

IZA DP No. 2121

**Get Training or Wait?
Long-Run Employment Effects of Training Programs
for the Unemployed in West Germany**

Bernd Fitzenberger
Aderonke Osikominu
Robert Völter

May 2006

Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany

Bernd Fitzenberger

*Goethe University of Frankfurt,
ZEW, IFS and IZA Bonn*

Aderonke Osikominu

Goethe University of Frankfurt

Robert Völter

*Goethe University of Frankfurt
and CDSEM, University of Mannheim*

Discussion Paper No. 2121
May 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany*

Long-term public sector sponsored training programs often show little or negative short-run employment effects and often it is not possible to assess whether positive long-run effects exist. Based on unique administrative data, this paper estimates the long-run differential employment effects of three different types of training programs in West Germany. We use inflows into unemployment for the years 1986/87 and 1993/94 and apply local linear matching based on the estimated propensity score to estimate the effects of training programs starting during 1 to 2, 3 to 4, and 5 to 8 quarters of unemployment. The results show a negative lock-in effect for the period right after the beginning of the program and significantly positive treatment effects on employment rates in the medium- and long-run. The differential effects of the three programs compared to one another are mainly driven by differences in the length of the lock-in periods.

JEL Classification: C14, J68, H43

Keywords: multiple treatments, training programs, employment effects, local linear matching, administrative data, active labor market programs

Corresponding author:

Bernd Fitzenberger
Department of Economics
Goethe-University
PO Box 11 19 32 (PF 247)
60054 Frankfurt am Main
Germany
Email: fitzenberger@wiwi.uni-frankfurt.de

* This study is part of the project "On the effectiveness of further training programs. An evaluation based on register data provided by the Institute of Employment Research, IAB" ("Über die Wirksamkeit von Fortbildungs- und Umschulungsmaßnahmen. Ein Evaluationsversuch mit prozessproduzierten Daten aus dem IAB"); IAB project number 6-531A). The data were compiled as part of the project jointly with the Swiss Institute for International Economics and Applied Economic Research at the University of St. Gallen (SIAW) and the Institut für Arbeitsmarkt- und Berufsforschung (IAB). We gratefully acknowledge financial support by the IAB. This paper has been presented at the Evaluation Conference at Paris, December 2005, and at the Rhein-Main-Neckar Labor Market Seminar at Mannheim, March 2006. We thank participants for helpful comments. All errors are our sole responsibility.

Contents

1	Introduction	1
2	Basic Regulation of Further Training	4
2.1	Programs	4
2.2	Financial Incentives for Participation	6
2.3	Participation	7
3	Data and Type of Treatment	7
3.1	Employment and Benefit Data	8
3.2	Monitoring Data for Training and the Merged Data Set	9
3.3	Contents and Types of Further Training	10
3.4	Three Important Programs	11
3.4.1	Practice Firm	12
3.4.2	Provision of Specific Professional Skills and Techniques	12
3.4.3	Retraining/Qualification via the Educational System	13
3.4.4	Identification of the Three Programs in the Data	13
3.5	Inflow Sample into Unemployment and Participation by Type of Training	14
4	Evaluation Approach	16
4.1	Multiple Treatment Framework	16
4.1.1	Extension to Dynamic Setting	17
4.1.2	Propensity Score Matching	18
4.1.3	Interpretation of Estimated Treatment Effect	20
4.2	Details of the matching approach	23
4.2.1	Local Linear Regression	24
4.2.2	Kernel Function and Bandwidth Choice	25
4.2.3	Bootstrapping	26
4.2.4	Balancing Test	26
5	Empirical Results	27
5.1	Estimation of Propensity Scores	27
5.2	Estimated Treatment Effects	30

5.2.1	Training versus Waiting	31
5.2.2	Pairwise Comparisons of Training Programs	33
6	Conclusions	35
	References	36
	Appendix	40
	Descriptive Statistics and Description of Data	40
	Estimated Employment Effects of Further Training Measures	43

1 Introduction

Public sector sponsored training has traditionally been a main part of active labor market policy (ALMP) in many countries and countries like Germany have devoted a lot of resources to training programs. During the last decade, there were many pessimistic assessments regarding the usefulness of public sector sponsored training programs in raising employment chances of the unemployed (see the surveys in Fay, 1996; Heckman et al., 1999; Martin and Grubb, 2001; Kluve and Schmidt, 2002). While the surveys emphasized that small scale training programs, which are well targeted to specific groups and which involve a strong on-the-job component, can show positive employment effects, these studies doubt that the large scale training programs in countries like Germany are successful in raising on average the employment chances of adult workers who became unemployed and who participate in such programs. Negative short run effects of these programs are attributed to a lock-in effect while being in the program.

Recently, OECD (2005) has emphasized that long-term labor market programs, such as training, often have little or negative short-run effects on outcomes. Also, it is clear that lock-in effects are worse for longer programs, because they keep the unemployed away from the labor market for a longer time. However, it could be the case that sizeable labor market effects are only to be expected from sufficiently long training programs (Fay, 1996). Therefore, it is crucial to assess program impacts in a longer term perspective in order to investigate whether the sizeable lock-in effects in the short run are compensated by positive long run effects. In fact, OECD (2005) reports positive long term results for some training programs. This paper adds to this literature by estimating the long-run employment effects of different types of training programs in Germany over at least six years since the beginning of the treatment.

The vast majority of the evaluation studies summarized in the aforementioned surveys used a static evaluation approach receiving treatment during a certain period of time against the alternative of not receiving treatment during this period of time. In a dynamic setting, the timing of events becomes important, see Abbring and van den Berg (2003, 2004), Fredriksson and Johansson (2003, 2004), and Sianesi (2003, 2004). Static treatment evaluations run the risk of conditioning on future outcomes leading to possibly biased treatment effects. This is because the nontreated individuals in the data might be observed as nontreated because their treatment starts after the end of the observation period or because they exit unemployment before

treatment starts (Fredriksson and Johansson 2003, 2004). This paper follows Sianesi (2003, 2004) and estimates the effects of treatment starting after some unemployment experience against the alternative of not starting treatment at this point of time and waiting longer.

For Germany, appropriate data for a long term evaluation of public sector sponsored training were not available for a long time and there existed serious scepticism in the German policy debate as to whether ALMP is actually effective (Hagen and Steiner, 2000). Until recently, basically all the evaluation studies¹ made use of survey data.² Although these data are rich with respect to informative covariates, the evaluation studies using survey data suffer from severe shortcomings with respect to the quality of the treatment information and to the precision of the employment history before and after treatment. The samples sizes in these studies are typically small. They do not allow the researcher to evaluate the effects of any heterogeneous treatment or of treatments targeted to specific groups of individuals.

Contributing to the debate on the effectiveness of ALMP, this paper analyzes the employment effects of three types of public sector sponsored training programs in West Germany. We use unique administrative data which have only recently become available. Using data on employment, periods of transfer payments, and participation in training programs, we carefully identify three types of public sector sponsored training programs for the unemployed. These programs are not associated with a regular job. The largest program among the three is the Provision of Specific Professional Skills and Techniques (SPST). SPST programs provide additional skills and specific professional knowledge in medium-term courses. The two other training programs are working in a Practice Firm (PF) and Retraining (RT). RT involves typically a two-year program providing complete vocational training in a new occupation. PF involves training in a work environment simulating a real job. PF tends to be a slightly shorter treatment than SPST. This classification of treatments is developed in this paper and in the earlier paper, Fitzenberger and Speckesser (2005). The three training programs considered here differ both in length and content (PF

¹See Speckesser (2004, chapter 1) and Wunsch (2006, section 6.5) as recent surveys for Germany. Previous studies based on survey data gave inconclusive evidence. For instance, for East Germany, Lechner (2000) found negative employment effects of training programs in the short run and insignificant effects in the long run based on survey data. In contrast, Fitzenberger and Prey (2000) found some positive employment effects of training programs in East Germany.

²Notable exceptions are the recent studies of Lechner et al. (2005a,b) and Fitzenberger and Speckesser (2005), which are all based on the same data set as our study. In fact, the data set is the outcome of a joint effort to merge administrative data for evaluation purposes, see Bender et al. (2005). The studies of Lechner et al. (2005a,b) and our study differ substantially regarding the exact treatment definition, the choice of valid observations, and the econometric methods used.

has the strongest on-the-job component, SPST involves typically of-the-job classroom training, RT involves both on-the-job and of-the-job training for a specific occupation). Based on the aforementioned evidence reported by Martin and Grubb (2001) and others, PF should be the most effective program, at least in the short run. In contrast, Lechner et al. (2005a) report quite favorable evidence for RT.

This paper takes advantage of unique administrative data which integrate register data on employment as well as data on unemployment and participation in active labor market programs generated by the Federal Employment Office (*Bundesanstalt für Arbeit*, BA). Our data set merges register data with benefit data and with survey data obtained from the local offices of the Federal Employment Office. This survey records all cases of participation in further training programs during the period 1980–1997 and offers rich information on heterogeneous courses. Our analysis evaluates the effects of training for inflows into unemployment for the years 1986/87 and 1993/94 in West Germany. These two inflow samples face very different labor market prospects due to changing business cycle conditions and the impact of German unification for the second sample. It is of interest to investigate whether the effects of ALMP differ by the state of the labor market due to business cycle effects. The 1986/87 sample faced a fairly favorable labor market in the years to come culminating in the unification boom in West Germany. In contrast, the 1993/94 sample entered unemployment during one of the most severe recessions in West Germany resulting in a prolonged period with bad labor market chances.

Since our analysis is based on administrative data, we have to use a non-experimental evaluation approach. We build on the conditional independence assumption purporting that for individuals with different treatment status the employment outcome in case of not receiving the treatment of interest is the same on average conditional on a set of covariates which cover socio-economic characteristics, the previous employment history of the individuals, the beginning of unemployment, and the elapsed duration of unemployment. The analysis uses the popular propensity score matching approach adjusted to a dynamic setting building on the recent work by Fredriksson and Johansson (2003, 2004) and Sianesi (2003, 2004). In fact, when the timing of treatment is a random variable, a static evaluation approach is not appropriate for nonexperimental data. We evaluate the employment effects of three multiple exclusive training programs both against the alternative of nonparticipation and in pairwise comparisons building on Lechner (2001) and Imbens (2000). Our matching estimator is implemented using local linear matching (Heckman, Ichimura, Smith, and Todd, 1998) based on the estimated propensity score. We run separate analyses conditional on elapsed duration of unemployment

at the beginning of treatment. We distinguish between training programs starting during quarters 1 to 2, 3 to 4, and 5 to 8 of unemployment.

Our analysis extends considerably upon the earlier work of Fitzenberger and Speckesser (2005) in several dimensions. The earlier paper evaluates the employment effects of SPST against the comprehensive alternative of nonparticipation in SPST for 36 months after the beginning of the treatment. The analysis is performed only for the 1993 inflow sample into unemployment, both for East and West Germany. This study analyzes the effects of three exclusive training programs for inflow samples in 1986/87 and 1993/94 in West Germany. The three programs are analyzed in a multiple treatment framework and we evaluate medium- and long-run treatment effects up to 25–31 quarters after the beginning of the treatment depending on the start date of the treatment.

The remainder of this paper is structured as follows: Section 2 gives a short description of the institutional regulation and participation figures for Active Labor Market Policy. Section 3 focuses on the different options of further training, their target groups, and course contents. Section 4 describes the methodological approach to estimate the treatment effects. The empirical results are discussed in section 5. Section 6 concludes. The final appendix provides further information on the data and detailed empirical results. An additional appendix, which is available upon request, includes further details on the data and on the empirical results.

2 Basic Regulation of Further Training

2.1 Programs

For the period covered by our data, further training in Germany is regulated on the basis of the Labor Promotion Act (*Arbeitsförderungsgesetz*, AFG) and is offered and coordinated by the German Federal Employment Office (formerly *Bundesanstalt für Arbeit*, BA). Originally, further training was conceived to improve occupational flexibility and career advancement and to prevent skill shortages. In response to unemployment becoming an increasingly persistent phenomenon during the 1980s and 1990s, further training changes its character from a rather preventive ALMP towards an intervention policy predominantly targeted to unemployed, to those at severe risk of becoming unemployed, and to problems groups in the labor market. With the increasing number of unemployed entering training programs, skill-upgrading courses

targeted to employed workers lose importance in favor of courses in which individuals are taught new technologies or are given the opportunity to enhance existing skills for the purpose of occupational reintegration.

The German legislation distinguishes three main types of training: further training, retraining, and integration subsidy. In addition, there is short-term training which only existed from 1979 to 1992. Although, during the 1980s and 1990s, there have been many changes concerning passive labor market policy – i.e. changes in benefit levels and eligibility criteria – the regulation of the traditional training schemes, further training, retraining, and integration subsidy, remained stable until the end of 1997 when the Labor Promotion Act was replaced by the Social Code III. In the following, we give a short description of these programs:

- **Further training** (*Weiterbildung*) includes the assessment, maintenance and extension of skills, including technical development and career advancement. The duration of the courses depends on individual predispositions, other co-financing institutions and adequate courses provided by the training suppliers.
- **Retraining** (*Umschulung*) enables vocational re-orientation if a completed vocational training does not lead to adequate employment. Retraining is supported for a period up to 2 years and aims at providing a new certified vocational education degree.
- As third program of further training, **integration subsidy** (*Einarbeitungszuschuss*) offers financial aid to employers who are willing to give employment to unemployed or to workers directly threatened by unemployment. The subsidy is paid for an adjustment period until the supported person reaches full proficiency in their job (up to 50% of the standard wage in the respective occupation).
- In 1979, **short-term training** was introduced under §41a AFG aiming at “increasing prospects of integration”. With this program, skill assessment, orientation and guidance should be offered to unemployed. The curricula under this program are usually short-term, lasting from two weeks up to two months and are intended to increase the placement rate of the unemployed.

2.2 Financial Incentives for Participation

Except for the integration subsidy which is a subsidy to a standard salary (according to union wage contracts), participants in training programs are granted an income maintenance (IM, *Unterhaltsgeld*) if they satisfy the conditions of entitlement. To qualify, they must meet a minimum work requirement of being previously employed during at least one year in contributory employment or they must be entitled to unemployment benefits or subsequent unemployment assistance.³

Starting 1986 until 1993, the income maintenance amounted to 73% of the relevant previous wage for participants with at least one dependent child and 65% otherwise. This was higher than the standard unemployment benefits (UB, *Arbeitslosengeld*) in this period which was at 68% and 63%, respectively. And IM was considerably higher for those unemployed whose UB expired and who were receiving the lower, means tested unemployment assistance (UA, *Arbeitslosenhilfe*) which amounted to 58% (with children) and 56% (without children).⁴ In 1994, income maintenance and unemployment benefits were both cut back to a common level of 67% (with children) and 60% (without children), taking away some of the financial incentives to join a training measure. Unemployment assistance was also lowered to 57% (with children) and 53% (without children).⁵

Receiving income maintenance while being in a training measure did not affect the entitlement period for unemployment benefit payments. Effectively, this means that the unemployed could defer the transition from unemployment benefits to unemployment assistance by taking part in a training measure. Additionally, while participating in a training measure, participants in some circumstances could requalify for unemployment benefits providing additional incentives for them to participate in programs.⁶

Summing up, for our time period under investigation, there are positive financial incentives for the unemployed to join a program. The income maintenance is at least

³If a person does not fulfil the requirement of previous employment, but had received unemployment assistance until the start of the program, an income maintenance may be paid as well.

⁴In the relevant, period the exhaustion of UB and transition from the higher UB to the lower UA took place between the 6th and the 32nd month of unemployment, depending on age and employment history (for details see Plassmann, 2002).

⁵For detailed descriptions of the changes in regulations over time see Bender et al. (2005) and Steffen (2005).

⁶This is because, until 1997, periods of income maintenance payments were counted on the minimum work requirement for receiving unemployment benefits, for details see Bender and Klose (2000).

as high or higher than unemployment benefits and it is always higher than unemployment assistance. Furthermore, participation allows to postpone the transition from unemployment benefits to the lower, means tested unemployment assistance and sometimes even allows to requalify for unemployment benefits. In addition, the BA bears all costs directly incurred through participation in a further training scheme, especially course fees.

2.3 Participation

Among the three main types of training programs distinguished by German legislation, the general further training scheme (*Berufliche Weiterbildung*) traditionally is the most important in West Germany, which can be seen from table 1. Starting with a total of 232,500 participants in 1980, 70% of all participants start a further training scheme, whereas only 14% (32,600) begin a program under the integration subsidy (*Eingliederungszuschuss*) scheme. New entrants into retraining (*Berufliche Umschulung*) sum up to 37,900, about 16% of the total. On average, the stock of participants is about 89,300 in 1980. In 1985, participant entries are 60% higher in total. By then, further training programs amount to 80% of all participant entries. Between 1980 and 1990, participation increases to 514,600, 74% of these are entries into further training programs. Participation in retraining increases from 37,900 in 1980 to 63,300 in 1990. In 1992, shortly after the German reunification, participation peaks at 574,700 in West Germany, then declines to 378,400 in 1996. The share of further training increases over time to 77%. The share of participants in retraining amounts to 20%.

3 Data and Type of Treatment

This evaluation study uses a database that integrates administrative individual data from three different sources. It contains spells on employment subject to social insurance contributions, on transfer payments by the Federal Employment Office during unemployment, and on participation in further training schemes. Relying both on the transfer payment and on the participation data, we are able to identify different types of training programs that are coherent within category. This is in contrast to earlier studies that bundle very heterogeneous types of treatment and which therefore provide much less informative evidence for policy makers.

3.1 Employment and Benefit Data

The core data for this evaluation are taken from the **IAB Employment Subsample** (*IAB Beschäftigtenstichprobe*, IABS) of the Institute for Employment Research (IAB), see Bender et al. (2000) and Bender et al. (2005, chapter 2.1). The IABS is a 1% random sample drawn from the mandatory employment register data for all employees who are covered by the social security system over the period 1975–97.⁷ Social insurance contributions are compulsory for dependent employees with earnings above a minimum wage that is free of social insurance contributions. Among the dependent employees specific groups such as individuals working on a marginal part-time basis and civil servants are excluded. Self-employed are also not contained. Although these groups are not sampled, the IABS covers up to 80% of the German labor force.

The second important source is the **Benefit Payment Register** (*Leistungsempfängerdatei*, LED) of the Federal Employment Office (BA), see Bender et al. (2005, chapter 2.2).⁸ These data consist of spells on periods of transfer payments granted to unemployed and program participants from the BA. Besides unemployment benefit or assistance, these data also record very detailed information about income maintenance payments related to the participation in further training schemes.

As already mentioned, the IABS is a random sample of all individuals who experience at least one spell of dependent employment subject to social security contributions between 1975–97. This sampling rule implies that one should restrict the analysis to entrants into unemployment who were previously employed. As we do not observe job seekers who do not draw some form of unemployment compensation, we are unable to observe large parts of the potential control group for treated individuals who did not experience employment subject to social insurance contributions before. Altogether, the IABS samples roughly 1% of the overall dependent employment and benefit receipt, resulting in 591,627 individuals and in 8,293,879 spells over the

⁷For the purposes of this study we obtained additional employment information from the IAB covering the years 1998–2002.

⁸In the format that is available from the German central archive for empirical social research, the IABS actually already contains information on transfer payments from the Federal Employment Office. However, these spells have been consolidated according to certain rules that do not correspond to the requirements of this project. For instance, periods of receipt of transfer payments from the BA are not reported if they lie parallel to employment records. However, participants in training measures who work during an internship may well be recorded as employed in the IABS. This implies a structural underreporting of treatment. Therefore, it has been necessary to merge the transfer spells with the employment data a second time. Now transfer payments parallel to employment spells are included.

period 1975–97 for both East and West Germany.

3.2 Monitoring Data for Training and the Merged Data Set

The training participation data (FuU–data) are collected for all participants in further training, retraining, integration subsidies and language courses in Germany for internal monitoring and statistical purposes, see Bender et al. (2005, chapter 2.3) for a description of the FuU–data. Aggregate statistics based on these data were published regularly by the BA. These data report information about the type of course, the intended integration objectives and – with less detail – about the contents of the courses with respect to the skills provided. In addition, they provide information about personal characteristics of the participants in training programs (like sex, age, nationality, educational attainment, employment status before treatment), the region in which the program takes place, the type of training (whether the training takes place in classrooms or “on the job”), the providing institution and the beginning and end dates. The data also indicate the type of income maintenance paid during participation. The sample size of the FuU–data amounts to 54,767 individuals corresponding to 72,983 treatment spells in the period 1980–97 (for West Germany, and 1991–97 for the new federal states). In principle, individuals in the combined IABS–LED data receiving subsistence payments, that are related to participation in a training program, should also be contained in the FuU–data.⁹

The FuU–data were merged with the combined IABS–LED data by social insurance number and additional covariates, see Bender et al. (2005, chapters 3–4). The merged data form the basis for an integrated evaluation database consisting of comparable longitudinal information for treatment and control groups that covers all participants in further training, retraining, integration subsidies, short–term training as well as language training.

In addition to merging the different files, numerous corrections have been implemented in order to improve the quality of the data. Inconsistencies in both files, which occurred with respect to the reported level of education and occupational status, the year of birth and the family status, were removed. In particular, a correct assessment of the level of education is important for the purpose of this study, because individual skills play an important role for the assignment into treatment. As

⁹However, there are some exceptions to this rule. In fact, we observe participants in the FuU–data who do not receive income maintenance. Using the merged data is the only option to fully identify treatments.

the information on the individual’s education is recorded by the employer and mis-reports do not have any consequences on social insurance entitlements, we assume that this variable rather reflects the level of education necessary to fulfil the tasks in the individual’s current job. The individual’s formal skill level may, however, very well lie above the education level reported by the employer.¹⁰ A detailed description of the correction procedure implemented in this study can be found in Bender et al. (2005, chapter 3.4).

3.3 Contents and Types of Further Training

The legislation of further training provides only a very elementary framework for the purpose of classifying different kinds of treatments. It does not define treatment types with respect to specific integration targets or target groups. In fact, very different kinds of treatments are administered under the same heading, e.g. training for career advancement or short-term courses for long-term unemployed are both reported as ‘further vocational training’. Earlier descriptive studies on the types of treatment often do not distinguish treatments providing basic social skills from treatments offering certified professional skills, which might have a very different impact on job search.¹¹

In contrast to earlier work, we are able to identify different types of further training according to economically meaningful criteria with the merged database we are using in this study. The combination of benefit and participation data allows us to identify whether a treatment is provided outside a firm specific labor market or within a firm, whether the course is general or occupation specific. Our paper is one of the first studies on further training in Germany that exploits all available information from administrative data, using the occupational status while being on training, the specific information about benefit payments (which can be related to specific types of interventions) and a variable recording the type of training in the monitoring information on further training (FuU-data).¹²

¹⁰Fitzenberger, Osikominu, and Völter (2006) analyze the quality of the education variable in German employment register data and provide imputation methods for correcting it.

¹¹An exception is the study by Blasche and Nagel (1995) that distinguishes whether the training is carried out as adjustment or retraining and whether it is a full-time or part-time treatment.

¹²The training data should actually be sufficient to identify all participations in further training since they should have been collected for all training spells started under the Labor Promotion Act. However, there are two reasons which do not permit to rely only on the variable of type of training from the FuU-data: First, these training data are sometimes incomplete because data collection is not directly linked to benefit payments and reporting errors are therefore unlikely to be detected. In such cases, one observes an income maintenance spell in the benefit data but no

In order to identify informative (and coherent) types of treatment, we apply a typology that relies on the type of training given in the FuU–data and the closeness to the demands of the labor market as indicated by the IABS–data on employment status.¹³ In fact, employment status and program information are especially important. While the program information ‘further vocational training’ might comprise both employed and unemployed participants, the employment status allows additionally to identify the target group (‘improving reintegration chances’ for unemployed or ‘career advancement’ for employees) or to indicate how close the program is related to an internal labor market. A combination of training and employment data is more informative than the unmodified information from the training data, because the latter data do not specify details of the implementation and the target group of a program.

Based on the information from both sources, a variety of different treatments can be identified, which range from the provision of social skills and basic general training over the provision of specific skills for the purpose of reintegration of the unemployed, the integration of unemployed into firm specific labor markets, retraining and the promotion of certified occupations up to career advancement training that is supplied to persons without the risk of unemployment, see Speckesser (2004) for a more detailed description.

3.4 Three Important Programs

In this study, we focus on three quantitatively important programs targeted at the unemployed: practice firm (PF), provision of specific professional skills and techniques (SPST), and retraining (RT). Whereas PF and SPST usually take six to nine months, RT is very time intensive and usually lasts two years. The next subsections will give more information about the contents of these programs.¹⁴ With little change in their basic design, the three programs considered here have still been in widespread use until recently, see the report by the Federal Commission for

training spell in the FuU–data. Thus, it is necessary to also consider the information in the benefit data to identify the full extent of participation in a program. Second – and equally important – the use of employment data and benefit data increases the precision of information on the type of training. It allows to find out whether a person was employed while being participant or whether a specific benefit was paid, both offering additional valuable information about the participant’s type of treatment.

¹³This typology has been developed by Speckesser (2004, chapter 3).

¹⁴A description of the other types of training as e.g. courses targeted at other groups like career advancement subsidies for the employed and wage subsidies for people who recently found a job can be found in the additional appendix.

Education Planning and Research BLK (2000).

3.4.1 Practice Firm

Practice firms are simulated firms in which participants practice everyday working activities. The areas of practice are whole fields of profession, not specific professions. Hence, practice firms mainly train general skills while provision of new professional skills is of less importance. Some of practice firms are technically oriented, the practice studios, whereas others are commercially oriented, the practice enterprises. One of the practice firm's goals is to evaluate the participant's aptitude for a field of profession. The programs usually last for six months and do not provide official certificates.

3.4.2 Provision of Specific Professional Skills and Techniques

Provision of specific professional skills and techniques intends to improve the starting position for finding a new job by providing additional skills and specific professional knowledge in medium-term courses. It involves refreshing specific skills, e.g. computer skills, or training on new operational practices. SPST is targeted at unemployed or persons at risk of becoming unemployed in order to facilitate integration into full employment. This program mainly consists of classroom training but an acquisition of professional knowledge through practical working experience may also be provided.

After successfully completing the course, participants usually obtain a certificate indicating the contents of the course, i.e. the refreshed or newly acquired skills and the amount of theory and practical work-experience. Such a certificate is supposed to serve as an additional signal to potential employers and to increase the matching probability since the provision of up to date skills and techniques is considered to be a strong signal in the search process. The provision of specific professional skills and techniques aims at achieving the sustained reintegration of unemployed or workers at risk of unemployment into the labor market by improving skills as well as providing signals. Compared to retraining, that is a far more formal and thorough training on a range of professional skills and that provides a complete occupational degree, the role of SPST for a participant's occupational knowledge is weaker. However, the amount of occupation specific knowledge imparted in SPST certainly exceeds the level provided in short-term programs that usually aim at improving job search

techniques or general social skills. Thus, SPST ranges in the middle between very formal (and very expensive) courses on the one hand and very informal and general courses on the other side. As can be seen from see table 2, SPST is quantitatively the most important type of training for the unemployment cohorts analyzed here.

3.4.3 Retraining/Qualification via the Educational System

This type of training consists of the provision of a new and comprehensive vocational training according to the regulation of the German dual system. It is offered to individuals who already completed a first vocational training and face severe difficulties in finding a new employment within their profession. It might however also be offered to individuals without a first formal training degree if they fulfil additional eligibility criteria. Up to 1994, this type of treatment is also accessible to individuals without the formal criterion of “necessity” for career advancement. Participants are then granted income maintenance as a loan.

Qualification via the educational system/retraining provides widely accepted formal certificates according to German dual system of vocational training. It comprises both, theoretical training and practical work experience. The theoretical part of the formation takes place in the public education system. The practical part is often carried out in firms that provide work experience in a specific field to the participants, but sometimes also in interplant training establishments. This type of treatment leads to a formal job qualification that is supposed to improve the job match.

3.4.4 Identification of the Three Programs in the Data

Participation in SPST and RT can be identified from either LED–data or FuU–data, participation in PF only from FuU–data. In the best case, both sources provide coherent information about participation and one can easily identify the type of treatment from both data sources. However, due to quality deficiencies in the participation data, many participants might not be recorded in the FuU–data. In this case, the LED–data helps to identify the treatment on the basis of the benefit variable which itself offers very detailed information about the type of training. In other cases, we observe individual records showing employment in the IABS and at the same time training in the FuU–data. This is for example the case if the treatment takes place in a firm and individuals are paid a normal salary (e.g.

integration subsidy) or if individuals are prepared for precise job offers. Since we have two separate sources of data, we make use of all available information and combine benefit information with participation data in order to identify all different types of training.¹⁵

3.5 Inflow Sample into Unemployment and Participation by Type of Training

The goal of this study is to analyze the effect of training programs on employment chances of unemployed individuals. Therefore, we base our empirical analysis on inflow samples into unemployment. We use the inflows into unemployment in the years 1986/87 and 1993/94 in West Germany, omitting Berlin and East Germany. Effectively, we consider individuals who experience a transition from employment to nonemployment and for whom a spell with transfer payments from the Federal Employment Office starts within the first twelve months of nonemployment or for whom the training data indicates a program participation before the unemployed individual finds a new job. In the following, we denote the start of the nonemployment spell as the beginning of the unemployment spell. We condition on receipt of unemployment compensation or program participation to exclude most of the individuals who move out of labor force after ending their job. This concerns especially individuals whose treatment status would be nonparticipation in any training program during their nonemployment spell. A treatment is associated with an unemployment spell if the individual does not start employment before the second month of treatment. Furthermore, we restrict our samples to the 25 to 55 years old in order to rule out periods of formal education or vocational training as well as early retirement.

We choose the years 1993/94 and 1986/87 to make comparisons between the eighties and the nineties. Figure 1 depicts the unemployment rate in West Germany. The dotted vertical lines mark the years 1986 and 1993, respectively. Whereas in 1986/87, we are at the end of a sequence of years with relatively high unemployment, the cohort 1993/94 enters during a period with increasing unemployment rates. Thus, the 1986/87 cohort faced a fairly favorable labor market in the years to come culminating in the unification boom in West Germany, while the 1993/94 cohort entered unemployment during one of the most severe recessions in West Germany resulting

¹⁵The additional data appendix describes in details, which variables were used for this purpose and how we identified the different program types giving a precise coding plan. Tables 2, 3 and 4 in the additional appendix show the shares by type of treatment that are identified from the participation data, from the benefit payment data or from a combination of both.

in a prolonged period with bad labor market chances. Our data allow to follow individuals entering unemployment in 1986/87 until December 1996/97 and individuals entering unemployment in 1993/94 until the end of 2001/02. A second reason for analyzing two different cohorts is that programs became more targeted over time, see section 2.1.

Table 2 gives information about the size of the inflow samples and the incidence of training. We focus on the three types of training programs which are most suitable for unemployed individuals and which do not involve on-the-job training (training while working in a regular job). These are (i) practice firm (PF), (ii) provision of specific professional skills and techniques (SPST), and (iii) retraining (RT). The total inflow sample comprises 20,902 spells for the 1986/87 cohort and 25,051 spells for the 1993/94 cohort. There are 1,714 training spells for the eighties and 2,727 for the nineties. Thus, about 10% of all unemployed participate in one of the three training programs considered. Among these, SPST represents the by far largest type of training with 64% and 72% of the training spells, respectively in the two samples. About one fifth of all training spells are RT, and PF represents the smallest group in both samples. In absolute numbers, there are 246 (325) PF spells in the 86/87 (93/94) inflow sample, 1,093 (1,944) SPST spells and 375 (458) RT spells. Table 3 shows the frequency of training by time window of elapsed unemployment.

Table 4 provides descriptive statistics on the elapsed duration of unemployment at the beginning of treatment. Our discussion focuses on quantiles because averages can be biased due to outliers. The median entrant in PF has been unemployed for 10 months in 1986/87 and 9 months 1993/94. Late starts (75%-quantile) of PF occur after 19 months in 1986/87 and much earlier in the other sample. For RT, the quantiles in the samples are very similar. With a median of 6 and 7 months, RT starts the earliest in both samples. For SPST, we find a reversed trend in comparison to PF. While SPST participation starts almost as early as RT in the 1986/87 sample, the starting dates are noticeably later in 1993/94, with the median increasing from 6 to 11 months.

Table 5 provides descriptive information on the duration of training spells. The average duration of practice firm is similar in both samples with 5.1 months in 1986/87 and 5.7 months in 1993/94. SPST has an average duration of 4.9 months for the 1986/87 sample and 6.3 months for 1993/94. For Retraining, we have the same picture as for SPST but at a higher level. This program lasts on average 13.1 months in the eighties and 14.9 months in the nineties.

4 Evaluation Approach

We analyze the employment effects of $K = 3$ different types of public sector sponsored training programs, where we compare the different treatments among one another as well as to nonparticipation in any of the programs. Our outcome variable is the average difference in employment rates measured at a quarterly frequency. The quarterly employment rate is obtained as the average of three monthly employment dummy variables. We estimate the average treatment effect on the treated (ATT), i.e. the differential impact the treatment shows for those individuals who participate in the treatment under consideration, in a context where individuals have multiple treatment options. Extending the static multiple treatment approach to a dynamic setting, we further distinguish three types of treatment depending upon the quarter in which the program starts in terms of the elapsed unemployment duration. Our dynamic evaluation approach following Sianesi (2003, 2004) applies the standard static treatment approach recursively depending on the elapsed unemployment duration. The estimated ATT parameter thus mirrors the decision problem of the case worker and the unemployed who recurrently during the unemployment spell decide whether to begin any of the programs now or to postpone participation to the future.

In the following, we first discuss our extension of the static multiple treatment approach to a dynamic setting. Then, we describe the implementation of the matching estimator for our problem.

4.1 Multiple Treatment Framework

Our empirical analysis is based upon the potential–outcome–approach to causality, see Roy (1951), Rubin (1974), and the survey of Heckman, LaLonde, Smith (1999). Lechner (2001) and Imbens (2000) extend this framework to allow for multiple, exclusive treatments. Lechner (2001) and Gerfin and Lechner (2002) show how to extend standard propensity score matching estimators for this purpose.

Let the $K + 1$ potential outcomes be $\{Y^0, Y^1, \dots, Y^K\}$, where Y^k , $k = 1, \dots, K$, represents the outcome associated with treatment k and Y^0 is the outcome when receiving none of the K treatments. To simplify the discussion, we will from now on refer to the nontreatment outcome Y^0 as one of the $K + 1$ treatment outcomes. For each individual, only one of the $K + 1$ potential outcomes is observed and the remaining K outcomes are counterfactual.

One can define pairwise average treatment effects on the treated (ATT)

$$(1) \quad \theta(k, l) = E(Y^k - Y^l | T = k) \text{ with } k, l = 0, 1, \dots, K \text{ and } k \neq l,$$

where $T = 0, 1, \dots, K$ represents the treatment received. The individual treatment effect is the difference between the outcome Y^k and the outcome Y^l , where the latter is not observed for the treated individuals in $\{T = k\}$. In the following, we call the individuals with treatment status $\{T = k\}$ the k -group and the individuals with $\{T = l\}$ the l -group. Note that in general $\theta(k, l) \neq \theta(l, k)$ because the characteristics of participants in treatment k differ from those for participants in treatment l .

4.1.1 Extension to Dynamic Setting

As already mentioned, we use the static multiple treatment framework in a dynamic context. Our basic samples consist of individuals who start an unemployment spell with transfer payments in 1986/87 or 1993/94 and who had been employed before. These individuals can participate in any of the three training programs at different points of time in their unemployment spell. Both the type of treatment and the selectivity of the treated individuals depend upon the exact starting date of the program. Abbring and van den Berg (2003) and Fredriksson and Johansson (2003, 2004) interpret the start of the program as an additional random variable in the “timing of events”. Unemployed individuals are not observed to participate in a program either because their participation takes place after the end of the observation period or because they leave the state of unemployment either by finding a job or by moving out of labor force.

Fredriksson and Johansson (2003, 2004) argue that it is incorrect to undertake a static evaluation analysis by assigning unemployed individuals to a treatment group and a nontreatment group based on the treatment information observed in the data. Consider the case of analyzing treatment irrespective of the actual starting date during the unemployment spell. If one assigns individuals to the control group who find a job later during the observation period, one effectively conditions on future outcomes when defining the treatment indicator. This might lead to a downward bias in the estimated treatment effect, which is the bias emphasized by Fredriksson and Johansson (2003, 2004). An upward bias can arise as well when future participants, whose participation starts after the end of the observation period, are assigned to the control group. Using duration analysis in discrete time, Fredriksson and Johansson (2004) suggest a matching estimator for the treatment effect based on a time-varying

treatment indicator. Treatment can only start at discrete points of time. In a similar vein, Sianesi (2004) argues for Sweden that all unemployed individuals are potential future participants in active labor market programs, a view which is particularly plausible for countries with comprehensive systems of active labor market policies like Sweden or Germany.

The above discussion implies that a purely static evaluation of the different training programs is not warranted.¹⁶ Following Sianesi (2003, 2004), we extend the static framework presented above in the following way.¹⁷ We analyze the employment effects of the first participation in a training program during the unemployment spell considered *conditional on the starting date of the treatment*.¹⁸ We distinguish between treatment starting during quarters 1 to 2 of the unemployment spell (stratum 1), treatment starting during quarters 3 to 4 (stratum 2), and treatment starting during quarters 5 to 8 (stratum 3). We do not analyze treatments starting later than quarter 8. This allows us to study treatment effects for a long time period after the beginning of the treatment. Moreover, for treatments starting after 8 quarters, it seems increasingly difficult to distinguish unemployment from being out-of-labor force. It is very likely that a number of individuals have left the labor force after two years of unemployment. Our data do not allow us to distinguish the labor market states unemployment and out-of-labor force, see the discussion in section 3.

4.1.2 Propensity Score Matching

We evaluate the differential effects of multiple treatments under the Conditional Independence Assumption (CIA), i.e. we assume that conditional on individual characteristics X the potential outcomes $\{Y^0, Y^1, \dots, Y^K\}$ are independent of treatment status T . Building on Rosenbaum and Rubin's (1983) result on the balancing property of the propensity score in the case of a binary treatment, Lechner (2001) shows

¹⁶In a static setting one has to deal, in addition, with the problem that the potential starting dates of the nonparticipants are unobserved. Drawing random starting times of the program is an alternative to use in this context, see e.g. Lechner (2000) and Lechner et al. (2005a,b) for this approach. However, this does not overcome the problems discussed here and we prefer to consider the timing of events explicitly. We do not introduce a random timing of the program starts among the nonparticipants for the following three reasons. First, random starting dates add noise to the data. Second, the drawn starting time might be impossible in the actual situation of the nontreated individual. Third, drawing random starting dates does not take the timing of events seriously.

¹⁷We do not follow Abbring and van den Berg (2003) and Fredriksson and Johansson (2004) in estimating hazard rate models because we are interested in the total employment effects irrespective of multiple transitions between employment and nonemployment.

¹⁸We do not analyze multiple sequential treatments, see Bergemann et al. (2004), Lechner and Miquel (2001), and Lechner (2004).

that the conditional probability of treatment k , given that the individual receives treatment k or l , exhibits an analogous balancing property for the pairwise estimation of the ATT's $\theta(k, l)$ and $\theta(l, k)$. Formally, we have

$$(2) \quad E(Y^l | T = k, P^{k|kl}(X)) = E(Y^l | T = l, P^{k|kl}(X))$$

and analogously

$$E(Y^k | T = l, P^{k|kl}(X)) = E(Y^k | T = k, P^{k|kl}(X)) .$$

$P^{k|kl}(X)$ is the conditional probability of treatment k , given that the individual receives treatment k or l , i.e.

$$P^{k|kl}(X) = \frac{P(T = k | X)}{P(T = k | X) + P(T = l | X)} \equiv \frac{P^k(X)}{P^k(X) + P^l(X)} .$$

Estimating $\theta(k, l)$ and $\theta(l, k)$ under the CIA, the balancing property in equation (2) allows one to apply standard binary propensity score matching based on the sample of individuals participating in either program k or l (Lechner, 2001; Gerfin and Lechner, 2002; Sianesi, 2003). In this subsample of the data, one simply estimates the probability of treatment k versus l , yielding an estimate of the conditional probability $P^{k|kl}(X)$, and then applies standard matching techniques for the binary case.

In order to account for the dynamic nature of the treatment assignment process, we estimate the probability of treatment k versus l given that unemployment lasts long enough to make an individual 'eligible'. For treatment during quarters 1 to 2, we take the total sample of unemployed, who participate in k or l during quarters 1 to 2, and estimate a Probit model for participation in k . If the comparison involves nonparticipation in any treatment, then this group includes those unemployed who either never participate in any program or who start treatment after quarter 2. For treatment during quarters 3 to 4 or quarters 5 to 8, the basic sample consists of those unemployed who are still unemployed in the first month of the stratum considered. We then estimate a Probit of beginning a program within the considered time interval of elapsed unemployment duration using all individuals who are still unemployed at the beginning of the stratum and who participate in k or l during the time interval. Sianesi (2004) estimates a separate Probit model for each of the different starting dates of the programs. In our case, the number of observations is too small for this. However, even if enough data were available, we think that it would not be advisable to estimate Probit regressions by quarter (or even by month).

The reason is that the starting date of the treatment is somewhat random (relative to the elapsed duration of the unemployment spell) due to available programs starting only at certain calendar dates. Therefore, we pool the treatment Probit for all eligible persons in unemployment in three strata assuming that the exact starting date is random within the time interval considered. However, when matching treated and nontreated individuals, we align by elapsed unemployment duration at the start of the program.

As already mentioned, we aggregate the relative starting dates into three strata while the employment status is measured at a quarterly frequency. We therefore implement a stratified matching approach in order to estimate the ATT for k relative to l . Effectively, we impose as a matching requirement that the comparisons with treatment l for an individual receiving treatment k are still unemployed in the quarter before treatment k starts. This way, we only match participants in l who might have started a treatment k in the same quarter as the respective participant in k . In the following, we call this subset of the l -group the eligible l -group. Second, within this group of eligible l -participants, we match individuals based on their similarity in the calendar month of the beginning of the unemployment spell and based on their similarity in the estimated propensity score by local linear matching. The expression for the counterfactual outcome in treatment l for the treated k -individual is then obtained through a local linear regression of the treatment outcome in l on the estimated propensity score and the starting month of the unemployment spell. For weighting, we use a bivariate product kernel (details will be described below). This local linear regression is estimated in the eligible subset of the l -group matched to the individual receiving treatment k . This way, we obtain a close alignment in calendar time as well as elapsed unemployment duration, thus avoiding drawing random starting times of the program.

4.1.3 Interpretation of Estimated Treatment Effect

Our estimated ATT parameter has to be interpreted in a dynamic context. We analyze treatment conditional upon the unemployment spell lasting at least until the start of the treatment k and this being the first treatment during the unemployment spell considered. Therefore, the estimated treatment parameter is (similar to

equation (1) in Sianesi, 2004)¹⁹

$$(3) \quad \theta(k, l; u, \tau) = E(Y^k(u, \tau) | T_u = k, U \geq u-1, T_1 = \dots = T_{u-1} = 0) \\ - E(Y^l(\tilde{u}, \tau - (\tilde{u} - u)) | T_u = k, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = \dots = T_{u-1} = 0) ,$$

where T_u is the treatment variable for treatment starting in quarter u of unemployment. $Y^k(u, \tau)$, $Y^l(u, \tau)$ are the potential treatment outcomes for treatments k and l , respectively, in periods $u + \tau$, where treatment starts in period u and $\tau = 0, 1, 2, \dots$, counts the quarters since the beginning of treatment. U is the duration of unemployment, \tilde{u} is the random quarter when alternative treatment l starts, and $\bar{u} = 2, 4, 8$ is the last quarter in the stratum of elapsed unemployment considered. Then, $\tau - (\tilde{u} - u)$ counts the quarters since start of treatment l yielding alignment of unemployment experience, because $u + \tau = \tilde{u} + (\tau - (\tilde{u} - u))$, and $Y^l(\tilde{u}, \tau - (\tilde{u} - u))$ is the outcome of individuals who receive treatment l between period u and \bar{u} .²⁰ For starts of l later than u , we have $\tilde{u} - u > 0$ and therefore, before l starts, $\tau - (\tilde{u} - u) < 0$. Then, these individuals are still unemployed, i.e. $Y^l(\tilde{u}, \tau - (\tilde{u} - u)) = 0$ when the second argument of $Y^l(., .)$ is negative. This way, we account for the fact that alternative treatments, for which the individual receiving treatment k in period u is eligible, might not start in the same quarter u . Also, each member of the eligible subset of the l -group is used in the pairwise comparisons for treatment k .

Conditioning on past treatment decisions and outcomes, the treatment parameter for a later treatment period (quarters 3 to 4 or quarters 5 to 8) is not invariant with respect to changes in the determinants of the exit rates from unemployment or the treatment propensity in the earlier phase of the unemployment spell. This is a direct consequence of modeling heterogeneity with respect to the starting time of the treatment relative to the length of elapsed unemployment. Both the k -group and the l -group at the start of the treatment are affected by the dynamic sorting effects taking place before, see Abbring and van den Berg (2004) for a recent dis-

¹⁹In contrast to equation (3), Sianesi (2004) conditions on being unemployed in period u , i.e. $U \geq u$. We use the restriction $U \geq u-1$ defining eligibility for treatment assuming that the assignment to treatment can occur up to one quarter before beginning of treatment. An unemployed in $u-1$ might anticipate obtaining a job in u . For this reason, our estimated treatment effect for treatment versus receiving no treatment at all might be conservatively downward biased.

²⁰Based on monthly data, Sianesi (2003) restricts the comparison to treatment l starting in the same month as treatment k . In our setup, where starting times are aggregated into three strata, that would leave eligible individuals for comparison with treatment k starting in period u not being used in any pairwise treatment combination. That is because if u lies before the end of the time window of elapsed unemployment considered, then some individuals in the eligible subset of the l -group receive treatment after u .

discussion of this problem in the context of estimating duration models. Thus, the estimated treatment parameter depends dynamically on treatment decisions and outcomes in the past when taking the timing of events seriously (Abbring and van den Berg, 2003; Fredriksson and Johansson 2003, 2004; Sianesi 2003, 2004). To avoid this problem, one often assumes a constant treatment effect over the duration of elapsed unemployment at the program start. Alternatively, other suitable uniformity or homogeneity assumptions for the treatment effect could be used. Such assumptions are not attractive in our context.²¹ By taking account of the dynamic sorting effects taking place before treatment, there is no simple relationship between our estimated treatment parameter in equation (3) and the static ATT in equation (1), the literature typically attempts to estimate.²²

Using propensity score matching in a stratified manner, we estimate the treatment parameter in (3) allowing for heterogeneity in the individual treatment effects and for an interaction of the individual treatment effects with dynamic sorting taking place. To make this a valid exercise, we assume the following dynamic version of the conditional mean independence assumption (DCIA) to hold for our inflow sample into unemployment

$$(4) \quad E(Y^l(\tilde{u}, \tau - (\tilde{u} - u)) | T_u = k, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = \dots = T_{u-1} = 0, X) \\ = E(Y^l(\tilde{u}, \tau - (\tilde{u} - u)) | T_{\tilde{u}} = l, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = \dots = T_{u-1} = 0, X) ,$$

where X are time-invariant (during the unemployment spell) characteristics, $T_{\tilde{u}} = l$ indicates treatment l between u and \bar{u} , the end of the stratum of elapsed unemployment considered, and $\tau \geq 0$, see equation (3) above and the analogous discussion in Sianesi (2004, p. 137). We effectively assume that conditional on X , conditional on being unemployed until period $u-1$, and conditional on not receiving treatment k or l before u (both referring to treatment in period u) individuals are comparable in their outcome for treatment l occurring between u and \bar{u} .

For $l = 0$, i.e. the comparison to the nontreatment alternative, the treatment parameter in (3) is interesting when each time period one decides whether to start treatment in the next quarter against the alternative to postpone possible treatment to the future (treatment now versus waiting, see Sianesi, 2004). For $l \neq 0$

²¹Sianesi (2003) reports ‘synthetic’ averages over the relative starting dates u , $\sum_u \frac{N_{k,u}}{N_k} \theta(k, l; u, \tau)$, to provide a summary statistic of the u specific treatments. These estimated averages have by themselves no causal interpretation.

²²Fitzenberger and Speckesser (2005) provide a more detailed discussion of the relationship between the static and dynamic treatment parameter in the binary treatment case $K = 1$.

and $k \neq 0$, the treatment parameter in (3) is interesting when each time period one decides whether to start treatment k in the next quarter against the alternative to receive treatment l at some point in the near future before the end of the considered time window in elapsed unemployment (treatment k versus l in a dynamic context, see Sianesi, 2003). In addition, exits from unemployment in a certain period are not known in the period until they take place. Anticipation effects might invalidate this analysis, when the actual job arrival or the actual treatment is known some time beforehand. The former might introduce a downward bias in the estimated treatment effect while the latter might introduce an upward bias. This is a problem in any of the analyses based on the timing-of-events approach. However, it will not be a problem, if individuals anticipate the chances or the determinants of one of these events as long as this occurs in the same way for individuals receiving different treatments or no treatment conditional on X and conditional on the duration of elapsed unemployment $u-1$.

As a further balancing test, we investigate whether treated and matched nontreated individuals should not differ significantly in their employment rates before the beginning of treatment. By construction, treated individuals in the k -group and their matched counterparts in the l -group exhibit the same unemployment duration until the beginning of treatment k . Therefore, we investigate whether they differ in time-invariant unobserved characteristics by analyzing employment differences during eight quarters before the start of the unemployment spell.

4.2 Details of the matching approach

Estimating the ATT for treatment k versus l requires estimating the expected counterfactual outcome for the individuals in group k , had they instead received treatment l in the same or a later quarter during the time window of elapsed unemployment considered. The estimation of the counterfactual outcome is based upon the observed outcomes of the eligible individuals in the l -group. For this, we use a matching approach (Rosenbaum and Rubin (1983), Heckman, Ichimura, Todd (1998), Heckman, LaLonde, Smith (1999), Lechner (2000)) based on the estimated dynamic propensity score, as described in the previous section, and apply local linear matching to estimate the average counterfactual outcome.

4.2.1 Local Linear Regression

Effectively, we run a nonparametric local linear kernel regression (Heckman, Ichimura, Smith, Todd (1998), Pagan and Ullah (1999) Bergemann et al. (2004)) which comprises a weight function $w_{N_l}(i, j)$ that gives the higher weight to individual j in the eligible subset of the l -group, $\{T_{\tilde{u}} = l\}$, the stronger his similarity to participant i in treatment k regarding the estimated propensity score and the starting month of the unemployment spell. Based on a local linear regression of the outcomes in the eligible l -group on the estimated propensity score and the starting month of unemployment, the estimated ATT can be written as

$$(5) \quad \frac{1}{N_k} \sum_{i \in \{T_{u_i} = k\}} \left\{ Y_{i, u_i, \tau}^k - \sum_{j \in \{T_{\tilde{u}} = l, u_i \leq \tilde{u} \leq \tilde{u}\}} w_{N_l}(i, j) Y_{j, \tilde{u}, \tilde{\tau}}^l \right\},$$

where u_i is the quarter in which the treatment for i starts falling in the time window considered. N_l is the number of eligible members of the l -group and N_k the number of participants i in treatment k depending on elapsed unemployment u . $Y_{i, u_i, \tau}^k$ and $Y_{j, \tilde{u}, \tilde{\tau}}^l = Y_j^l(\tilde{u}, \tau - (\tilde{u} - u_i))$ are the outcomes in the post treatment period $u_i + \tau$, where $\tilde{\tau} = \tau - (\tilde{u} - u_i)$.

Kernel matching has a number of advantages compared to nearest neighbor matching, which is widely used in the literature (Lechner 2000, Lechner et al. 2005a,b, Sianesi 2003, 2004). The asymptotic properties of kernel based methods are straightforward to analyze and bootstrapping provides a consistent estimator of the sampling variability of the estimator in (5) even if matching is based on closeness in generated variables (this is the case with the popular method of propensity score matching which will be discussed below), see Heckman, Ichimura, and Todd (1998), Heckman, Ichimura, Smith, and Todd (1998) or Ichimura and Linton (2001) for an asymptotic analysis of kernel based treatment estimators.²³ In contrast, Abadie and Imbens (2006) show that matching estimators based on a fixed number of matches are not root-N consistent and that the bootstrap is in general not valid due to their extreme nonsmoothness.

²³Heckman, Ichimura, and Todd (1998) and Heckman, Ichimura, Smith, and Todd (1998) discuss the asymptotic distribution of estimated treatment effects based on local linear matching taking account of the sampling variability of the estimated propensity score. This asymptotic result justifies the application of bootstrapping.

4.2.2 Kernel Function and Bandwidth Choice

Matching estimators differ with respect to the weights attached to members of the comparison group. In our case, the weights are implied by the nonparametric local linear kernel regression of the outcome in the eligible l -group, $Y_{j,\bar{u},\bar{\tau}}^l$, on the estimated propensity score and the starting month of unemployment, see Heckman, Ichimura, Smith, and Todd (1998). We use a product kernel (see Racine and Li (2004)) in the estimated propensity score and the calendar month of entry into unemployment

$$(6) \quad KK(p, c) = K\left(\frac{p - p_i}{h_p}\right) \cdot h_c^{|c - c_i|}$$

with $K(z) = \exp(-z^2/2)/\sqrt{2\pi}$, i.e. the Gaussian kernel function.²⁴ p_i and c_i are the estimated propensity score and the calendar month of entry into unemployment, respectively, for treated individual i in the k -group, p and c for eligible individuals in the l -group. h_p and h_c are the bandwidths which are determined by the crossvalidation procedure described in the following.

In our case, standard bandwidth choices for pointwise estimation are not advisable because the estimation of the treatment effect is based on the average expected outcome in treatment l for the treated individuals in the k -group. To choose the bandwidths h_p and h_c , we use a two dimensional variant of the leave-one-out cross-validation procedure suggested in Bergemann et al. (2004) and Fitzenberger and Speckesser (2005) mimicking the estimation of the average expected outcome in the alternative treatment l for each period. First, for each participant i in the k -group, we identify the nearest neighbor $nn(i)$ in the eligible subset of the l -group, i.e. the individual in that group whose propensity score is closest to that of i . Second, we choose the bandwidths to minimize the sum of the period-wise squared prediction errors

$$(7) \quad \sum_{\tau=0}^{\tau_{\max}} \left[\frac{1}{N_k} \sum_{i=1}^{N_k} \left(Y_{nn(i),u_i,\tau}^l - \sum_{j \in \{T_{\bar{u}(i)}=l, u_i \leq \bar{u} \leq \bar{u}\} \setminus nn(i)} w_{(N_l(i)-1)}(i, j) Y_{j,\bar{u},\bar{\tau}}^l \right) \right]^2$$

where $\tau_{\max} = 33 - \underline{u}$ and $\underline{u} = 1, 3, 5$ is the first quarter of the time window 1-2, 3-4, 5-8 during which treatment starts. τ counts the number of quarters since quarter u_i , and $N_l(i)$ represents the size of the eligible l -group for i , $\{T_{\bar{u}(i)} = l\}$. The estimation of the employment status for $nn(i)$ is not based on $nn(i)$ itself. However, the individual

²⁴Note that $h_c \in [0, 1]$, where $h_c = 0$ amounts to only considering matches whose unemployment spell starts in the same calendar month.

$nn(i)$ is used for the local linear regression for other treated individuals in the k -group, provided it is in the eligible l -group and unless it happens to be also the nearest neighbor in this case. Therefore, the local linear regression in (7) always depends on $(N_l(i) - 1)$ observations. The optimal bandwidths h_p and h_c affecting the weights $w_{(N_l(i)-1)}(i, j)$ through the local linear regression are determined by a two-dimensional search.²⁵

4.2.3 Bootstrapping

We take account of the sampling variability in the estimated propensity score by bootstrapping the standard errors of the estimated treatment effects. We use partly a parametric bootstrap by resampling the coefficients of the probit estimates for the propensity scores based on their estimated asymptotic distribution. To account for autocorrelation over time, we use the entire time path for each individual as block resampling unit. All the bootstrap results reported in this paper are based on 200 resamples. Since the crossvalidation in (7) is computationally expensive, the sample bandwidths are used in all resamples.

4.2.4 Balancing Test

As a balancing test, we use the regression test suggested in Smith and Todd (2005) to investigate whether the time-invariant (during the unemployment spell) covariates are balanced sufficiently by matching on the estimated propensity score $\hat{P}(X)$. The test involves regressing the respective covariate X_g on a flexible polynomial in $\hat{P}(X)$ with order δ and interactions with the treatment dummy variable

$$(8) \quad X_g = \sum_{d=0}^{\delta} \beta_d \hat{P}(X)^d + \sum_{d=0}^{\delta} \gamma_d D_k \hat{P}(X)^d + \eta_{kl},$$

where X_g is one component in the covariate vector X and $D_k = I(T = k)$ is a dummy variable for treatment k . The regression in (8) is estimated separately based on the sample of those individuals receiving either treatment k or l in the respective interval for unemployment duration (1–2, 3–4, 5–8). If the estimated propensity score balances the covariate X_g in the treatment and the control sample, then the coefficients on all terms involving the treatment dummy, $\gamma_0, \gamma_1, \dots, \gamma_\delta$, should be

²⁵Since the controls consist of the large group of nonparticipants in the respective stratum, it turns out for our application that there is no need to smooth over the calendar month of entry into unemployment as well. In this case, h_c is set to zero and a one dimensional search is performed.

zero. We test this joint hypothesis both for cubic ($\delta = 3$) and quartic ($\delta = 4$) polynomials to see whether the test result is sensitive to the choice of the order of the polynomial, a problem mentioned by Smith and Todd (2005, p. 373). If the test does not reject, then the treatment dummy D_k does not provide any significant information about the covariate X_g conditional on the estimated propensity score. For each specification of the propensity score, we report the number of covariates for which the balancing test passes, i.e. the zero hypothesis is not rejected.²⁶

5 Empirical Results

5.1 Estimation of Propensity Scores

Our empirical analysis is performed separately for the two samples of inflows from employment into unemployment, associated with transfer payment or program participation. To estimate the propensity scores, we run Probit regressions for training starting during the three time intervals for elapsed unemployment duration, i.e. 1–2 quarters (stratum 1), 3–4 quarters (stratum 2), and 5–8 quarters (stratum 3). Instead of estimating a multinomial choice model for entry in one of the three programs or no entry at all for each window of elapsed unemployment duration and sample we estimate a series of binary Probit regressions. In the additional appendix, we report our preferred specifications for the 1986/87 and 1993/94 samples which are obtained after extensive specification search.

The covariates considered are all defined for the beginning of unemployment and are thus time-invariant for an individual during the unemployment spell. Personal characteristics considered are age (as dummies for five-year intervals), dummy variables for gender, marital status, having kids, being a foreigner and formal education (no vocational training degree, vocational training degree, tertiary education degree). In addition, we use information about the last employer, namely industrial sector and firm size dummies, and a number of characteristics of the previous job as employment status and information on earnings in the previous job. Furthermore, we use three variables containing information on earnings. Due to reporting errors and censoring, we do not know the exact earnings for all observations. Therefore, we distinguish the following three cases. First, we use a dummy variable that is equal to one if daily earnings are above 15 Euro (in 1995 Euros), roughly the minimum

²⁶The results of the propensity score estimations are contained in the additional appendix.

level to be subject to social security taxation.²⁷ Second, we have a dummy variable that indicates whether daily earnings are topcoded at the social security taxation threshold (*Beitragsbemessungsgrenze*). Third, we have a variable that records log daily earnings in the range between 15 Euro and the topcoding threshold and zero otherwise.

Regarding the employment and program participation history, we consider the following covariates. We use dummies indicating employment status in month 6, 12, and 24 before the beginning of unemployment. We also consider the number of months in regular employment during five years before the beginning of unemployment. The previous program participation history of an individual is captured by dummy variables that indicate participation in an ALMP program in year(s) 1, 2, and 3–5 before the beginning of unemployment. Differences in regional labor market conditions as well as supply of programs are the reason to include regional variables in the specification. We use the federal state of last employment and the unemployment rate as well as the population density at the district level. Finally, we also use the calendar month of the beginning of the unemployment period, either as a variable counting elapsed months since a given reference date (e.g. January 1960) or as dummies for the respective years and quarters.

Our specification search starts by using as many as possible of the covariates mentioned above without interactions. The specification search is mainly led by the following two criteria: (i) single and joint significance, and (ii) balance of the covariates according to the Smith–Todd (2005) test implemented as described in the last section. As regards the qualitative variables, like state, firm size and industry, which are split up into dummies for the different categories in the regression, we usually test for joint significance. When insignificance is found, the covariates are dropped. Furthermore, we test for the significance of interaction effects, in particular interactions with gender and age. In order to achieve balance of covariates, we test different functional forms (e.g. the square of a variable) and interaction effects. In a few cases, we keep insignificant covariates or interactions if they help to achieve balance. As we find the balancing test to be somewhat sensitive to small cell sizes we occasionally aggregate small groups that have similar coefficients. One example

²⁷Monthly earnings below e.g. DEM 410 in 1986 and DEM 500 in 1992 in West Germany for marginal part–time employees (*geringfügig Beschäftigte*) were not subject to social security taxation and should therefore not be present in the data. In addition, it was possible to earn at most twice as much in at most two months of the year without contributing to the social insurance. Probably due to recording errors, the data shows a number of employment reports with zero or very low earnings. Since this information is not reliable, we only use the information for daily earnings reported above 15 Euro as a conservative cut–off point.

is the aggregation of two federal states.

The results for the Probit estimates in the additional appendix show that the final specifications vary considerably over the inflow cohorts and the three time intervals even keeping the k/l -comparison constant. On the one hand, this emphasizes the necessity to treat all 36 k/l -pairs separately. On the other hand, it makes it impossible to present and discuss all the specifications in detail. In general, the number of covariates decreases with elapsed unemployment duration. This is not surprising because many covariates contain information about the previous job, which should characterize someone in a better way who has only recently become unemployed compared to a long-term unemployed.

Age effects are significant in most estimations. In particular, participants in retraining are younger than individuals in other groups. This reflects the assignment policy of the employment agency. In fact, the very comprehensive and expensive retraining schemes are preferably assigned to individuals who have a long time horizon of working life. Gender effects are also relevant in most estimations, but cannot be easily summarized. In cases where the foreign dummy is significant, it shows that foreigners have a lower probability to participate in any program. The employment history is important in most estimations. Previous participation in an ALMP program is sometimes significant. If so, it increases the probability of another program participation. The industrial sector of the previous job is sometimes significant and the firm size only rarely. In most estimations regional effects and the calendar date of unemployment entry (seasonal effects) are contained.

The results of the balancing tests for each specification are also reported in the additional appendix. Using a cubic in the estimated propensity score the test almost always does not reject for all or for all except one variable in the respective propensity score specification. Only in two out of 36 cases the test rejects two variables. Considering both variants, i.e. the cubic and the quartic in the propensity score, the test does not reject for more than one variable in the specification in 20 out of 36 specifications. Overall, we are confident to have achieved a sufficient degree of balance between treatment and control groups in order for matching on the propensity score to be a valid exercise.

A graphical examination of the common support requirement for estimating the average treatment effect on the treated (ATT) can also be found in the additional appendix. When the control group consists of those not participating in any program, for very small estimated participation probabilities, there is sometimes only a positive density for the nonparticipants. This does not matter since we only es-

timate the ATT for the participants but not the ATT for the nonparticipants (i.e. treatment on the untreated) in this case. Moreover, as we perform kernel matching we need not be concerned about closeness of treated individuals and potential matches because the kernel weights account for this. Overall, we are satisfied with the overlap of support for all k/l -pairs and proceed without restricting the samples.

5.2 Estimated Treatment Effects

The outcome variable is the average of monthly employment dummies in a quarter. We match participants in treatment k and participants in treatment l by their similarity in the estimated propensity scores²⁸ and the starting month of the unemployment spell. For matching, we use only eligible participants in l who are still unemployed in the quarter before treatment starts and we align them by the quarter of elapsed unemployment duration. The ATT is then estimated separately for quarters $\tau = 0, \dots, \tau_{\max}$ since the beginning of program k according to equation (5), where $\tau_{\max} = 31, 27, 25$, respectively, for stratum 1, 2, and 3. The expected counterfactual employment outcome for l is obtained by means of a local linear regression on the propensity score and the starting month of the unemployment spell among the eligible l -group. We obtain an estimate for the variance of the estimated treatment effects through bootstrapping the entire observation vector for a spell in our inflow sample. This way, we take account of possible autocorrelation in the outcome variable. Inference is based on 200 resamples. As a further test of the matching quality, we estimate in the same way the differences between participants and matched nonparticipants during quarters 1 to 8 before the beginning of unemployment. By construction, participants in k and matched eligible members of the l -group are unemployed between the beginning of their unemployment spell and the beginning of the treatment in the k -group.

Figures 2–7 graphically represent the evaluation results. Each figure contains a panel of three times three graphs, where each row represents one pairwise comparison of two treatments and each column represents one of the three intervals of elapsed duration of unemployment at the beginning of the treatment, i.e. 1–2 (stratum 1), 3–4 (stratum 2), or 5–8 (stratum 3) quarters since the start of the unemployment spell. The graphs display the estimated average treatment effect for the treated during quarters 0 to τ_{\max} since the beginning of the treatment and the differences during 8 quarters before the beginning of the unemployment spell. We put pointwise

²⁸We use the fitted index $X_i\beta$ from the Probit estimates.

95%–confidence intervals around the estimated treatment effects. The vertical gap at zero reflects the variable time between beginning of the unemployment spell and the beginning of the treatment.

In order to contrast the initial negative lock-in effects of the programs with the later positive program effects, we calculate the cumulated effects of the program 8, 16, and 24 quarters after the beginning of the program. The cumulated effects are calculated as the sum of the effects depicted in figures 2–7 starting in quarter 0 and summing up over the first 8, 16, and 24 quarters, respectively. Table 6 provides the results. The estimated standard errors are based on the bootstrap covariance estimates for the quarter specific treatment effects.

5.2.1 Training versus Waiting

We first discuss the effects of the three training programs against the alternative of waiting, i.e. no treatment during the time interval (stratum) of elapsed unemployment duration, displayed in figures 2 and 5.

We do not find significant pre-unemployment employment differences in any case. Since all individuals become eventually unemployed, this test for matching quality should focus on the differences during the earlier quarters. There is no evidence of systematic differences in employment rates between treated and associated matched individuals. This indicates that time-invariant unobserved heterogeneity does not invalidate our matching approach.

The results for 1986/87 in figure 2, show positive medium-run (1–3 years) and long-run (4–6 years) post treatment effects of all three training programs after a negative lock-in-effect in the program right after the beginning of treatment. These effects are typically of the magnitude 10 to 20 percentage points (ppoints) and significant. They are smaller and not significant for PF in the second and third stratum. For SPST and RT the medium-run effects lie even above 20 ppoints for strata 2 and 3 and are larger than the long-run effects. As expected, the lock-in-periods are shortest for PF (typically the shortest treatment), lasting at most 3 quarters, and longest for RT, lasting up to two years. SPST lies in between for strata 1 and 2 with a lock-in period of about 1 year and shows a very short lock-in-period of 2 quarters for stratum 3. The positive effects for SPST show similar patterns for the three strata (similar to the results for SPST in Fitzenberger and Speckesser, 2005), with the effects being slightly higher in strata 2 and 3. For RT the positive medium-run effects are larger for strata 2 and 3 compared to stratum 1 and the long-run effects

are larger for stratum 2 compared to both strata 1 and 2.

For the 1993/94 cohort, figure 5 shows similar patterns for training versus waiting. For PF, we find shorter lock-in periods for strata 2 and 3 and small positive but insignificant treatment effects after the lock-in period in stratum 1. For strata 2 and 3, we now find significantly positive medium- and long-run treatment effects of 10 to 15 ppoints. Again, the lock-in period is longer for SPST and even longer for RT. The significantly positive medium- and long-run effects for SPST lie between 10 and 20 ppoints and are more persistent than for the earlier cohort. The positive medium- and long-run effects for RT in stratum 1 are below 10 ppoints and barely significant. The effects are somewhat stronger for strata 2 and 3.

Next, we discuss the cumulated effects of the different programs against the alternative of waiting, which are reported in table 6. This allows for a simple comparison of the ATT effects across programs, though it is important to recall that these effects for the treated cannot be compared because they are based on the separate groups of participants in the different programs. It will be interesting to contrast these effects to the results of the pairwise program comparisons reported in the next subsection.

For the 1986/87 cohort, the cumulated long-run effects after 24 quarters are significantly positive at the 10%-level for all cases, except PF in stratum 3. Overall, SPST shows the largest long-run effects with the highest value of 4.2 for stratum 3, i.e. during the 24 quarters after the beginning of the treatment the treated individuals are employed on average for about 4 quarters more than had they not been treated. For SPST and RT, the long-run effects are higher for later strata, though one can not put a causal interpretation to this because the selection of individuals in the different strata changes. There are less cases with significantly positive cumulated effects after 16 quarters. After 8 quarters, the cumulated effects are still negative for RT due to the longer lock-in period, mostly positive for SPST and PF, and significant in strata 2 and 3 for SPST.

For the 1993/94 cohort, the cumulated long-run effects after 24 quarters are significantly positive in all strata for SPST, for strata 2 and 3 for PF, and for stratum 2 for RT. For SPST, the pattern is similar to the earlier cohort. For PF, the effect is higher for strata 2 and 3 and much lower for stratum 1. Also for RT, the effects are lower and even significantly negative in stratum 1. Early treatments for PF or RT in stratum 1 show worse effects for 1993/94 compared to 1986/87. The effects at 8 and 16 quarters for RT show stronger lock-in effects for the later cohort. For PF in strata 2 and 3, there are stronger positive effects already at 8 and 16 quarters.

Summing up, our results on training versus waiting show that most training programs yield significantly positive and fairly persistent medium- and long-run treatment effects. There are strong lock-in effects, with RT showing the longest lock-in periods (up to 8 quarters). The treatment effects for stratum 1 deteriorate from 1986/87 to 1993/94, which could reflect the stronger targeting in the later years. The cumulated effects are significantly positive for most programs. Overall, SPST seems to show the best results for the treated individuals.

5.2.2 Pairwise Comparisons of Training Programs

As mentioned before, the results in the previous subsection should not be misinterpreted in saying that one program is better than another program because we only reported estimated treatment effects for the treated individuals in one stratum and the selection of treated individuals differs by stratum and program type. Therefore, we now estimate the effects of a treatment k versus the alternative l , where l is one of the other training programs, for the participants in treatment k . This pairwise comparison addresses the issue whether the different programs are well targeted. With individual heterogeneity of treatment effects, it could very well be the case that the participants in SPST fare better on average through participating in SPST as compared to RT even though the participants in RT also fare better on average through participating in RT as compared to SPST. This example is used because we find some evidence for such effects, though they are often not significant.

The quarterly treatment effects for the pairwise comparisons are displayed in figures 3, 4, 6 and 7. After a short description of these effects, our discussion focuses on the cumulated effects in table 6. Note that for the pairwise comparisons, the control groups used for local linear matching are considerably smaller compared to evaluating one training program versus nonparticipation, see tables 2 and 3.

In the majority of cases, we do not find significant pre-unemployment employment differences. In a small number of cases, there are significant (but barely so) employment differences for some quarters before the beginning of unemployment.²⁹ Therefore, we conclude that there are no systematic differences in employment rates left between treated and associated matched individuals.

We find significant short-run treatment effects in a number of cases reflecting the different lock-in periods of the three training programs. RT performs worse than

²⁹These differences in employment history become insignificant, if larger bandwidths are used. Further details are available upon request.

the two other programs during the first two years and PF tends to perform better during the first year. However, we do not find this for all cases. We do not find persistent medium- and long-run effects. In a number of cases, the treatment effects in the medium and long run are significant over a short time period and display quite erratic movements. We prefer to discuss the estimated cumulated effects to assess these effects.

The estimated cumulated effects in table 6 suggest that for the cohort 1986/87 most significant effects are caused by the differential lock-in periods. When comparing SPST with RT for those treated in SPST, we find strong significant positive effects after 8 quarters. When RT is compared with SPST for those treated in RT, we find smaller such effects and the effect is insignificant for stratum 3. When comparing RT with PF, we find analogous effects due to the longer lock-in period for RT but there are only significant for PF versus RT in strata 1 and 3. The medium- and long-run effects at 16 and 24 quarters are rarely significant. For participants in SPST, SPST seems to outperform RT at 16 quarters for strata 1 and 3 but the cumulated effects are reduced at 24 quarters and not significant any more. For participants in RT, SPST seems to outperform RT as well at 16 quarters for stratum 1 but again the effect at 24 quarters is reduced and not significant any more. PF seems to outperform SPST for participants in SPST in stratum 1 after 24 quarters, whereas the cumulated long-run effects are insignificant for participants in PF. The long-run cumulated effects for RT in comparison to PF for participants in RT are positive and sizeable, but not significant. The long-run cumulated effects of PF in comparison to RT are also positive but not significant.

For the cohort 1993/94, the cumulated effects at 8 quarters are qualitatively similar reflecting again the different lock-in periods. Both PF, for all strata, and SPST, for stratum 1, seem to outperform RT in the medium- and long-run for the participants in PF and SPST, respectively. RT is also outperformed in the medium- and long-run by SPST and PF even for participants in RT, though the effects are only strongly significant at 16 quarters (the effects are of similar size at 24 quarters). Comparing SPST and PF, the medium- and long-run effects are not significant but the point estimates suggest both programs outperform the respective other program for the own participants.

Summing up, our results on the pairwise comparisons are much weaker compared to the comparison of training versus waiting, because the standard errors for the pairwise comparisons are much higher. Nevertheless, we can draw some conclusions. The significant cumulated effects after 8 quarters reflect the different lock-in periods

for the three training programs. Most medium- and long-run cumulated effects are insignificant which suggests that in these cases, the employment outcome of the treated individuals could not have been improved on average in the medium- or long-run by reallocating them to a different training program. There is, however, weak evidence for SPST outperforming RT in the medium-run and in the long-run even for the participants in RT, especially for the 1993/94 cohort. Also for the 1993/94 cohort, PF outperforms RT in the medium- and the long-run even for the participants in RT. The point estimates for SPST versus PF suggest for stratum 1 in 1986/87 that the cumulated employment effect would have been better if participants in SPST would have instead participated in PF. For 1993/94, the point estimates suggest that SPST outperforms PF in the medium- and the long-run even for participants in PF. However, none of these effects in 1993/94 are significant.

6 Conclusions

Based on a unique administrative data set, which has only recently been made available, we analyze the long-run employment effects of three types of public sector sponsored training in West Germany, which do not involve a job for the participants. The three types of training are Practice Firm (PF), Retraining (RT), and the Provision of Specific Professional Skills and Techniques (SPST). Specifically, we estimate the average treatment effect on the treated (ATT) against the alternative of nonparticipation in any program as well as for pairwise comparisons among the three programs. We take inflow samples into unemployment for West Germany in 1986/87 and 1993/94. We use the approach for multiple treatment evaluation suggested by Lechner (2001) and Imbens (2000) and apply it to a dynamic setup. Slightly modifying the approach suggested by Sianesi (2003, 2004), we distinguish three types of treatment depending upon the elapsed duration of unemployment when treatment starts, i.e. treatment starts during the first two quarters (stratum 1), during the third and fourth quarter (stratum 2), and in the fifth to eighth quarter (stratum 3). An overall assessment of the microeconomic effects is not possible, because necessary information for a comprehensive cost-benefit-analysis is lacking in our data set.

When comparing treatment against nonparticipation, the estimated treatment effects in almost all cases involve first a lock-in period with negative treatment effects and significantly positive treatment effects in the medium- and long-run. The lock-in period is shortest for PF (at most 2 quarters) and longest for RT (around 2 years).

SPST lies in between with a lock-in period around of 4 to 6 quarters. The treatment effects for stratum 1 deteriorate from 1986/87 to 1993/94, which might reflect stronger targeting of the programs over time. The cumulated effects are significantly positive for most programs. Overall, SPST seems to show the best results for the treated individuals.

The pairwise comparisons of the three treatments, one against another, shows first the differences in the lock-in periods and in most cases insignificant treatment effects in the medium- and long-run. There is, however, weak evidence for SPST and PF outperforming RT in the medium-run and in the long-run, especially for the 1993/94 cohort. For 1993/94, SPST tends to outperform PF, but the effect is not significant. We do not find evidence for business cycle effects on the performance of the programs.

Comparing our results to the study by Lechner et al. (2005a) based on the same data source, our general results for the 1993/94 cohort differ only in some ways from those obtained in their study, though the exact treatment definition, the choice of valid observations, and the employed econometric methods differ substantially from ours. In fact, we find significantly positive effects for all treatments relative to nonparticipation much earlier after the treatment starts, but our results for RT in comparison to other training programs are often negative.

Our study draws a somewhat more positive picture of large scale public sector sponsored training programs compared to the previous literature. However, an overall assessment of the microeconomic effects is not possible since various necessary information for a comprehensive cost-benefit-analysis are lacking in our data. Since the relative performance of SPST tends to improve over time and PF does not seem to dominate the other two programs, our evidence is in contrast to some of the conclusions in the surveys by Martin and Grubb (2001), Kluge and Schmidt (2002), and OECD (2005) advocating a strong on-the-job component for public sector sponsored training to show positive employment effects.

References

- Abadie, and G. Imbens (2006). "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74, 235-267.
- Abbring, J. and G.J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71, 1491-1517.

- Abbring, J. and G.J. van den Berg (2004). "Social Experiments and Instrumental Variables with Duration Outcomes." Unpublished Manuscript, Free University Amsterdam and Tinbergen Institute.
- Bender, S., A. Bergemann, B. Fitzenberger, M. Lechner, R. Miquel, S. Speckesser, C. Wunsch (2005). "Über die Wirksamkeit von Fortbildungs- und Umschulungsmaßnahmen", *Beiträge zur Arbeitsmarkt- und Berufsforschung*, IAB, Nürnberg.
- Bender, S., A. Haas, and C. Klose (2000). "IAB employment subsample 1975–1995", *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 120, 649–662.
- Bender, S. and C. Klose (2000) "Berufliche Weiterbildung für Arbeitslose – ein Weg zurück in die Beschäftigung? Analyse einer Abgängerkohorte des Jahres 1986 aus Maßnahmen der Fortbildung und Umschulung mit der ergänzten IAB–Beschäftigtenstichprobe 1975–1990." *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 33, 421–444.
- Bergemann, A. B. Fitzenberger, and S. Speckesser (2004). "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences." *ZEW Discussion Paper*, Mannheim.
- BLK ([Bund/Länder/Kommission für Bildungsplanung und Forschungsförderung] 2000), *Erstausbildung und Weiterbildung, Materialien zur Bildungsplanung und Forschungsförderung* (83), Bonn: BLK.
- Blasche, D. and E. Nagel (1995). "Beschäftigungssituation von Teilnehmern an AFG-finanzierter beruflicher Weiterbildung," *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 28, 195–213.
- Bundesanstalt für Arbeit (1993, 1997, 2001). *Berufliche Weiterbildung*. Nürnberg: Bundesanstalt für Arbeit (various issues).
- Bundesanstalt für Arbeit (2003), *Geschäftsbericht 2002*, Einundfünfzigster Geschäftsbericht der Bundesanstalt für Arbeit. Nürnberg: Bundesanstalt für Arbeit.
- Fay, R. (1996) "Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD countries." *Labour Market and Social Policy Occasional Papers*, 18, OECD, Paris.
- Fitzenberger, B., A. Osikominu, and R. Völter (2006). "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample." Forthcoming *Schmollers Jahrbuch (Journal of Applied Social Science Studies)*.
- Fitzenberger, B. and S. Speckesser (2005) "Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany." *ZEW Discussion Paper*.
- Fredriksson, P. and P. Johansson (2003) "Program Evaluation and Random Program Starts." Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Working Paper, 2003:1.
- Fredriksson, P. and P. Johansson (2004) "Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Data." *IZA Discussion Paper No. 1062*.

- Gerfin, M. and M. Lechner (2002) Microeconomic Evaluation of the Active Labor Market Policy in Switzerland. *Economic Journal*, 112(482), 854–893.
- Hagen, T. and V. Steiner (2000) “Von der Finanzierung der Arbeitslosigkeit zur Förderung der Arbeit - Analysen und Handlungsempfehlungen zur Arbeitsmarktpolitik.” *ZEW Wirtschaftsanalysen*, 51, Nomos, Baden-Baden.
- Heckman, J. H. Ichimura, and P. Todd (1998). “Matching as an Econometric Evaluation Estimator.” *Review of Economic Studies* 65, 261–294.
- Heckman, J. H. Ichimura, J.A. Smith and P. Todd (1998). “Characterizing Selection Bias using Experimental Data.” *Econometrica* 65, 1017–1098.
- Heckman, J. R.J. LaLonde, and J.A. Smith (1999). “The Economics and Econometrics of Active Labor Market Programs.” In: O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3 A, Amsterdam: Elsevier Science, 1865–2097.
- Ichimura, H. and O. Linton (2001). “Asymptotic Expansions for some Semiparametric Program Evaluation Estimators.” Discussion paper, London School of Economics and University College London.
- Imbens, G. (2000). “The Role of the Propensity Score in Estimating Dose-Response Functions” *Biometrika* 87, 706-710.
- Kluve, J. and C. Schmidt (2002) “Can training and employment subsidies combat European unemployment?”, *Economic Policy*, 441–448.
- Lechner, M. (2000). “An evaluation of public sector sponsored continuous vocational training programs in East Germany.” *Journal of Human Resources*, 35, 347–375.
- Lechner, M. (2001). “Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption.” In: M. Lechner and F. Pfeifer (eds.) (2000), *Econometric Evaluation of Active Labor Market Politics in Europe*, Heidelberg: Physica-Verlag.
- Lechner, M. (2004). “Sequential Matching Estimation of Dynamic Causal Models.” Discussion Paper 2004–06, University of St. Gallen.
- Lechner, M. and R. Miquel (2001). “A Potential Outcome Approach to Dynamic Program Evaluation – Part I: Identification.” Discussion Paper 2001–07, SIAW, University of St. Gallen.
- Lechner, M., R. Miquel, and C. Wunsch (2005a). “Long-Run Effects of Public Sector Sponsored Training in West Germany.” IZA Discussion Paper No. 1443.
- Lechner, M., R. Miquel, and C. Wunsch (2005b). “The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification.” Discussion Paper, University of St. Gallen.
- Martin, J.P. and Grubb, D. (2001) “What works and for whom: A review of OECD countries’ experiences with active labour market policies.” *Swedish Economic Policy Review*, 8, 9–56.
- OECD (2005) “Labour Market Programmes and Activation Strategies: Evaluating the Impacts.” Chapter 4 of *Employment Outlook*, OECD, Paris.

- Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
- Plassmann, G. (2002). “Der Einfluss der Arbeitslosenversicherung auf die Arbeitslosigkeit in Deutschland — eine mikroökonomische und empirische Untersuchung,” *Beiträge zur Arbeitsmarkt- und Berufsforschung*, IAB, Nürnberg.
- Racine, J. and Q. Li (2004). “Nonparametric estimation of regression functions with both categorical and continuous data.” *Journal of Econometrics* 119, 99–130.
- Rosenbaum, P.R. and D.B. Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70, 41–55.
- Roy, A.D. (1951). “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers* 3, 135–146.
- Rubin, D.B. (1974). “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66, 688–701.
- Sianesi, B. (2003) “Differential Effects of Swedish Active Labour Market Programs for Unemployed Adults in the 1990s.” Discussion Paper, Institute for Fiscal Studies, London.
- Sianesi, B. (2004) “An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s.” *Review of Economics and Statistics* 86, 133–155.
- Smith, J.A. and P. Todd (2005). “Rejoinder.” *Journal of Econometrics* 125, 365–375.
- Steffen, J. (2005). “Sozialpolitische Chronik Arbeitslosenversicherung seit 1969”, Arbeitnehmerkammer Bremen, Bremen.
- Wunsch, C. (2006). “Labour Market Policy in Germany: Institutions, Instruments and Reforms since Unification.” Discussion Paper, University of St. Gallen.

Appendix

Descriptive Statistics and Description of Data

Table 1: Participation in Further Training in West Germany until 1997

Year	Annual entries			Annual average stocks
	Total	Further training	Retraining	
1980	232.5	162.4	37.9	89.3
1985	371	298.2	45.1	114.9
1990	514.6	383.4	63.3	167.6
1991	540.6	421.2	70.5	189
1992	574.7	464.5	81.5	180.6
1993	348.1	266	72.2	176.8
1994	306.8	224.9	73.1	177.9
1995	401.6	309.7	81.8	193.3
1996	378.4	291.6	77.3	203.6

Remark: All numbers in thousands. Source: Amtliche Nachrichten der Bundesanstalt für Arbeit, several volumes

Figure 1: Unemployment Rate in West Germany

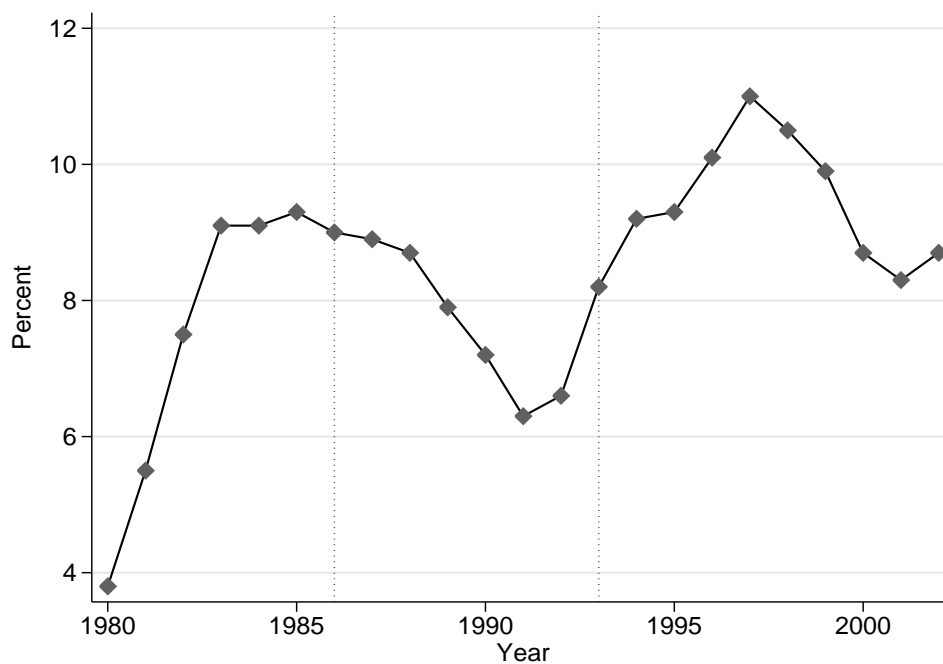


Table 2: Participation in First Training Program for the Inflow Samples into Unemployment

Training Program	Frequency	Percent of inflow sample	Percent among treated
Cohort 86/87			
Practice Firm	246	1.2	14.4
SPST	1,093	5.2	63.8
Retraining	375	1.8	21.9
No training program above	19,188	91.8	–
Total inflow sample	20,902	100	100
Cohort 93/94			
Practice Firm	325	1.3	11.9
SPST	1,944	7.8	71.3
Retraining	458	1.8	16.8
No training program above	22,324	89.1	–
Total inflow sample	25,051	100	100

Remark: Programs that start before a new job is found are considered. We exclude training programs which start together with a job (like integration subsidies) or which involve a very small number of participants since they are not targeted on inflows into unemployment (as career advancement and German language courses). Furthermore, we do not consider the very short programs according to §41a of the Labor Promotion Act, which are only offered to the 1986/87 inflow sample as separate programs, but treat them as open unemployment. This improves the comparability of the inflow samples since comparable very short-term programs offered to the 1993/94 inflow sample are not recorded as programs but as open unemployment in our data. Thus, a participation in retraining after a §41a program is counted as the first program.

Table 3: Number of Training Spells and Length of Unemployment before Program Start

	Cohort 86/87	Cohort 93/94
Practice Firm		
1–2 quarters	74	102
3–4 quarters	60	102
5–8 quarters	69	86
>8 quarters	43	35
Total	246	325
SPST		
1–2 quarters	503	528
3–4 quarters	257	481
5–8 quarters	176	669
>8 quarters	157	266
Total	1,093	1,944
Retraining		
1–2 quarters	172	198
3–4 quarters	101	138
5–8 quarters	71	106
>8 quarters	31	16
Total	375	458

Remark: The time intervals indicate the quarter of program start relative to the beginning of the unemployment spell.

Table 4: Elapsed Duration of Unemployment in Months at Beginning of Training Spell

	Cohort 86/87	Cohort 93/94
Practice Firm		
Average	15.8	11.4
25%–Quantile	5	5
Median	10	9
75%–Quantile	19	15
SPST		
Average	13.3	12.9
25%–Quantile	3	5
Median	6	11
75%–Quantile	14	18
Retraining		
Average	10.2	8.1
25%–Quantile	3	3
Median	6	7
75%–Quantile	12	12

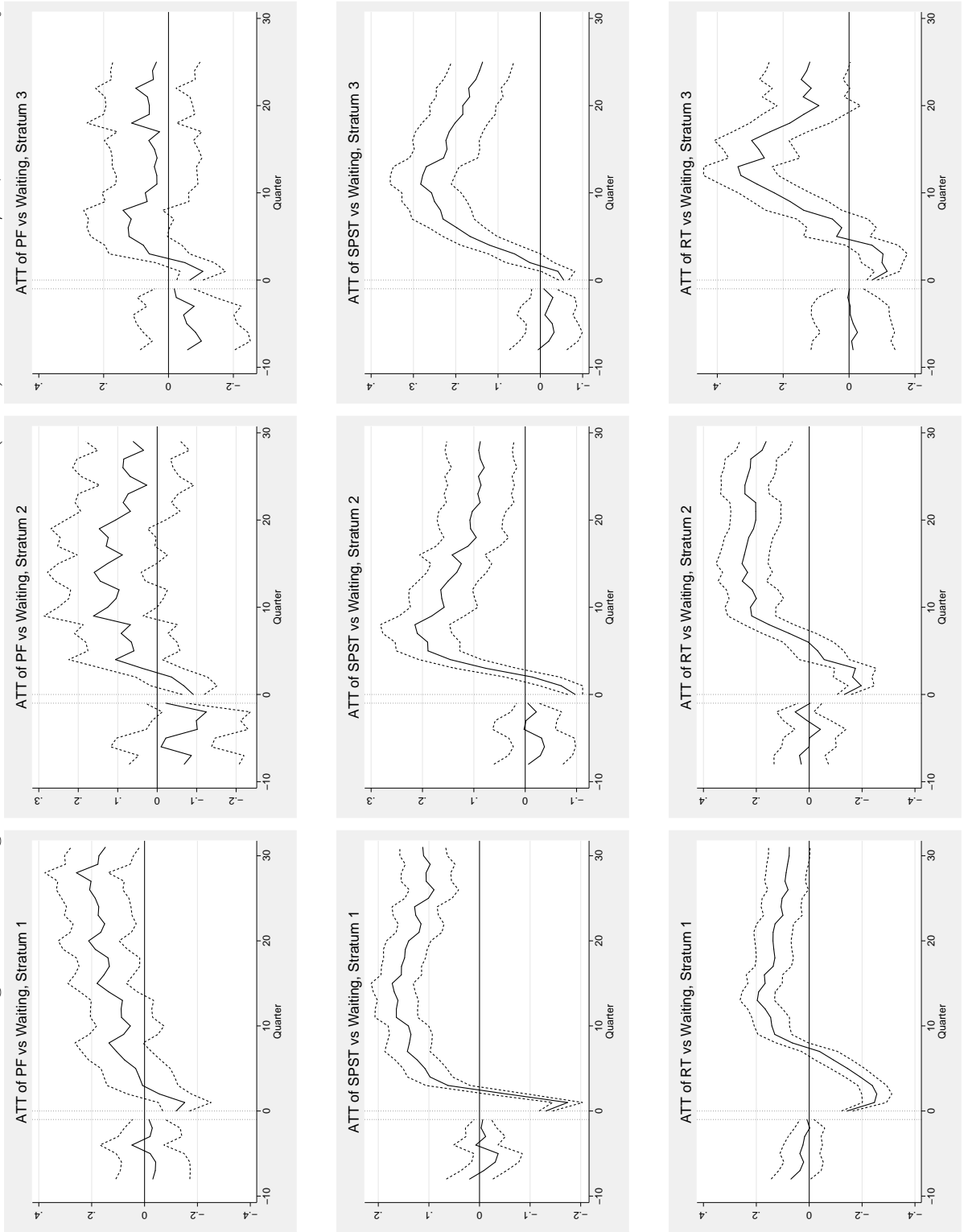
Table 5: Realized Duration of Training Spells in months

	Cohort 86/87	Cohort 93/94
Practice Firm		
Average	5.1	5.7
25%-Quantile	2	3
Median	5	6
75%-Quantile	6	8
SPST		
Average	4.9	6.3
25%-Quantile	2	3
Median	4	6
75%-Quantile	7	8
Retraining		
Average	13.1	14.9
25%-Quantile	5	6
Median	12	16
75%-Quantile	22	21

Remark: The duration of the training spell is defined as the number of months of continuous training. No interruptions are allowed. If in any month we do not identify the program we assume the program has ended the month before.

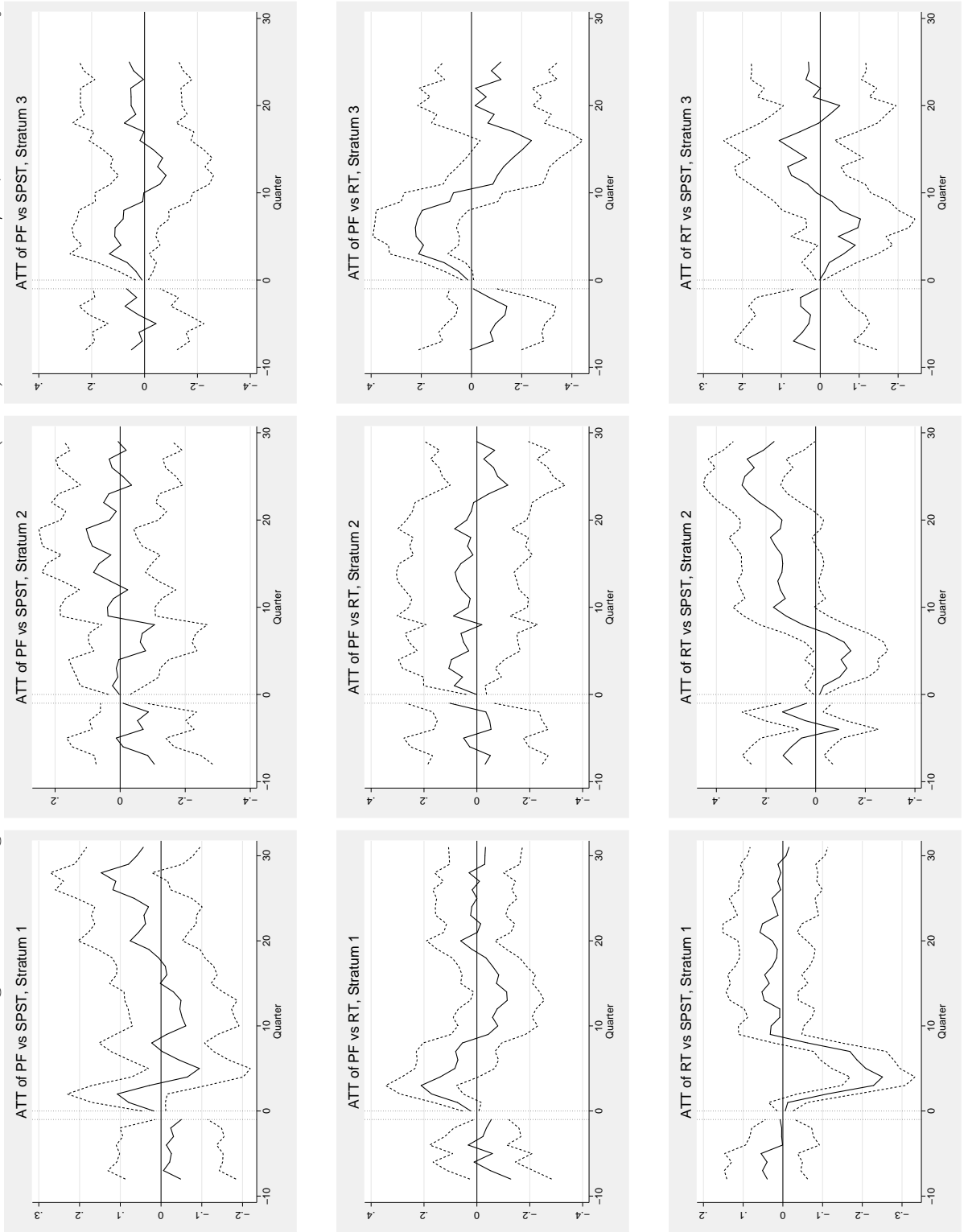
Estimated Employment Effects of Further Training Measures

Figure 2: Average Treatment Effect on the Treated (ATT) for Cohort 86/87, West Germany



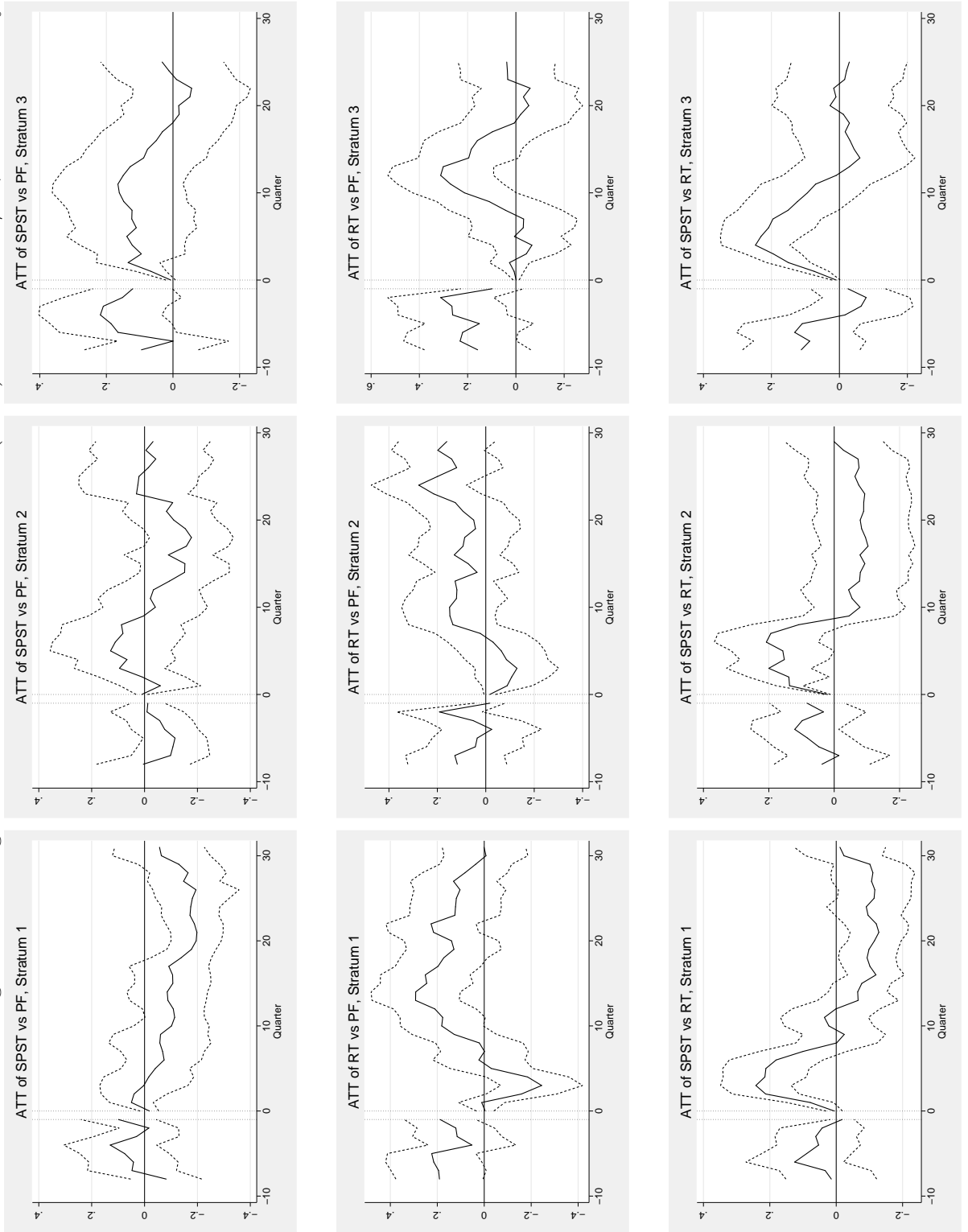
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Figure 3: Average Treatment Effect on the Treated (ATT) for Cohort 86/87, West Germany



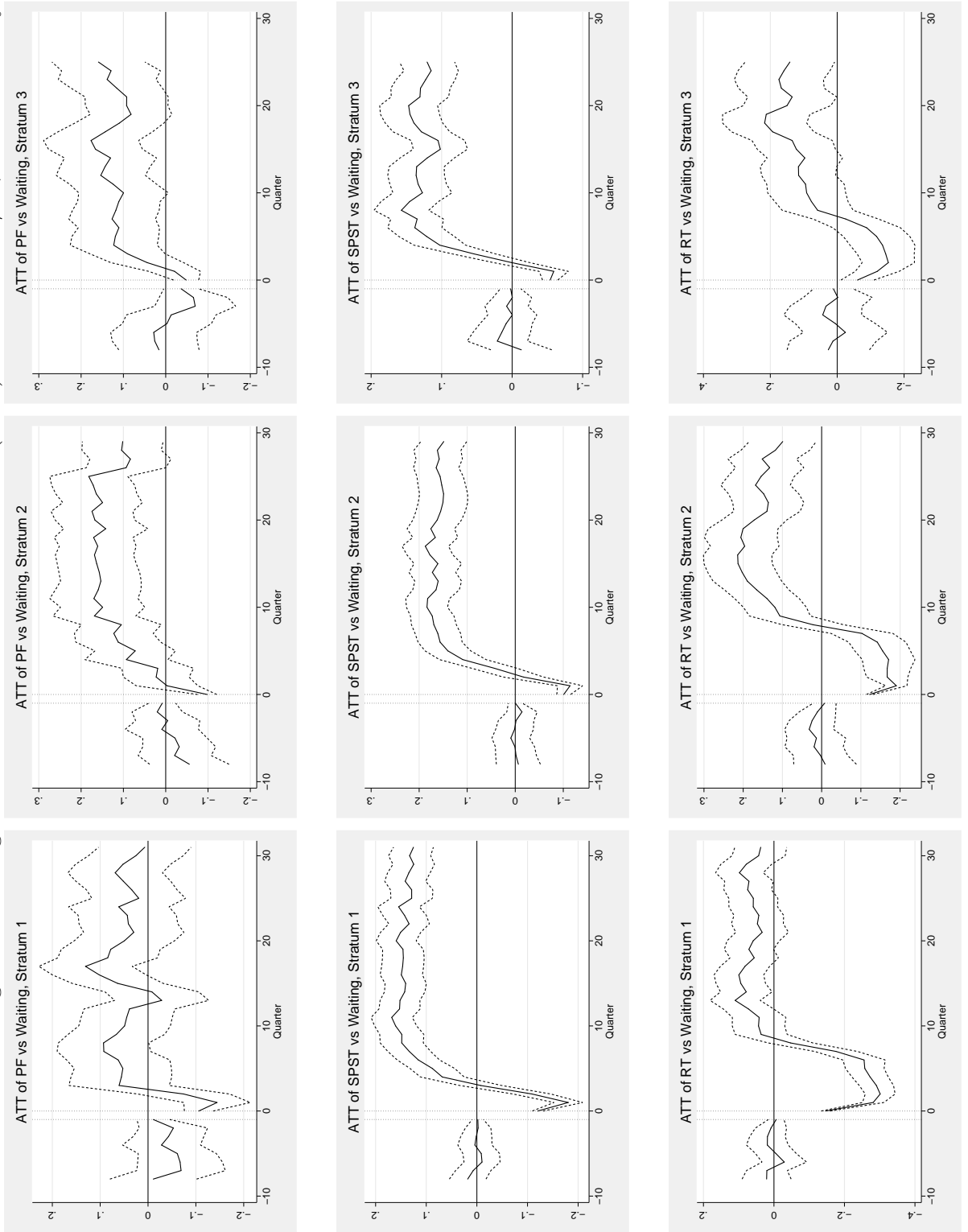
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Figure 4: Average Treatment Effect on the Treated (ATT) for Cohort 86/87, West Germany



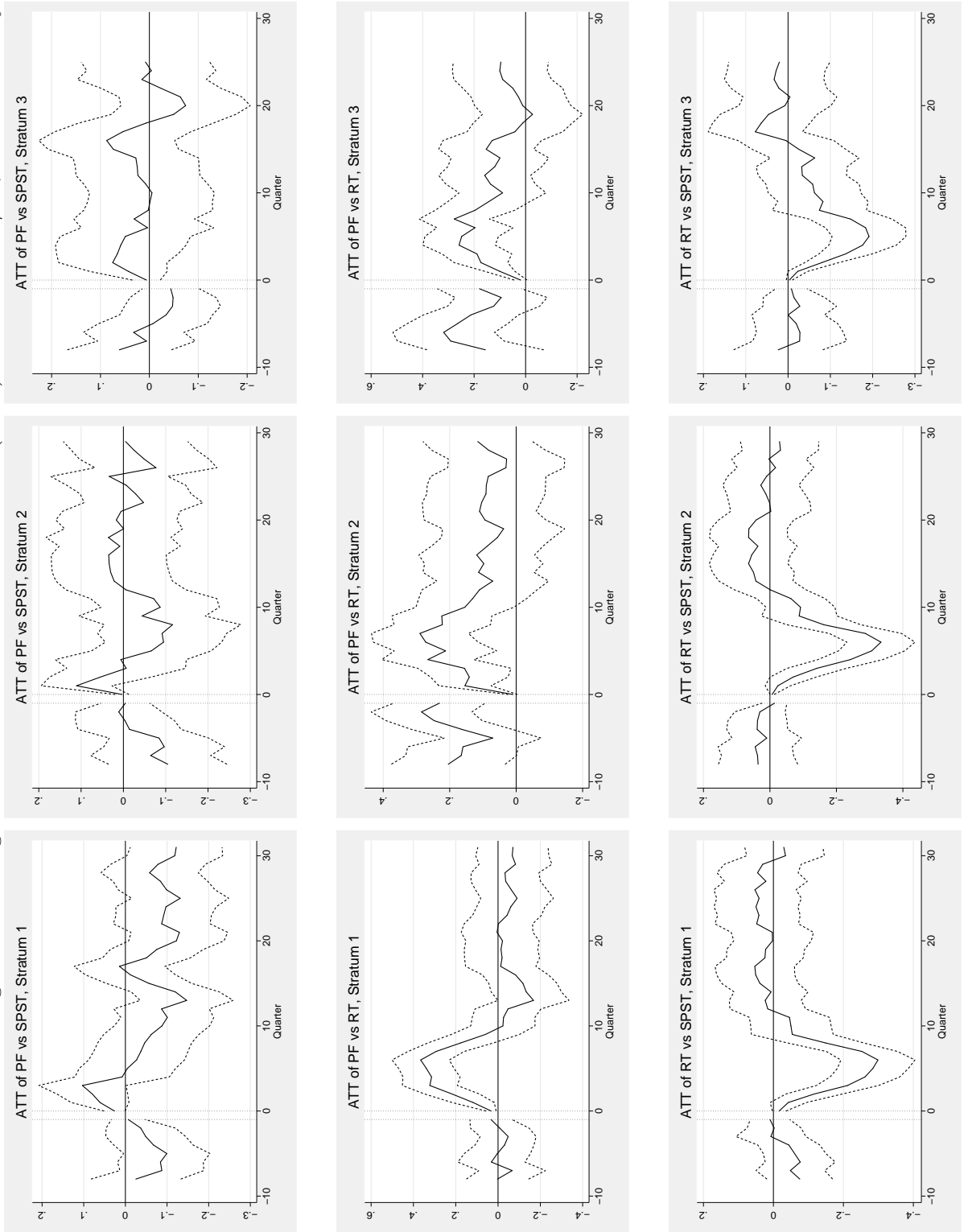
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Figure 5: Average Treatment Effect on the Treated (ATT) for Cohort 93/94, West Germany



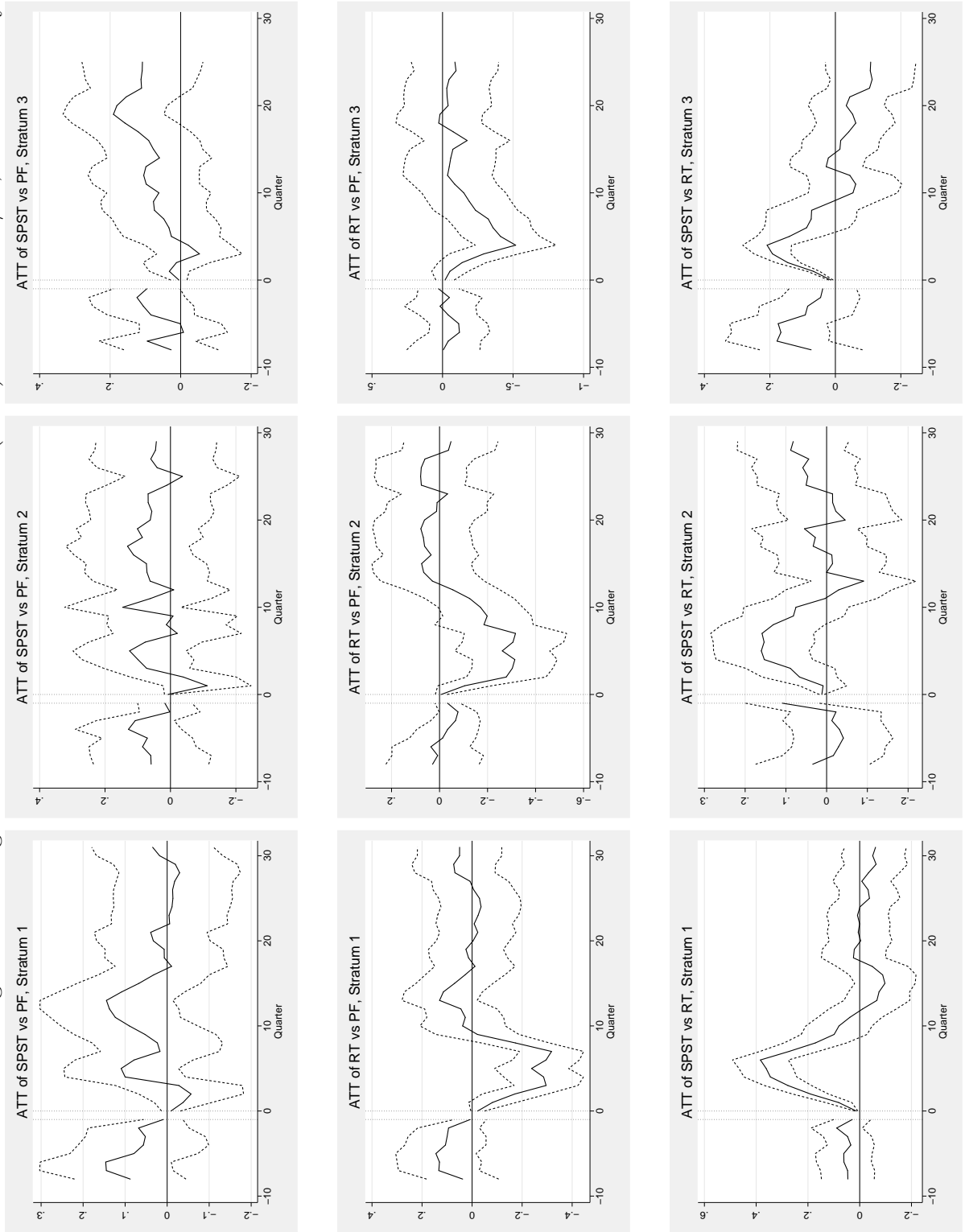
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Figure 6: Average Treatment Effect on the Treated (ATT) for Cohort 93/94, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Figure 7: Average Treatment Effect on the Treated (ATT) for Cohort 93/94, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

Table 6: Cumulated differences in employment rates – sum of quarterspecific average treatment effects on the treated since beginning of treatment

Cumulated Treatment Effects, PF vs Waiting, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-0.085 (0.385)	0.758 (0.706)	2.072 (1.016)**
Stratum 2	0.164 (0.316)	1.150 (0.653)*	1.971 (1.009)*
Stratum 3	0.276 (0.304)	0.748 (0.685)	1.280 (1.115)
Cumulated Treatment Effects, SPST vs Waiting, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	0.174 (0.118)	1.420 (0.241)***	2.524 (0.373)***
Stratum 2	0.631 (0.173)***	1.920 (0.353)***	2.766 (0.536)***
Stratum 3	0.702 (0.173)***	2.725 (0.406)***	4.221 (0.649)***
Cumulated Treatment Effects, RT vs Waiting, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-1.353 (0.169)***	-0.150 (0.326)	0.921 (0.511)*
Stratum 2	-0.678 (0.252)***	1.069 (0.501)**	2.842 (0.761)***
Stratum 3	-0.347 (0.216)	1.673 (0.533)***	3.017 (0.808)***
Cumulated Treatment Effects, PF vs SPST, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	0.028 (0.355)	-0.199 (0.686)	0.023 (1.036)
Stratum 2	-0.159 (0.426)	-0.014 (0.833)	0.431 (1.224)
Stratum 3	0.635 (0.348)*	0.435 (0.876)	0.722 (1.499)
Cumulated Treatment Effects, PF vs RT, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	0.853 (0.395)**	0.348 (0.736)	0.259 (1.117)
Stratum 2	0.485 (0.526)	0.887 (1.165)	1.072 (1.868)
Stratum 3	1.237 (0.350)***	0.907 (0.836)	0.140 (1.402)
Cumulated Treatment Effects, SPST vs PF, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-0.125 (0.339)	-0.848 (0.714)	-2.114 (1.041)**
Stratum 2	0.442 (0.606)	0.039 (1.148)	-0.810 (1.556)
Stratum 3	0.798 (0.406)**	1.837 (1.022)*	1.768 (1.601)
Cumulated Treatment Effects, SPST vs RT, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	1.246 (0.354)***	1.072 (0.599)*	0.199 (0.789)
Stratum 2	1.208 (0.372)***	0.842 (0.708)	0.126 (1.062)
Stratum 3	1.310 (0.286)***	1.625 (0.771)**	1.575 (1.240)

Cumulated Treatment Effects, RT vs PF, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-0.590 (0.476)	0.957 (1.064)	2.413 (1.728)
Stratum 2	-0.496 (0.498)	0.413 (1.022)	1.252 (1.528)
Stratum 3	-0.133 (0.431)	1.498 (1.104)	1.632 (1.654)

Cumulated Treatment Effects, RT vs SPST, Cohort 86/87, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-1.173 (0.227)***	-1.024 (0.440)**	-0.774 (0.698)
Stratum 2	-0.674 (0.376)*	0.354 (0.848)	1.778 (1.345)
Stratum 3	-0.430 (0.269)	-0.207 (0.691)	-0.066 (1.098)

Cumulated Treatment Effects, PF vs Waiting, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-0.001 (0.293)	0.317 (0.606)	0.876 (0.924)
Stratum 2	0.340 (0.235)	1.566 (0.499)***	2.862 (0.744)***
Stratum 3	0.544 (0.276)**	1.590 (0.600)***	2.540 (0.899)***

Cumulated Treatment Effects, SPST vs Waiting, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-0.012 (0.113)	1.201 (0.235)***	2.375 (0.348)***
Stratum 2	0.378 (0.130)***	1.745 (0.266)***	3.070 (0.421)***
Stratum 3	0.439 (0.097)***	1.495 (0.217)***	2.544 (0.338)***

Cumulated Treatment Effects, RT vs Waiting, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-1.982 (0.149)***	-1.552 (0.340)***	-1.061 (0.535)**
Stratum 2	-1.218 (0.192)***	-0.059 (0.395)	1.352 (0.649)**
Stratum 3	-0.845 (0.257)***	-0.074 (0.556)	1.310 (0.905)

Cumulated Treatment Effects, PF vs SPST, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	0.209 (0.282)	-0.498 (0.605)	-1.054 (0.930)
Stratum 2	-0.085 (0.354)	-0.324 (0.741)	-0.300 (1.136)
Stratum 3	0.333 (0.376)	0.485 (0.782)	0.439 (1.165)

Cumulated Treatment Effects, PF vs RT, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	2.002 (0.376)***	1.723 (0.763)**	1.534 (1.234)
Stratum 2	1.500 (0.387)***	2.623 (0.795)***	3.322 (1.296)**
Stratum 3	1.463 (0.355)***	2.559 (0.879)***	2.893 (1.408)**

Cumulated Treatment Effects, SPST vs PF, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	0.174 (0.391)	0.920 (0.824)	1.017 (1.240)
Stratum 2	0.210 (0.366)	0.620 (0.828)	1.306 (1.374)
Stratum 3	0.081 (0.370)	0.733 (0.898)	1.852 (1.378)

Cumulated Treatment Effects, SPST vs RT, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	1.926 (0.263)***	2.065 (0.614)***	1.984 (0.981)**
Stratum 2	0.801 (0.311)**	0.958 (0.616)	0.963 (0.950)
Stratum 3	0.929 (0.215)***	0.886 (0.560)	0.420 (0.860)

Cumulated Treatment Effects, RT vs PF, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-1.707 (0.374)***	-1.477 (0.805)*	-1.481 (1.164)
Stratum 2	-1.890 (0.445)***	-2.453 (1.017)**	-2.158 (1.678)
Stratum 3	-2.112 (0.743)***	-2.988 (1.713)*	-3.341 (2.694)

Cumulated Treatment Effects, RT vs SPST, Cohort 93/94, West Germany			
	8 quarters	16 quarters	24 quarters
Stratum 1	-1.485 (0.257)***	-1.698 (0.540)***	-1.453 (0.848)*
Stratum 2	-1.411 (0.250)***	-1.661 (0.536)***	-1.389 (0.869)
Stratum 3	-0.940 (0.201)***	-1.372 (0.519)***	-1.122 (0.825)

Remark: *, **, and *** denote significance at the 10%-, 5%-, and 1%-significance level, respectively.