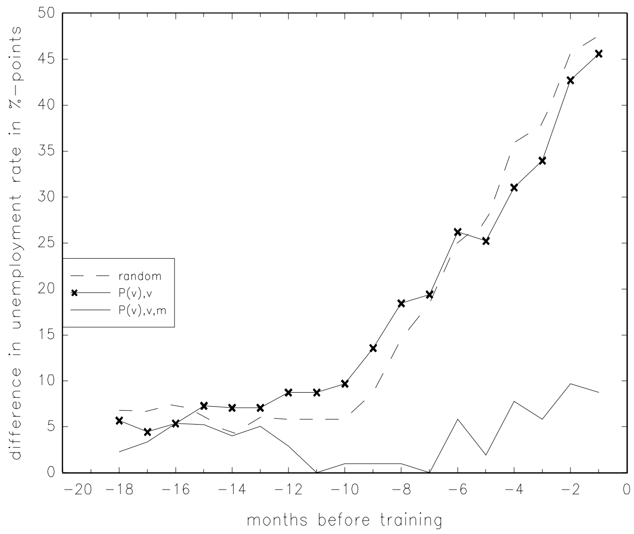
When using the partial propensity score for matching note that conditioning is on $v_n\hat{\beta}$ instead of $v_n\beta^0$. The asymptotic standard error, computed using the delta method, of $v_n\hat{\beta}$ resulting from the estimation of $\hat{\beta}$ can be considerable and ranges from 0.13 to 0.47 in the CTRT sample, and from 0.17 to 0.48 in the control sample. The mean in the CTRT (control) sample is 0.27 (0.26), the median 0.27 (0.26), and the empirical standard deviation 0.07 (0.06). 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. The details of the matching algorithm used are described in Appendix B.1. It follows Rosenbaum and Rubin (1985a) 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 are *gender*, *university*, *12 and 8 years of schooling*, *women in highly qualified or management job positions* (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.

[29] 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.

[30] 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 OPG matrix and the matrix of the expected hessian.

As has been argued in the main body of the paper conditioning on monthly employment information to capture the impact of temporary shocks could be important. Figure D.1 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 CTRT and different control samples relative to the number of months before CTRT. The two lines that are highest in the right hand part of the plot are based on the matching method mentioned so far plus a random draw in the control pool.[31] They are very similar and reveal unemployment rates that are up to 45%-points lower than for the CTRT sample. Conditioning additionally on the yearly and monthly pre-training employment information reduces the bias significantly. The additional variables used are *expectation of losing own job in the next two years (yearly), expectation of a declining career in the next two years (yearly), monthly wage / salary (yearly), training (unspecified, yearly), self-employment (yearly), highly qualified or management job positions (yearly), unemployment (monthly), STW (monthly), full-time work (monthly).* Although there is still a small upward bias, figures in the main body of the text show that it is not significantly different from zero. Therefore, all the following evaluations are based on this matched sample.

Figure D.1: Difference of pre-training unemployment between CTRT and matched control groups in %-points: a comparison of different matching algorithms



Note: *random*: CTRT observations are matched with random controls; *P(v),v*: CTRT observations are matched with controls using P(v) and selected v-variables; *P(v),v,m*: CTRT observations are matched with controls using P(v), selected v-variables and m-variables. See note to Tables 3 and D.1.

Table D.1 gives additional descriptive statistics for control groups based on different or no matching procedure and the CTRT group. The sample size used in the estimation of the coefficients of the partial propensity score of 125 is now reduced to 103 because censored spells

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

and variables that have missing values in variables necessary for this step of the analysis are deleted.

Table D.1: Descriptive statistics of selected variables (1990) of CTRT and control sample: different matching algorithms

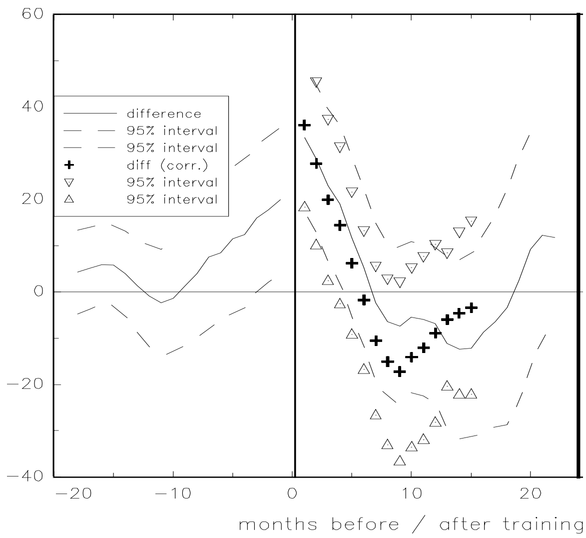
	Controls			CTRT
	all (1063)	matched on $v\hat{\beta}$ (103) and select. v-variables	matched on $v\hat{\beta}$ (103) and select. v-variables and m-var.	(103)
(1)	(2)	(3)	(4)	(5)
Variable	mean (std), share in %	mean(std), share in %	mean(std), share in %	mean(std), share in %
$v\hat{\beta}$	-1.50 (.55)	-0.91 (.52)	-1.01 (.51)	-0.87 (.56)
Gender: female	40	57	53	58
Federal states (Länder) in 1990: Berlin	7	16	8	11
Years of schooling (highest degree) in 1990				
12	18	26	19	28
10	60	65	67	63
Highest professional degree in 1990: university	13	19	17	19
Job position in 1990: highly qualified, manag.	22	26	17	26
Job characteristics in 1990: already fired	4	10	10	13
Profession in 1990 (ISCO)				
services, incl. trade, office	24	21	24	19
trade	4	7	5	6
Employer character. in 1990: industrial sector				
agriculture	13	13	13	17
construction	8	13	6	6
trade	8	12	8	8
communication, transport	9	5	3	5
other services	13	4	8	6
education, science	12	8	9	11
Optimistic about the future in general in 1990	13	10	14	13
Income very import. for subjective well-being	56	17	25	15
Expectations for the next 2 years in 1990				
redundancies in firm: certainly	50	48	57	52
losing the job: certainly	13	6	11	15
decline in professional career: certainly	2	1	3	3

Note: (2) no matching; (3) matched on $v\hat{\beta}$ (103) and selected v-variables; (4) matched on $v\hat{\beta}$ (103), selected v-variables and m (monthly, yearly)-variables; 1990 relates to the date of the interview that for almost all cases was completed before July 1990 (EMSU). v-variables used for the additional conditioning are: *gender, university, 12 and 8 years of schooling, women in highly qualified or management job positions* (1990), m-variables are: *expectation of losing own job in the next two years (yearly), expectation of a declining career in the next two years (yearly), monthly wage / salary (yearly), training (unspecified, yearly), self-employment (yearly), highly qualified or management job positions (yearly), unemployment (monthly), STW (monthly), full-time work (monthly);* see also note to Table A.1.

Appendix E: More evaluation results

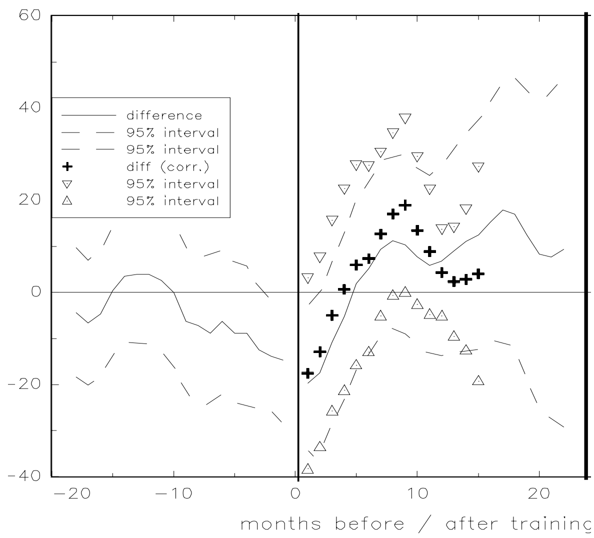
This appendix contains additional results for various outcome variables that have been omitted from the main body of the paper.

Figure E.1: Unemployment: only CTRT participants unemployed or STW before CTRT



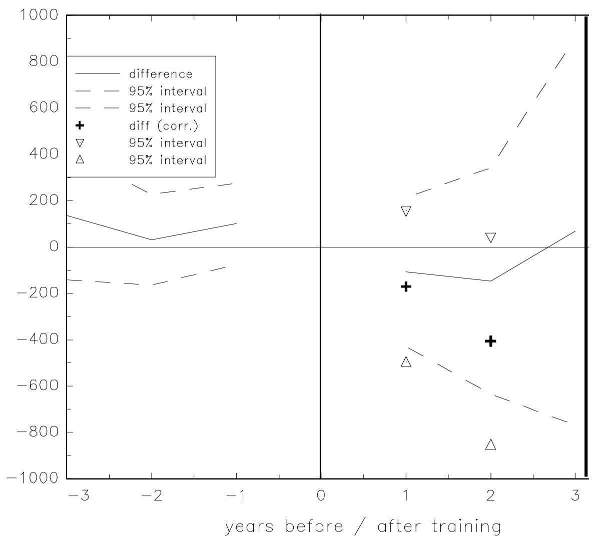
Note: N_{-1}^{t} = 75. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure E.2: Full-time employment: only CTRT participants unemployed or STW before CTRT



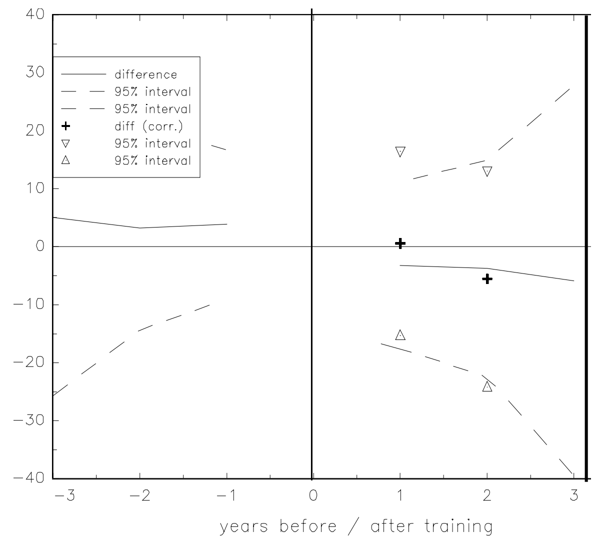
Note: N_{-1}^{t} = 75. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure E.3: Gross earnings (in 1993 DM)



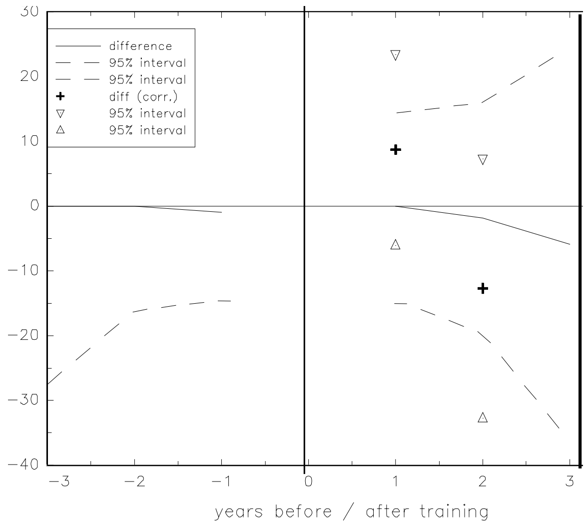
Note: N_{-1}^{t} = 103. Earnings when not employed coded as unemployment benefit or social assistance, whichever is higher. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure E.4: Very worried about possibility of future job loss (or unemployed)



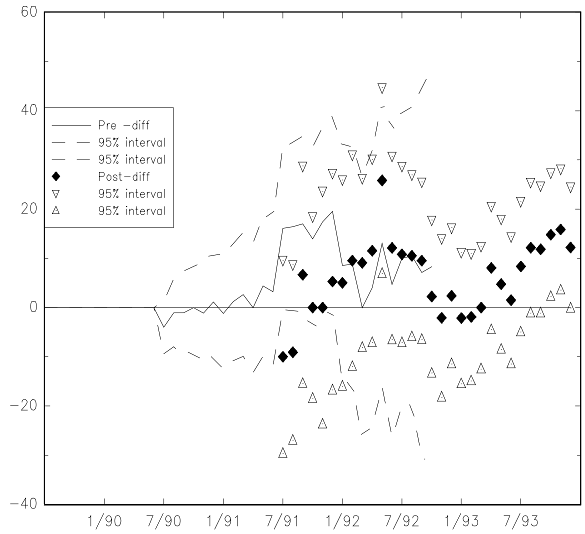
Note: N_{-1}^{t} = 103. Nonemployment coded as being very worried. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure E.5: Expected decline in the professional career in the next two years



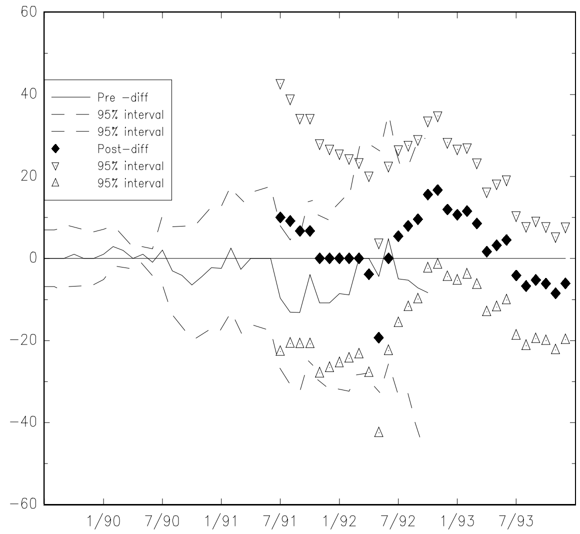
Note: N^{t}_{-1} = 103. Nonemployment coded as expecting decline. Smoothed using 3 month moving averages for $|\tau| > 1$.

Figure E.6: Registered unemployment (date)



Note: N^{t}_{-1} = 103.

Figure E.7: Full-time employment (date)



Note: N^{t}_{-1} = 103.

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 E6 and E7 show the development of pre-training (lines) and post-training outcomes (unconnected symbols) over time.[32] 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

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

CTRT 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.

References

Angrist, J.D. (1995): "Introduction to the JBES Symposium on Program and Policy Evaluation", *Journal of Business & Economic Statistics*, 13, 133-136.

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

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings", *The Review of Economics and Statistics*, 60, 47-57.

Ashenfelter, 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.

Bell, S.H., L.L. Orr, J.D. Blomquist, and G.G. Cain (1995): *Program Applicants as a Comparison Group in Evaluating Training Programs*, Upjohn: Kalamazoo.

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

Blaschke, D. und E. Nagel (1995): "Beschäftigungssituation von Teilnehmern an AFG-finanzierter beruflicher Weiterbildung", *MittAB*, 2/95, 195-213.

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, 1991): *Förderung der beruflichen Weiterbildung: Bericht über die Teilnahme an beruflicher Fortbildung, Umschulung und Einarbeitung im Jahr 1990*, Nürnberg.

Bundesanstalt für Arbeit (BA, 1992): *Förderung der beruflichen Weiterbildung: Bericht über die Teilnahme an beruflicher Fortbildung, Umschulung und Einarbeitung im Jahr 1991*, Nürnberg.

Bundesanstalt für Arbeit (BA, 1993a): *Förderung der beruflichen Weiterbildung: Bericht über die Teilnahme an beruflicher Fortbildung, Umschulung und Einarbeitung im Jahr 1992*, Nürnberg.

Bundesanstalt für Arbeit (BA, 1993b): *Geschäftsbericht 1993*, Nürnberg.

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

Bundesanstalt für Arbeit (BA, 1994b): *Geschäftsbericht 1994*, Nürnberg.

Bundesanstalt für Arbeit (BA, 1995): *Berufliche Weiterbildung: Förderung beruflicher Fortbildung, Umschulung und Einarbeitung im Jahr 1994*, 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.

Bundesministerium für Wirtschaft (1995), *Dokumentation, Nr. 382*.

Burtless, G. (1995): "The Case for Randomized Field Trials in Economic and Policy Research", *Journal of Economic Perspectives*, 9, 63-84.

Buttler, F. (1994): "Finanzierung der Arbeitsmarktpolitik", *IAB-Werkstattbericht*, Nr. 8, 31.8.1994.

Buttler, F. (1994): "Berufliche Weiterbildung als öffentliche Aufgabe", *MittAB*, 1/94, 33-42.

Buttler, F. and K. Emmerich (1994): "Kosten und Nutzen aktiver Arbeitsmarktpolitik im ostdeutschen Transformationsprozeß", *Schriften des Vereins für Sozialpolitik*, 239 / 4, 61-94.

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

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 Labour Economics: Theory and Public Policy*, 5th ed., New York: HarperCollins.

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

Fitzenberger, B. and H. Prey (1996): "Training in East Germany: An Evaluation of the Effects on Employment and Earnings", *Unpublished manuscript*, University of Konstanz.

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

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 Labour Market Data*, New York: Cambridge University Press.

Heckman, J.J. and J.A. Smith (1995a): "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, 9, 85-110.

Heckman, J.J. and J.A. Smith (1995b): "Ashenfelter´s Dip and the Determinats of Participation in a Social Program: Implications for a Simple Program Evaluation Strategies", *Research Report* # 9505, The University of Western Ontario.

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).

Infratest Sozialforschung (1990, 1991, 1992, 1993, 1994): *Das sozio-ökonomische Panel - Ost, Welle 1, Welle 2, Welle 3, Welle 4, Welle 5*, Anlagenbände zum Methodenbericht, München.

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

Institut für Arbeitsmarkt- und Berufsforschung (IAB, 1995): *Zahlen-Fibel 1995* (BeitrAB 101), Nürnberg.

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.

LaLonde, R.J. (1995): "The Promise of Public Sector-Sponsored Training Programs", *Journal of Economic Perspectives*, 9, 149-168.

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

Lechner, M. (1995a): "Effects of Continuous Off-the-job Training in East Germany after Unification", *Discussion Paper*, # 95-27, Zentrum für Europäische Wirtschaftsforschung, Mannheim.

Lechner, M. (1995b): "Some Specification Tests for the Panel Probit Model", *Journal of Business and Economic Statistics*, 13, 475-488.

Maddala, G.S. (1983): *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Müller, K. (1994): "Weiterbildung in Ostdeutschland: Ein Markt wird transparenter", *IAB-Werkstattbericht*, Nr. 4, 28.4.94.

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. (1995): *Weiterbildungsaktivitäten und Erwerbsbiographie*, Campus: Frankfurt.

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*.

Rosenbaum, P.R. (1984): "From Association to Causation in 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 (1985a): "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, 39, 33-38.

Rosenbaum, P.R. and D.B. Rubin (1985b): "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. (1977): "Assignement of Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 2, 1-26.

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.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-25.