Discussion Paper No. 06-039

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

Bernd Fitzenberger, Aderonke Osikominu
and Robert Völter

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 06-039

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

Bernd Fitzenberger, Aderonke Osikominu
and Robert Völter

Non–technical Summary

Public sector sponsored training has traditionally been a main part of active labor market policy (ALMP) in many countries. As part of ALMP, countries like Germany have implemented large scale training programs. It is often argued, that long–term public sector sponsored training programs show little or negative short–run employment effects and often it is not possible to assess whether positive long–run effects exist. For Germany, appropriate data for an evaluation of the long–term effects of public sector sponsored training were not available for a long time.

Based on unique administrative data, which have only recently become available, this paper estimates the long–run differential employment effects of three different types of training programs in West Germany. 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). Typically, RT involves an up to two–year program providing complete vocational training in a new occupation and lasts longer than an SPST program. PF involves training in a work environment simulating a real job. PF tends to be a slightly shorter treatment than SPST.

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 employment effects of training programs starting during 1 to 2, 3 to 4, and 5 to 8 quarters of unemployment. 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.

When comparing treatment against nonparticipation in any training program, 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 cumulated effects are significantly positive for most programs. Overall, against the alternative of nonparticipation, SPST seems to show the best results for the treated individuals.

The pairwise comparisons of the three treatments, one against another, show first the differences in the lock–in periods and in the medium– and long–run mostly insignificant treatment effects. In some cases for the 1993/94 inflows into unemployment, SPST seems to outperform RT.

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

Bernd Fitzenberger[†], Aderonke Osikominu[‡], Robert Völter[§]

This version: October 2007
First version: April 2006

*We are grateful for very helpful comments by two anonymous referees. We thank Reinhard Hujer, Peter Mueser, and Stefan Speckesser for helpful discussions as well as Stefan Bender for support in obtaining the data used here. This study is part of the project "On the effectiveness of further training programs. An evaluation based on register data provided by the Institute for 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 was presented at the Evaluation Conference at Paris, December 2005, at the Rhein-Main-Neckar Labor Market Seminar at Mannheim, March 2006, at the ESEM in Vienna, August 2006, and at the 3rd Conference on Policy Evaluation at ZEW, October 2006. We thank participants for helpful comments. All errors are our sole responsibility.

[†]Albert Ludwigs–University Freiburg, ZEW, IZA, IFS. Address: Dept. of Economics, Albert Ludwigs–University Freiburg, 79085 Freiburg, Germany. E–mail: bernd.fitzenberger@vwl.uni-freiburg.de

[‡]Albert Ludwigs-University Freiburg. E–mail: aderonke.osikominu@vwl.uni-freiburg.de

[§]CDSE University of Mannheim. E–mail: voelter@econ.uni-mannheim.de.

Abstract: 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 programs 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.

Keywords: multiple treatments, training programs, employment effects, local linear matching, administrative data, active labor market programs
JEL: C 14, J 68, H 43

Contents

1 Introduction

Public sector sponsored training has traditionally been a main part of active labor market policy (ALMP) in many countries like Germany. 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 the 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. Our paper adds to this literature by estimating the long–run employment effects of three 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, 2005), Fredriksson and Johansson (2003, 2004), and Sianesi (2003, 2004). Static treatment evaluations implicitly condition 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 studies[1] 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 the precision of the employment history before and after treatment. The sample 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 courses with a median duration between 4 and 6 months. The two other training programs are working in a Practice Firm (PF) and Retraining (RT). RT involves a long program, which lasts up to 2 years and which provides complete vocational training in a new occupation. PF involves training in a work environment simulating a real job and is of similar length as SPST. This classification of treatments is developed in this paper and in the earlier paper, Fitzenberger and Speckesser (2007). The three training programs considered here differ both in length and content. PF has the strongest on–the–job component, SPST involves typically off–the–job classroom training and RT involves both

[1] See 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.

[2] Notable exceptions are the recent studies of Lechner et al. (2005a,b) and Fitzenberger and Speckesser (2007), 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).

on–the–job and off–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 faced very different labor market prospects due to changing business cycle conditions and the impact of German unification. It is of interest to investigate whether the effects of ALMP differ by the state of the labor market. 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 which says that for treated and nontreated individuals the expected potential employment outcome in case of not receiving the treatment of interest (which is observed for the nontreated and counterfactual for the treated) is the same conditional on a set of observable covariates. In our case, these covariates involve socio–economic characteristics, previous employment history, beginning of unemployment, and 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 depending upon elapsed duration of unemployment, a static evaluation approach does not seem appropriate. 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. In fact, kernel matching has a number of advantages compared to nearest neighbor matching (Heckman, Ichimura, Smith and Todd (1998), Ichimura and Linton (2001), Abadie and Imbens (2006)), which is widely used in the literature (e.g. Lechner et al. (2005a,b), Sianesi (2003, 2004)). 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 (2007) 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.

Comparing our work to the study by Lechner et al. (2005a), who also estimate multiple treatment effects based on the same data, there are the following notable differences. First, Lechner et al. only consider a sample for the 1990s, while we also consider an inflow sample from the 1980s. Second, their study uses a different approach regarding the construction of the sample and the choice of valid observations. The definition of treatment types and the identification of treatments from the data differ as well. Third, taking into account the dynamic assignment into programs our paper also comprises an important methodological difference compared to Lechner et al. (2005a) who apply a static approach. Nonetheless, as far as comparable, the results are quite similar in most cases.

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 from our web page, 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 and to those at severe risk of becoming unemployed. 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 vocational training, retraining, and integration subsidy. In addition, there is short–term training which only existed from 1979 to 1992 (§41a AFG). 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 vocational 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 the main programs:[3]

- **Further vocational training** (*Berufliche Fortbildung*) includes the assessment, maintenance and extension of skills, including technical development and career advancement. The duration of the courses depends on the participants' individual predispositions, other co–financing institutions and on the supply of adequate training courses by the training providers.
- **Retraining** (*Umschulung*) enables vocational re–orientation if a completed vocational training does not lead to adequate employment. Retraining is

[3] The complete list with descriptions of the different training schemes that are regulated by the Labor Promotion Act can be found in the additional appendix.

supported for a period up to 2 years and aims at providing a new certified vocational education degree.

- The program **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 (up to 50% of the standard wage in the respective occupation) is paid for an adjustment period until the supported person reaches full proficiency in the job.

- 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. This type of training was abolished in 1992.

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 a job subject to social insurance contributions or they must be entitled to unemployment benefits or subsequent unemployment assistance.[4]

Starting 1986 until 1993, the income maintenance amounted to 73% of the relevant previous net earnings 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).[5] In 1994, income maintenance and unemployment benefits were both cut back to a common level of 67% (with children) and 60% (without children), reducing the financial incentives to

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

[5]In 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).

join a training program. Unemployment assistance was also lowered to 57% (with children) and 53% (without children).[6]

IM recipiency during a training program 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 program. Additionally, participants in training programs could requalify for unemployment benefits providing additional incentives to participate.[7]

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 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

Participation in further training programs in West Germany is large, see table 2. In 1980, 247,000 participants enter such programs. In the late 1980s and early 1990s, annual entries peak at almost 600,000 and then decline to 378,400 in 1996. Among the three main types of training programs distinguished by German legislation, the general further vocational training schemes traditionally are the most important in West Germany with about 70–80% of the entries. Roughly 20% enter retraining. The small remaining share are integration subsidies.

3 Data and Treatment Types

We use a database which integrates administrative individual data from three different sources (see Bender et al. (2005) for a detailed description). The data contain

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

[7]This 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).

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.

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 employment register data for all employees who are covered by the social security system over the period 1975–97. Because sampling is based on employment, we restrict the analysis to inflows from employment to unemployment.[8] For this study, we obtained additional information from the IAB for 1998–2002, which we merged to the basic data.

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). 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.

The third data source are the data on training participation (FuU-data). The employment agency collects these data for all participants in further training, retraining, and other training programs for internal monitoring and statistical purposes, see Bender et al. (2005, chapter 2.3). For every participant, the FuU-data contains detailed information about the program and about the participant.

The FuU–data were merged with the combined IABS–LED data by social insurance number and additional covariates. Numerous corrections were implemented in order to improve the quality of the data, see Bender et al. (2005, chapters 3–4) and the additional appendix for details. The IABS provides information on personal characteristics and employment histories. The combination of the transfer payment information and the participation information is used to identify the likely participation status regarding the different types of training programs.

[8]Restricting the analysis to inflows from employment, we exclude program participants who have not been employed before registering as unemployed. This restriction of the data only concerns a small share of training participants. Statistics from the BA show that the share of those who have never worked or have not worked within the last 6 years before they enter a program is between 4 and 8 percent. Because of the sample design, it would be impossible to construct an appropriate control group for such participants.

3.1 Evaluated Programs

We evaluate the three quantitatively important training programs targeted at the unemployed. Further vocational training is a very broad legal category and consists of quite heterogeneous programs. Hence we utilize a classification developed in Fitzenberger and Speckesser (2007) and evaluate two specific further vocational training programs: Practice Firms (PF) and provision of specific professional skills and techniques (SPST). We also evaluate Retraining (RT). We do not evaluate short–term training, since this program only existed until 1992.[9] We also do not evaluate integration subsidies, because they condition on having found a job involving a potentially difficult endogeneity issue.[10] Next, we describe the evaluated programs. Table 1 gives an overview of the evaluated programs.

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 the provision of new professional skills is less important. Some of the practice firms are technically oriented, the practice studios, whereas others are commercially oriented, the practice enterprises. One of the practice firms' 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.

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 mainly consists of classroom training but an acquisition of professional knowledge through practical work 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 sustained reintegration

[9] After 1992 comparable programs were offered, but they are not recorded in the data. In order to analyze both inflow samples in a comparable way we ignored information about short–term training in the eighties.

[10] A detailed description of how the different treatment types are identified from the data is given in the additional appendix.

Table 1: Overview of Evaluated Programs

	PF	SPST	RT
Name	Practice Firm	Provision of Specific Skills and Techniques	Retraining
Description	training on the job in a simulated firm	classroom courses	complete new vocational training
Orientation	practical	theoretical	practical and theoretical
Median duration 86/87 (93/94)	5 (6) months	4 (6) months	12 (16) months

into the labor market by improving skills as well as providing signals.

Compared to retraining, which is a far more formal and thorough training on a range of professional skills and which provides a complete vocational training degree, SPST provides a smaller, specific addition to the occupational knowledge. However, this addition certainly exceeds the level provided in short–term programs (not evaluated here) 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 and very informal and short courses (improving general human capital).

Retraining consists of the provision of a new and comprehensive vocational training according to the regulation of the German apprenticeship system. It is targeted to individuals who had already completed a vocational training degree and face severe difficulties in finding a new job within their profession. It might however also be offered to individuals without a first formal training degree if they fulfill additional eligibility criteria.

Retraining provides widely accepted formal certificates. 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 certified job qualification in order to improve the job match. Ideally, the training occupation in retraining corresponds to qualifications which are in high demand in the labor market.

3.2 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 indicate a program participation before the unemployed individual finds a new job.[11] 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 exiting from 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 starts training before possibly exiting to employment. In our monthly data, this means that the individual should still be recorded as nonemployed in the month when treatment starts. 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 93/94 and 86/87 to allow for a comparison between the 1980s and the 1990s. Figure 1 depicts the unemployment rate in West Germany. The dotted vertical lines mark the years 1986 and 1993, respectively. Whereas 86/87 mark the end of a sequence of years with relatively high unemployment, the cohort 93/94 enters during a period with increasing unemployment rates. Thus, the 86/87 cohort faced a fairly favorable labor market in the years to come culminating in the unification boom in West Germany, while the 93/94 cohort entered unemployment during one of the most severe recessions in West Germany resulting in a prolonged period with bad labor market chances. Our data allow to follow individuals entering unemployment in 86/87 until December 1996/97 and individuals entering unemployment in 93/94 until the end of 2001/02.

Table 3 gives information about the size of the inflow samples and the incidence of

[11] We allow the same individual to appear in the sample more than once if he or she exhibits more than one transition from employment to unemployment during the relevant time period.

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 86/87 cohort and 25,051 spells for the 93/94 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 by far the 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 4 shows the frequency of training by time window of elapsed unemployment.

Table 5 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 the 86/87 sample and 9 months in the 93/94 sample. Late starts (75%–quantile) of PF occur after 19 months in the 86/87 sample and much earlier in the 93/94 sample. For RT, the quantiles in the samples are very similar. With a median of 6 and 7 months, RT starts the earliest. For SPST, we find a reversed trend in comparison to PF. While SPST participation starts almost as early as RT in the 86/87 sample, the starting dates are noticeably later in 93/94, with the median increasing from 6 to 11 months.

Table 6 provides descriptive information on the realized duration of training spells. The average duration of practice firm is similar in both samples with 5.1 months in 86/87 and 5.7 months in 93/94. SPST has an average duration of 4.9 months for the 86/87 sample and 6.3 months for 93/94. Retraining is by far the longest program. It lasts on average for 13.1 months in the 86/87 sample and 14.9 months in the 93/94 sample. Note that some participants drop out of the programs early. So the realized durations can be shorter than the planned durations. In our samples about 70% of the participants complete the programs, 10% drop out because they have found a job and 20% drop out for other reasons. Our analysis does not condition on program completion since program dropout is likely to be endogenous. So, strictly speaking, the programs we evaluate are the starts of the respective programs, as is common in most of the recent literature.

The final question about the samples which we want to discuss is the incidence of other programs. Our basic approach is to ignore the (relatively rare) participation in other programs and classify such spells as spells without program participation. For 86/87, 1.2% of the participants in evaluated programs participated in short–term training before starting an evaluated program. Also 1.2% of the nonparticipants took part in short–term training during their defining unemployment spell.[12] For 93/94, there existed no short–term training. The share of nonparticipants in the evaluated programs who took part in another, not evaluated further vocational training program[13] is 0.5% in the 86/87 sample and 0.3% in the 93/94 sample. Integration subsidies are paid to 1.2% of the nonparticipants in the 86/87 sample. Only 0.3% of the training participants in the 86/87 sample finish their unemployment spell with a subsidized job. And in the 93/94 sample among both, the treated and the controls, this share is 0.3% or lower. Concluding, we argue that participation in other, not evaluated training programs is small enough to be neglected in our analysis.

4 Evaluation Approach

Our goal is to analyze the effect of $K = 3$ different training programs on the quarterly employment rate at the individual level, which is measured as an average of three monthly employment dummy variables.[14] In a situation where individuals have multiple treatment options, we estimate the average treatment effect on the treated (ATT) of one training program against nonparticipation in any of the three programs and of pairwise comparisons of two programs. Extending the static multiple treatment approach to a dynamic setting, we follow Sianesi (2003, 2004) and apply the standard static treatment approach recursively depending on the elapsed unemployment duration. The implementation builds upon the approach for binary treatment in Fitzenberger and Speckesser (2007). We first present our evaluation approach and then compare it to recent alternative proposals in the literature.

[12] As we do with the evaluated programs we look at program participation during the defining unemployment spell and in the case of integration subsidies at payment of a subsidy for the first job after the defining unemployment spell.

[13] These other programs are mainly career advancement programs targeted at the employed.

[14] The quarterly employment rate can take the four values 0, 1/3, 2/3, and 1.

4.1 Multiple Treatments in a Dynamic Context

Our empirical analysis is based upon the potential–outcome–approach to causality, see Roy (1951), Rubin (1974), and the survey of Heckman et al. (1999). Lechner (2001) and Imbens (2000) extend this framework to allow for multiple, mutually exclusive treatments. Let the 4 potential outcomes be $\{Y^0, Y^1, Y^2, Y^3\}$, where $Y^k, k = 1, ..., 3$, represents the outcome associated with training program k and Y^0 is the outcome when participating in none of the 3 training programs. For each individual, only one of the $K + 1$ potential outcomes is observed and the remaining K outcomes are counterfactual. We estimate the average treatment effect on the treated (ATT) of participating in treatment $k = 1, 2, 3$ against nonparticipation $k = 0$ (treatment versus waiting) and the differential effects of the programs (program k versus program l where $k, l \neq 0$), see Lechner (2001).

Fredriksson and Johansson (2003, 2004) argue that a static evaluation analysis, which assigns unemployed individuals to a treatment group and a nontreatment group based on the treatment information observed in the data, yields biased treatment effects. This is because the definition of the control group conditions on future outcomes or future treatment. For Sweden, Sianesi (2004) argues 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 Germany). In former West Germany, active labor market programs were implemented at a fairly large scale in international comparison. This discussion implies that a purely static evaluation of the different training programs is not warranted. Following Sianesi (2003, 2004), we analyze the effects of the first participation in a training program during the unemployment spell considered *conditional on the starting date of the treatment.* 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 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 ATT parameter (comparing treatments k and l) of interest is

$$\text{(1)} \qquad \theta(k, l; u, \tau) = E(Y^k(u, \tau)|T_u = k, U \geq u-1, T_1 = ... = T_{u-1} = 0)$$

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

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 $\tau = 0, 1, 2, ...,$ counts the quarters since the beginning of treatment. When $l = 0$, we compare treatment k versus waiting (nonparticipation in the stratum) and when $l \geq 1$, we do a pairwise comparison between treatment k and l. 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 $\bar{u}$. 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.[15]

The treatment parameter we actually estimate is the average within a stratum

$$\theta(k, l; \tau) = \sum_u g_u \theta(k, l; u, \tau) ,$$

with respect to the distribution g_u of starting dates u within the stratum.

Our estimated treatment parameter (1) mirrors the decision problem of the case worker and the unemployed who recurrently during the unemployment spell decide whether to start any of the programs now or to postpone participation to the future.

We evaluate the differential effects of multiple treatments assuming the following dynamic version of the conditional mean independence assumption (DCIA)

$$(2) \quad E(Y^l(\tilde{u}, \tau - (\tilde{u} - u))|T_u = k, u \leq \tilde{u} \leq \bar{u}, U \geq u{-}1, T_1 = ... = T_{u-1} = 0, X)$$

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

[15] Admittedly, the notation in equation (1) is cumbersome because we do not follow Sianesi (2003) and allow the alternative treatment l to start only in the same quarter u as treatment k. The problem is that one has to decide how to assign individuals who receive an alternative treatment later in the stratum considered. We think that program participation $l \in \{1, 2, 3\}$ in a later quarter $\tilde{u} > u$ (with $\tilde{u} < \bar{u}$) should not be interpreted as no participation (treatment 0) but rather we suggest to add such a case to the l–alternative for treatment k in quarter u. This is reflected in the definition of the parameter of interest.

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

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 that the conditional probability of treatment k, given that the individual receives treatment k or treatment l, $P^{k|kl}(X)$, exhibits an analogous balancing property for the pairwise estimation of the ATT's of program k versus l. This allows to apply standard binary propensity score matching based on the sample of individuals participating in either program k or in program l. For this subsample, we simply estimate the probability of treatment k and then apply a bivariate extension of standard propensity matching techniques. Implicitly, we assume that the actual beginning of treatment within a stratum is random conditional on X.

To account for the dynamic treatment assignment, we estimate the probability of treatment k 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 (stratum 1), and estimate a Probit model for participation in k. This group includes those unemployed who either never participate in any program or who start some treatment after quarter 2. For treatment during strata 2 and 3, the basic sample consists of those unemployed who are still unemployed in the first month of the stratum.

We implement a stratified local linear matching approach by imposing that the matching partners for an individual receiving treatment k are still unemployed in the quarter before treatment k starts, i.e. we exactly align treated and nontreated individuals by elapsed unemployment duration in quarters. The expected counterfactual employment outcome for nonparticipation is obtained by means of a bivariate local linear regression on the propensity score and the starting month of the unemployment spell. We use a bivariate crossvalidation procedure to obtain the bandwidths in both dimensions (propensity score and beginning of unemployment spell) minimizing the squared prediction error for the average of the l–outcome for the nearest neighbors of the participants in program k.[16] An estimate for the variance

[16] This method is an extension of the crossvalidation procedure suggested in Bergemann et al.

of the estimated treatment effects is obtained through bootstrapping based on 200 resamples. This way, we take account of the sampling variability in the estimated propensity score.

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 $P^{k|kl}(X)$ using a flexible polynomial approximation. For each specification of the propensity score, the additional appendix reports the number of covariates for which the balancing test passes, i.e. the zero hypothesis is not rejected. Furthermore, we investigate whether treated and matched nontreated individuals differ significantly in their outcomes before the beginning of unemployment, in addition to those variables already used as arguments of the propensity score. We estimate these differences in the same way as the treatment effects after the beginning of the program. By construction, treated individuals and their matched counterparts exhibit the same unemployment duration until the beginning of treatment.

Finally, we need to discuss the plausibility of the DCIA (2) for our application. As Sianesi (2004), we argue that the participation probability depends upon the variables determining re-employment prospects once unemployment began. Consequently, all individuals are considered who have left employment in the same two years (matching controls for beginning of unemployment) and who have experienced the same unemployment duration before program participation. Furthermore, observable individual characteristics and information from the previous employment spell have been included in the propensity score estimation. E.g., we consider skill information, regional information, occupational status, and industry which should be crucial for re-employment chances. Unfortunately, our data lack subjective assessments of labor market chances of the unemployed (e.g. by case workers). We argue that these are proxied sufficiently by the observed covariates in so far as they affect selection into the program. This is particularly plausible, since participation occurred at a fairly large scale, assignment was not very targeted and driven by the supply of programs, and case workers had little guidance on 'what works for whom'. Supporting our point of view, Schneider et al. (2006) argue that until 2002 assignment to training was strongly driven by the supply of available courses.[17]

(2004) and also used in Fitzenberger and Speckesser (2007). A detailed description of the implementation of the two dimensional bandwidth search can be found in Fitzenberger, Osikominu and Völter (2006).

[17]For the evaluation of the employment effects of job creation schemes in 1999/2000 based on administrative data for Germany, Caliendo et al. (2004) were able to use a survey asking about the motivation of participants (such information is not available for our data). It turned out that

4.2 Comparison to Alternative Dynamic Approaches

Abbring and van den Berg (2003), Lechner and Miquel (2005) as well as Lechner (2004), and Heckmann and Navarro (2007) propose three important alternative approaches to estimate treatment effects in a dynamic context. We now compare our approach building on Sianesi (2004), as described in the previous section, to these approaches. Under stronger assumptions than we are willing to make for our analysis, all three alternative approaches would allow to estimate more comprehensive treatment effects than estimated in this paper.

The timing–of–events approach by Abbring and van den Berg (2003) uses a continuous time duration model with unobserved heterogeneity, where time until treatment start and unemployment duration constitute two competing risks. The goal is to estimate the causal treatment effect on the hazard to leave unemployment. Identification of the causal effect of entering a program relies on the conditional randomness of program starts and a non–anticipation condition as well as functional form assumptions involving e.g. a mixed proportional hazard model and a tight specification of the joint dependence between duration until treatment and the outcome variable unemployment duration. Our approach also considers the variation of starting dates during the unemployment spell, but relying on a selection on observables strategy, we estimate flexible discrete time hazards into the program where covariates are fully interacted with the elapsed duration of unemployment. By conditioning on elapsed unemployment duration by strata, we account for the endogenous selectivity of the group of individuals eligible for treatment at different points of time. In contrast to Abbring and van den Berg (2003), we allow for heterogeneity of treatment effects and our outcome variable employment is static. The two approaches have in common that the estimated treatment parameter corresponds to the effect of starting a program at a given point in time versus postponing it.

Our estimated treatment parameter (1) can be cast into the sequential treatment framework proposed by Lechner and Miquel (2005) and Lechner (2004). These studies consider the identification and estimation of dynamic treatment effects with matching methods in a context where selection into and out of programs takes place sequentially from one period or stage to the next. Lechner and Miquel distinguish two versions of the conditional independence assumption: strong and weak dynamic conditional independence. The strong version is inappropriate in our case, because

both using administrative data and controlling for these motivational variables did not result in noticeably different estimated program effects compared to using administrative data only. This evidence also supports our point of view.

it effectively rules out to match on elapsed unemployment duration which is affected by earlier treatments. Under the weak dynamic conditional independence assumption, it is possible to identify and estimate the effect of joining versus waiting for those who join at the period in question.[18] This assumption allows to identify other more comprehensive parameters as well. However, it involves common support requirements which are infeasible in our case because treatment or exit to employment at some point excludes later treatment.

Heckman and Navarro (2007) consider the semiparametric identification of dynamic treatment effects in structural dynamic discrete choice models. Similar to Abbring and van den Berg (2003), the treatment status is allowed to depend on unobserved factors in the outcome equation. Heckman and Navarro require the existence of instruments that affect choices but not outcomes for semiparametric identification of causal effects. Matching methods, in contrast, rely on a rich set of conditioning variables that affect both the selection into treatment as well as the outcome such that any dependence between treatment assignment and outcome is netted out.[19] In a dynamic context, one needs instrumental variation at each stage of the sequential selection process or variation in the impact of time–invariant instruments (see Heckman and Navarro, 2007, Theorem 1). This variation must not be fully anticipated by the agents.[20] Using the reduced form model of Heckman and Navarro (section 2 of their paper), it is possible to identify the counterfactual outcome of waiting until a later period for those who join in an earlier period. Using the approach to identify other more comprehensive parameters in our case, more stringent modelling assumptions are necessary than we are willing to make and the necessary data requirements regarding instruments are not likely to hold. Unfortunately, we lack time–varying exogenous variables which affect the assignment process into treatment (see section 3 above).

[18]For instance, the effect of training versus waiting in the second stratum for the group of participants corresponds to the following dynamic average treatment effect on the treated in the Lechner/Miquel framework: nonparticipation in period one and training in period two versus nonparticipation in both periods for the population of those who participate in the second period.

[19]Heckman, Ichimura, and Todd (1998) compare matching to conventional selection models and show how to exploit information on exclusion restrictions and additive separability of the outcome equation for matching. See Heckman and Navarro-Lozano (2004) for a comparison of matching, instrumental variables and control function methods in a static context.

[20]Abbring and van den Berg (2003, 2005) argue that it is often difficult to maintain exclusion restrictions in dynamic settings with forward-looking agents.

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. The additional appendix reports 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. In particular, 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 level to be subject to social security taxation.[21] 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.

[21] Monthly 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.

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. 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 is the aggregation of two federal states.

The results for the Probit estimates show that the final specifications vary considerably over the inflow cohorts and the three time intervals even keeping the k/l–comparison constant.[22] 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. Furthermore, since the 'better' types leave unemployment earlier, the

[22] The tables with the propensity score estimations, the balancing tests and the figures showing the support of the propensity scores are displayed in the additional appendix.

long–term unemployed tend to be a more homogeneous subgroup.

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. 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 cases, but they do not follow a common pattern. 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.

For the balancing test in almost all cases, using a cubic in the estimated propensity score, we reject for at most one variable in the respective propensity score specification. Only in two out of 36 cases the test rejects for 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 valid.

The graphical examination of the common support requirement for estimating the average treatment effect on the treated (ATT) for training versus waiting reveals that lack of common support is a problem only in some cases. In these cases, it occasionally happens, that for very small estimated participation probabilities there are only nonparticipants, but no participating counterparts with such low estimated participation probabilities. This poses no problem for estimating the treatment effect on the treated for treatment versus waiting. Only in two of the treatment versus waiting comparisons we excluded 1.3% and 1.8% of the treated observations, respectively, from the estimation. Overall, we are also satisfied with the overlap of support for treatment versus treatment (k/l–pairs with $k, l \geq 1$). Though the graphical inspection of common support seems to reveal slight differences in support in a few cases, these differences mostly lie within the close neighborhood of the respective treated observation determined by the bandwidth. Therefore, we proceed without restricting the samples except for four cases where we drop about 2% to 4% of the treated observations, respectively. Detailed results of the balancing tests and common support graphics are shown in the additional appendix.

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 scores[23] 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 $\tau = 0, ..., \tau_{\max}$ since the beginning of program k, where $\tau_{\max} = 31, 29, 25$, 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 $\tau_{\max}$ since the beginning of the treatment and the differences during 8 quarters before the beginning of the unemployment spell. We put pointwise 95%–confidence intervals around the estimated treatment effects. The vertical gap at zero reflects the variable length of time between the start of the unemployment spell and the start of the treatment.

In order to contrast the initial negative lock–in effects of the programs with the

[23] We use the fitted index $X_i\hat{\beta}$ from the Probit estimates.

later positive program effects, we calculate the cumulated effects of every program 8, 16, and 24 quarters after the beginning of the program. The cumulated effects ($\equiv$ sum of quarter specific treatment 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. Tables 7 and 9 provide the results. These effects show the change in the total number of quarters in employment since the beginning of treatment. When the cumulated effects become positive then a negative lock–in effect is compensated by positive effects afterwards. The estimated standard errors are based on the bootstrap covariance estimates for the quarter specific treatment effects. A potential drawback of considering cumulated effects is that many of them are rather imprecisely estimated because they are summed over all quarters such that negative short–run effects and positive medium– to long–run effects are lumped together. Therefore, we also include a table (table 8) for training versus waiting with average ATT's, that are averages of the quarter specific treatment effects during the first three years and from year four onwards after the beginning of treatment. Table 8 allows to assess in a parsimonious way whether persistent significantly positive effects exist after the end of the lock–in period.

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 (cohort 86/87) and 5 (cohort 93/94).

We do not find significant pre–unemployment employment differences in any case. Since all individuals become unemployed eventually, 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 suggests that time–invariant unobserved heterogeneity does not invalidate our matching approach.

The results for 86/87 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 prove 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, 2007), 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 3.

For the 93/94 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 7. 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 86/87 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 in 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 in 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 significantly positive in strata 2 and 3 for SPST.

For the 93/94 cohort, the cumulated long–run effects after 24 quarters are significantly positive in all strata for SPST, in strata 2 and 3 for PF, and in stratum 2 for RT. For SPST, the pattern is similar to the earlier cohort. For PF, the effect is higher in strata 2 and 3 and much lower in 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 93/94 compared to 86/87. 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.

Table 8 shows the yearly averages of the ATT's which are typically more precisely estimated than the quarter specific treatment effects and the cumulated effects. For SPST, all average ATT's are significantly positive from year 2 onwards and slightly smaller for the cohort 93/94. For RT, all effects from year 3 onwards are significantly positive for the cohort 86/87 and only from year 4 onwards significantly positive for the cohort 93/94. RT shows longer significantly negative lock–in effects for the cohort 93/94 compared to 86/87. The pattern of the estimated effects for PF is less clear cut. For the cohort 86/87, there are significantly positive effects for one or two of the periods following year 1. For 93/94, PF shows significantly positive effects for all periods following year 1 in strata 2 and 3 but no significantly positive effects in stratum 1.

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 cumulated effects and the average treatment effects during years 2 to 4 are significantly positive for most programs. Overall, SPST seems to show the best results for the treated individuals. The positive effects of SPST deteriorate little for the later cohort. For RT, there is a noticeable increase in the lock–in period and a noticeable decline in the treatment effects for the later cohort and, for PF, the treatment effects deteriorate for stratum 1 and improve for later program starts.[24] The slight deterioration of the treatment effects for the later cohort could be caused by the worse business cycle conditions in the 90s.

5.2.2 Pairwise Comparisons of Training Programs

Next, we estimate pairwise ATT's both of the treatment k versus the alternative l for the treated in k and the treatment l versus k for the treated in l. As mentioned above,

[24]Following the suggestion of a referee, we also investigated whether there are heterogeneous treatment effects by gender, age, and qualification. In the matched samples, we regressed outcome differences on these covariates. However, based on bootstrapped standard errors we did not find any significant differences. These results are available upon request.

the first ATT does not necessarily coincide with the negative of the second ATT because of effect heterogeneity and the different composition of the two treatment groups (Lechner, 2001). The pairwise comparison allows to investigate whether the different programs are well targeted on average. 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 often are 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 9. 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 3 and 4.

In the vast 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.[25] 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 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.

The estimated cumulated effects in table 9 suggest that for the cohort 86/87 most significant effects are caused by the differential lock–in periods. Comparing SPST with RT for those treated in SPST ('SPST vs RT'), we find strong significantly positive effects after 8 quarters in stratum 1 and 3. Comparing RT with SPST and RT with PF both for those treated in RT, we find no significantly positive effects and the point estimates are even negative in a number of cases. 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

[25]These differences in employment history often become insignificant, if larger bandwidths are used. Further details are available upon request.

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 in stratum 2 and 3, but not significant. The long–run cumulated effects of PF in comparison to RT are also positive in stratum 1 and 3 but not significant.

For the cohort 93/94, the cumulated effects at 8 quarters are qualitatively similar reflecting again the different lock–in periods. Both PF, for stratum 1 and 3, and SPST, for all strata, seem to outperform RT in the short and medium run for the participants in PF and SPST, respectively. In the long run we only find significant effects for participation in SPST compared to RT in stratum 1. RT is also outperformed by SPST and PF even for participants in RT, though the effects are only strongly significant at 8 and at 16 quarters (the effects are of similar size at 24 quarters). Comparing SPST and PF, the cumulated effects are not significant but the point estimates suggest that SPST outperforms PF at least 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, some evidence for SPST and PF outperforming RT in the medium and long run even for the participants in RT, for the 93/94 cohort. The point estimates for SPST versus PF suggest for stratum 1 in 86/87 that the cumulated employment effect would have been better, if participants in SPST had instead participated in PF. For 93/94, 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 for 93/94 are significant.

6 Conclusions

Based on a unique administrative data set, which has only recently become 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 or fourth quarter (stratum 2), and between the fifth and the eighth quarter (stratum 3).

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 of around 4 to 6 quarters. The treatment effects deteriorate slightly from 1986/87 to 1993/94 in a number of cases, especially for RT and especially for treatments starting in stratum 1. For RT, the length of the lock-in period increases considerably for the later cohort. Both could reflect the worse business cycle conditions in the 1990s. The cumulated effects are significantly positive for most programs.

The pairwise comparisons of the three treatments, one against another, show first the differences in the lock–in periods and in most cases insignificant treatment effects in the medium and long run. There is, however, some evidence for SPST and PF outperforming RT in the medium and long run for the 1993/94 cohort. For 1993/94, SPST tends to outperform PF, but the effect is not significant.

Overall, SPST shows the best results for the treated individuals and the positive treatment effects for SPST are almost at the same level for 1993/94 compared to 1986/87. Note that SPST is by far the largest program and its share is even higher in 1993/94 compared to 1986/87. It is remarkable how little the effectiveness of

SPST differs between the two time periods despite the differences in business cycle conditions and the apparent change in the timing and length of treatments.

In comparison to the study by Lechner et al. (2005a) based on the same data source, our general results for the 1993/94 cohort are quite similar in most cases, even though the exact treatment definition, the choice of valid observations, and the employed econometric methods differ substantially. Notable differences from the results reported in Lechner et al. (2005a) are that we find significantly positive effects for treatments relative to nonparticipation much earlier after the treatment starts and that 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), Kluve 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, A. 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. van den Berg (2005). "Social Experiments and Instrumental Variables with Duration Outcomes." Tinbergen Institute Discussion Paper No. 05-047/3.

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* 289, 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 No. 04-41.

Bundesanstalt für Arbeit (1987, 1992, 1997). *Förderung der beruflichen Bildung 86, 91, Berufliche Weiterbildung 96.* Nürnberg: Bundesanstalt für Arbeit (various issues).

Caliendo, M., R. Hujer, S.L. Thomsen, and C. Zeiss (2004). "Einfluss von motivationalen Merkmalen aus Befragungsdaten auf die Schätzung des Beschäftigungseffektes von ABM–Maßnahmen." Unpublished Manuscript, Goethe University Frankfurt.

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. and H. Prey (2000). "Evaluating Public Sector Sponsored Training in East Germany." *Oxford Economic Papers* 52, 497–520.

Fitzenberger, B., A. Osikominu and R. Völter (2006). "Get Training or Wait? Long–Run Employment Effects of Training Programs for the Unemployed in West Germany." ZEW Discussion Paper No. 06-039.

Fitzenberger, B. and S. Speckesser (2007). "Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany." *Empirical Economics* 32(2), 529-573.

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.

Hagen, T. and V. Steiner (2000). "Von der Finanzierung der Arbeitslosigkeit zur Förderung der Arbeit – Analysen und Handlungsempfehlungen zur Arbeitsmarktpolitik." ZEW Wirtschaftsanalyszen, 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.

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

Heckman, J. and S. Navarro (2007). "Dynamic Discrete Choice and Dynamic Treatment Effects." *Journal of Econometrics* 136, 341–396.

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* 35, 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. Pfeiffer (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 (2005). "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions" Discussion Paper, 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.

Plassmann, G. (2002). "Der Einfluss der Arbeitslosenversicherung auf die Arbeitslosigkeit in Deutschland — eine mikroökonomische und empirische Untersuchung." *Beiträge zur Arbeitsmarkt- und Berufsforschung* 255, IAB, Nürnberg.

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.

Schneider, H., K. Brenke, D. Hess, L. Kaiser, J. Steinwede und A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission – Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen." IZA Research Report, No. 7, Bonn.

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.

Speckesser, S. (2004). "Essays on Evaluation of Active Labour Market Policy." Dissertation, University of Mannheim.

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

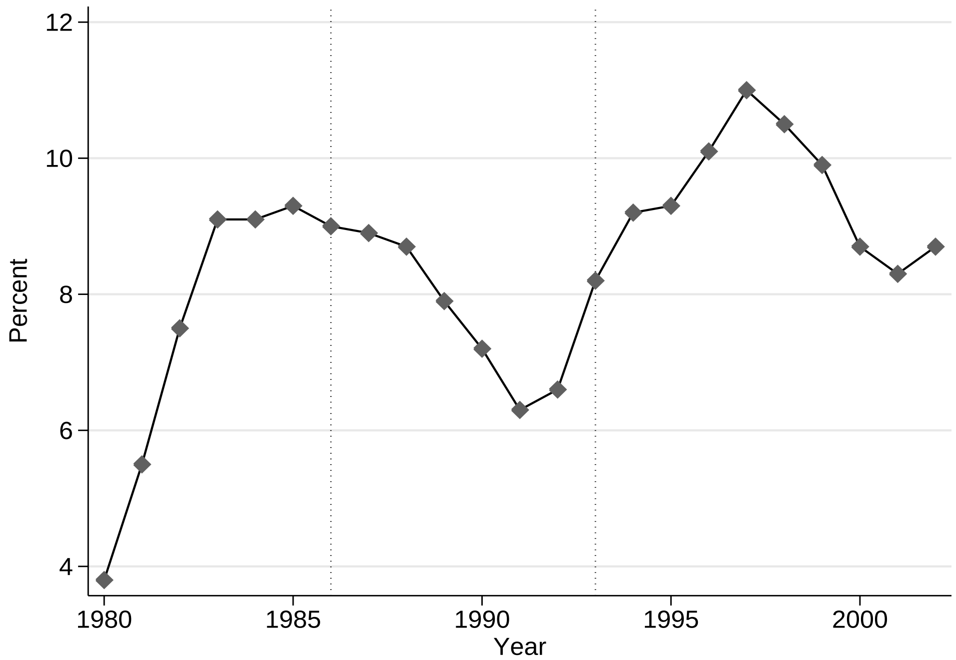
Descriptive Statistics and Description of Data

Table 2: Participation in Further Training in West Germany until 1997

Year	Annual entries				Annual average stocks
	Total	Further vocational training	Retraining	Integration subsidy	
1980	247.0	176.5	37.9	32.6	177.1
1985	409.3	336.5	45.1	27.7	245.8
1986	530.0	426.0	59.1	44.9	308.1
1987	596.3	482.6	64.5	49.2	346.1
1988	565.6	448.7	65.7	51.2	361.5
1989	489.9	388.4	61.0	40.8	357.9
1990	574.0	442.8	63.4	67.9	349.7
1991	593.9	474.5	70.5	48.9	364.5
1992	574.7	464.5	81.5	28.7	372.1
1993	348.1	266.0	72.2	9.9	348.4
1994	306.8	224.9	73.1	8.8	308.8
1995	401.6	309.7	81.8	10.0	304.3
1996	378.4	291.6	77.3	9.5	306.6

Remark: All numbers in thousands. Source: Bundesanstalt für Arbeit (1987, 1992, 1997).

Figure 1: Unemployment Rate in West Germany



Source: Statistikangebot der Bundesagentur für Arbeit,
http://www.pub.arbeitsamt.de/hst/services/statistik/detail_2004/d.html

Table 3: 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 4: 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 5: 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

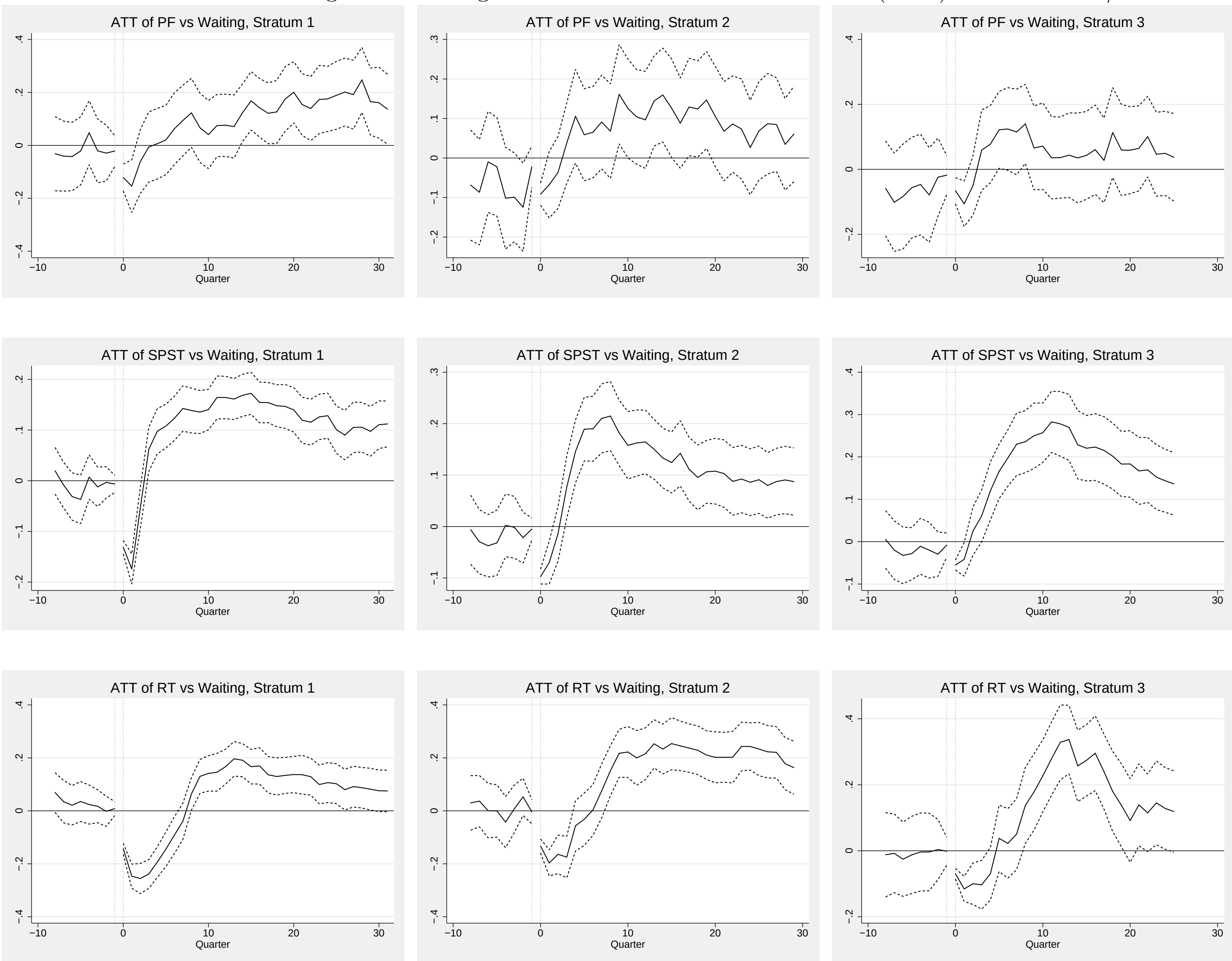
Table 6: Realized Duration of Training Spells in months

	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 uninterrupted training.

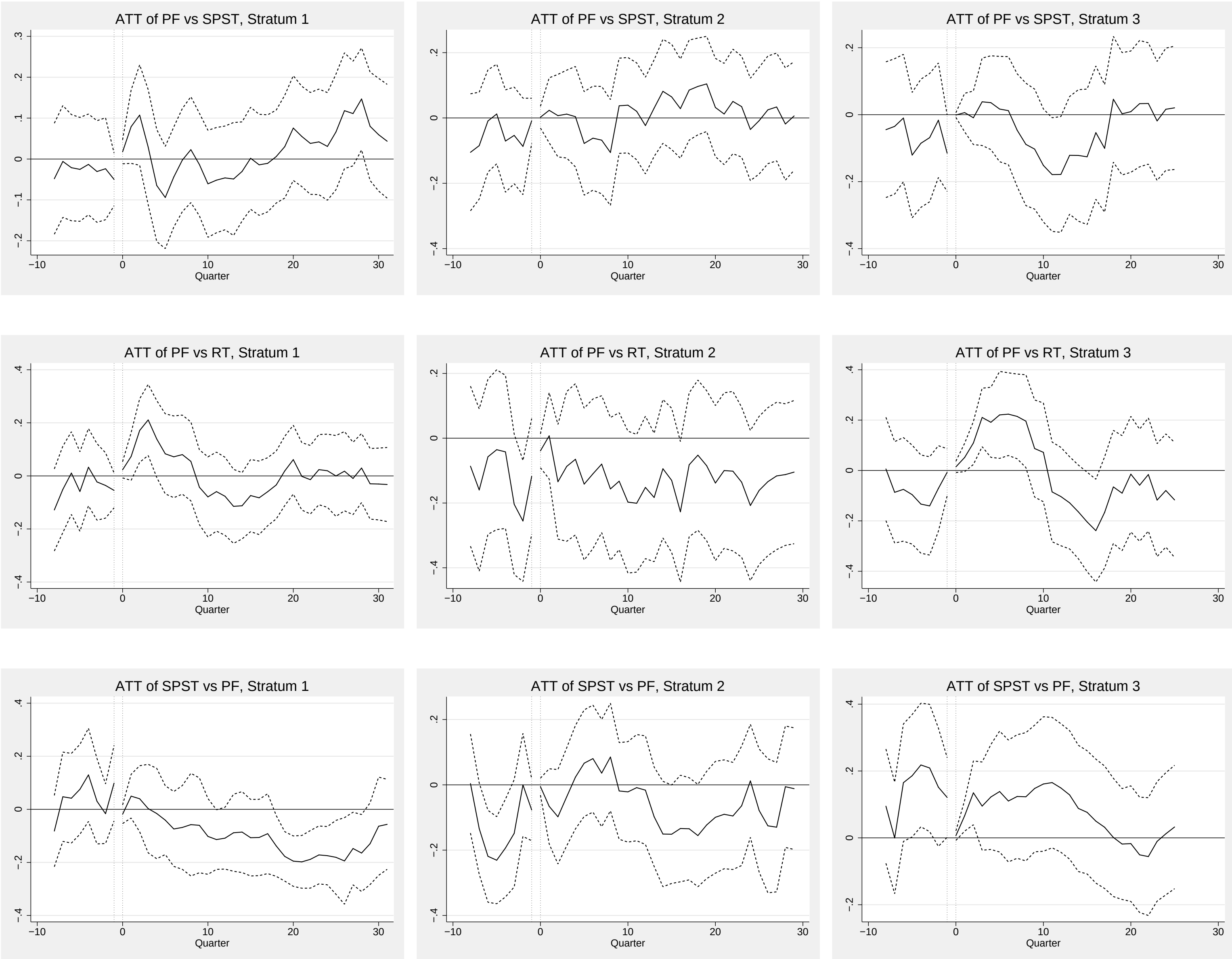
Estimated Employment Effects of Further Training Programs

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



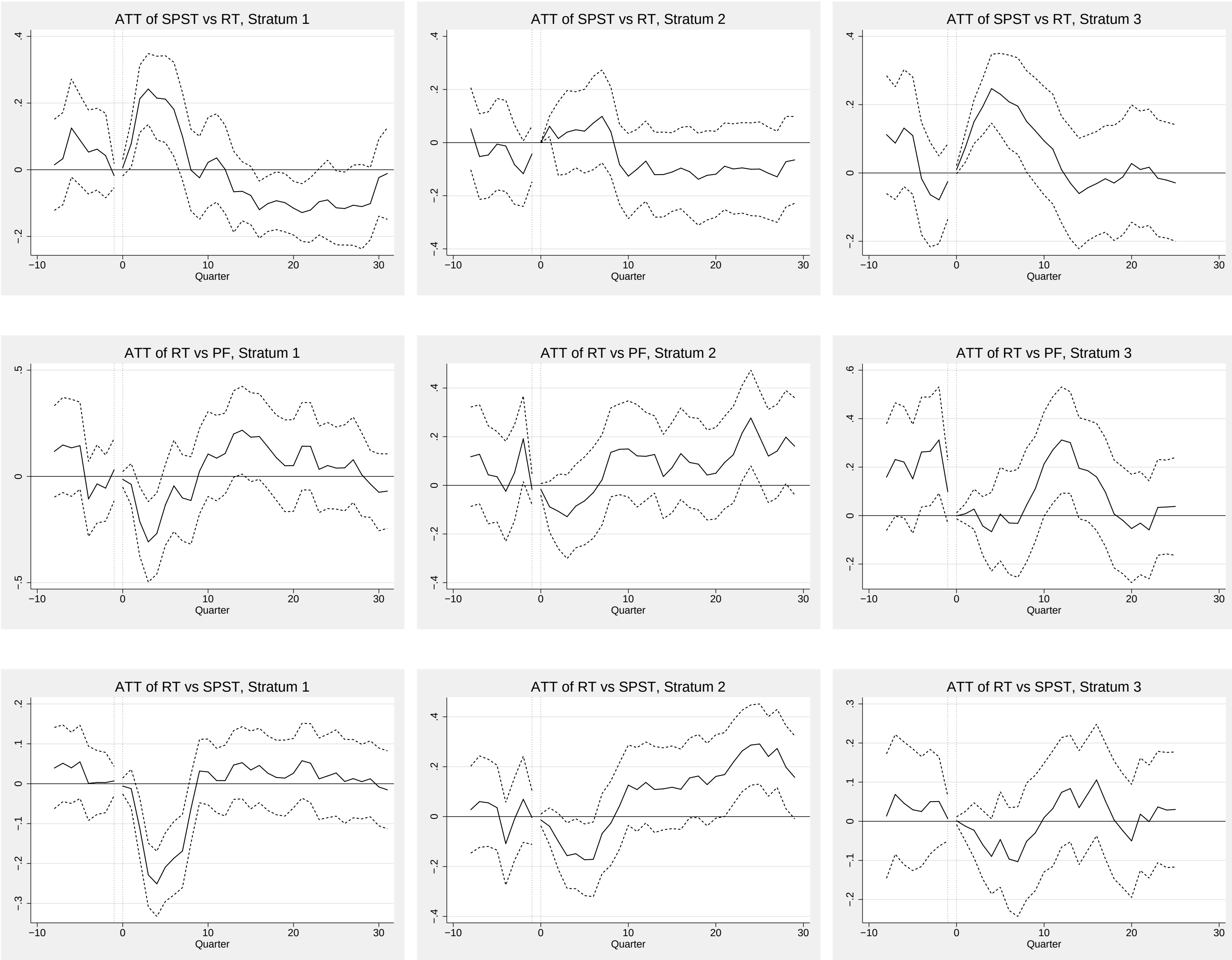
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

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



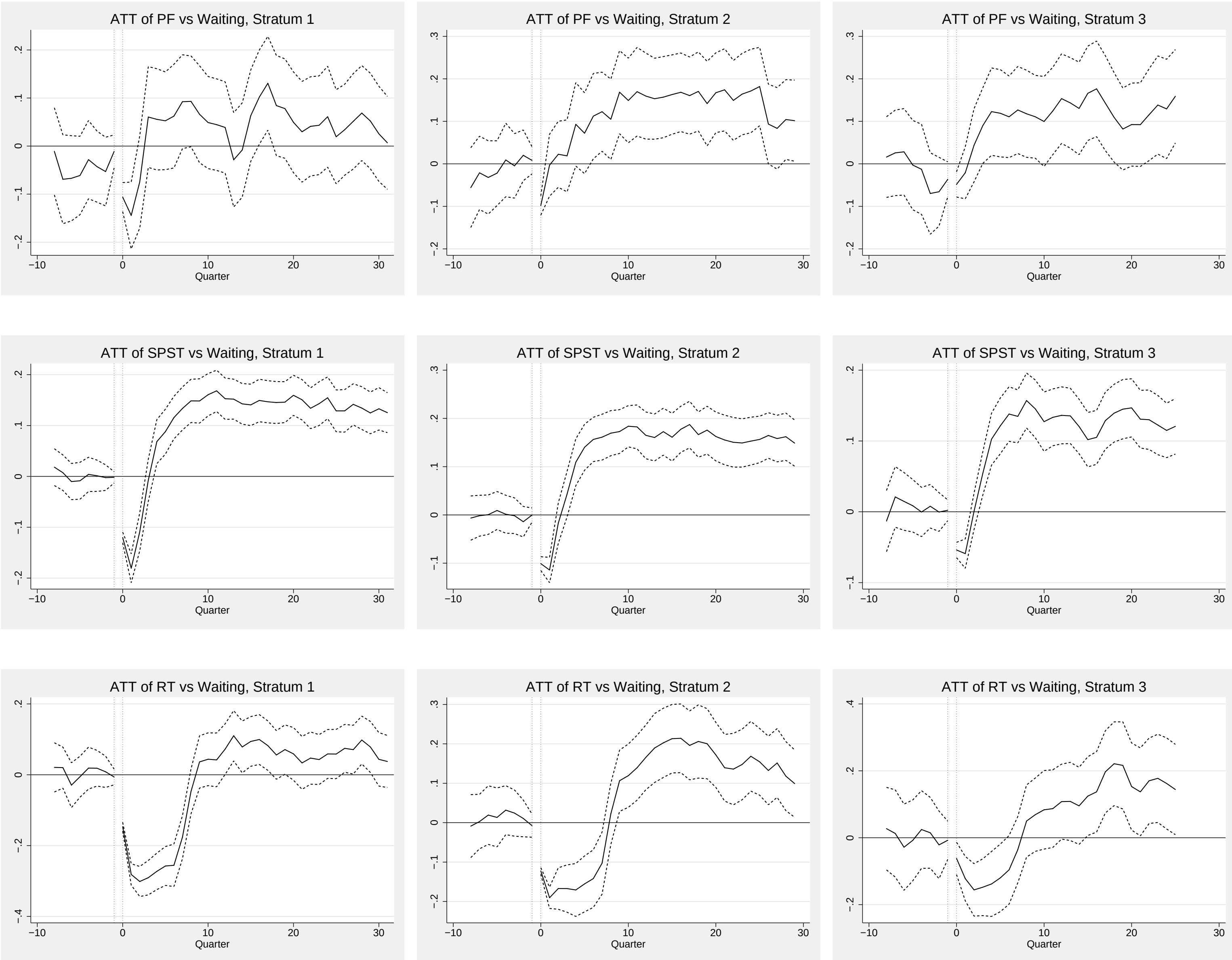
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

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



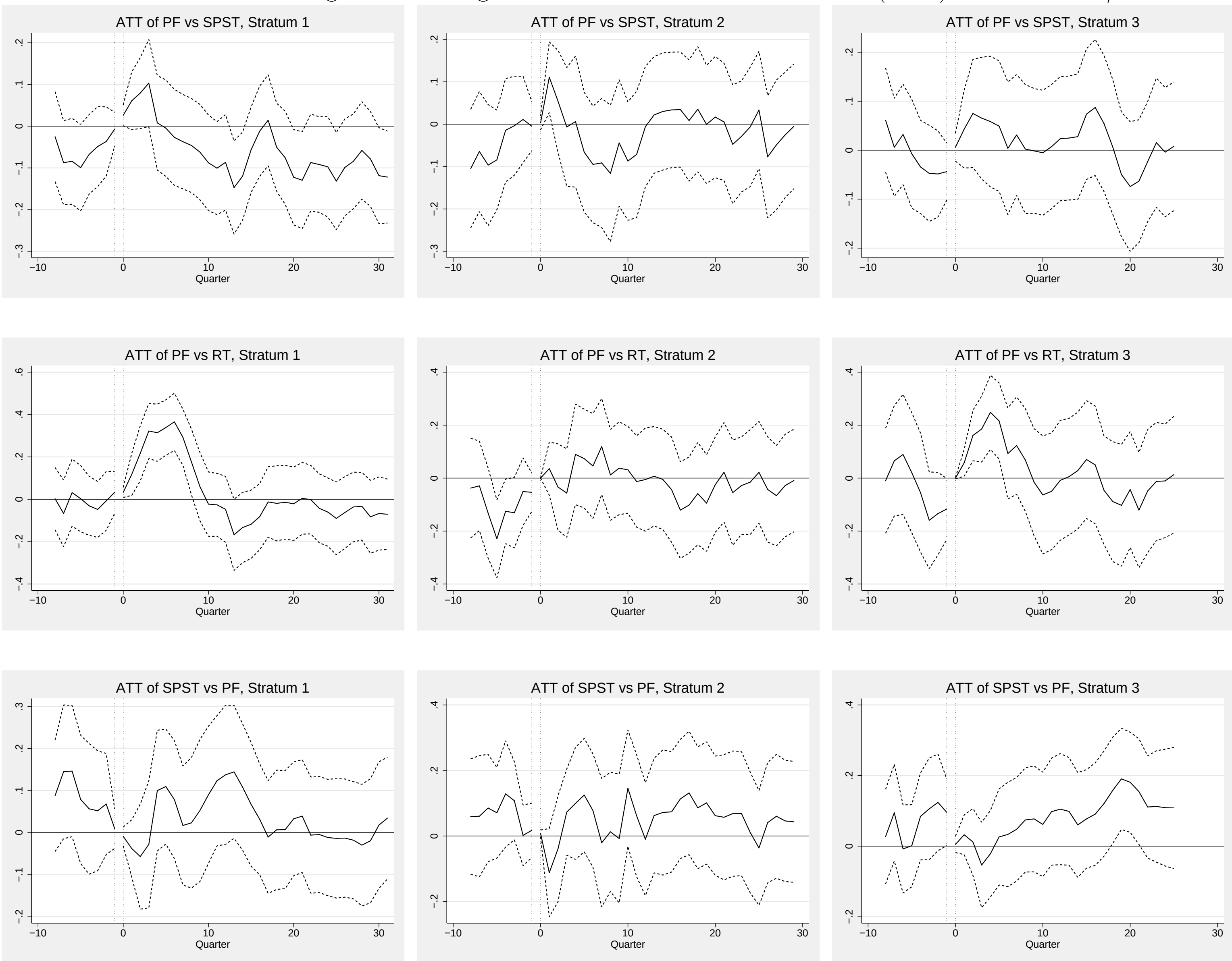
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

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



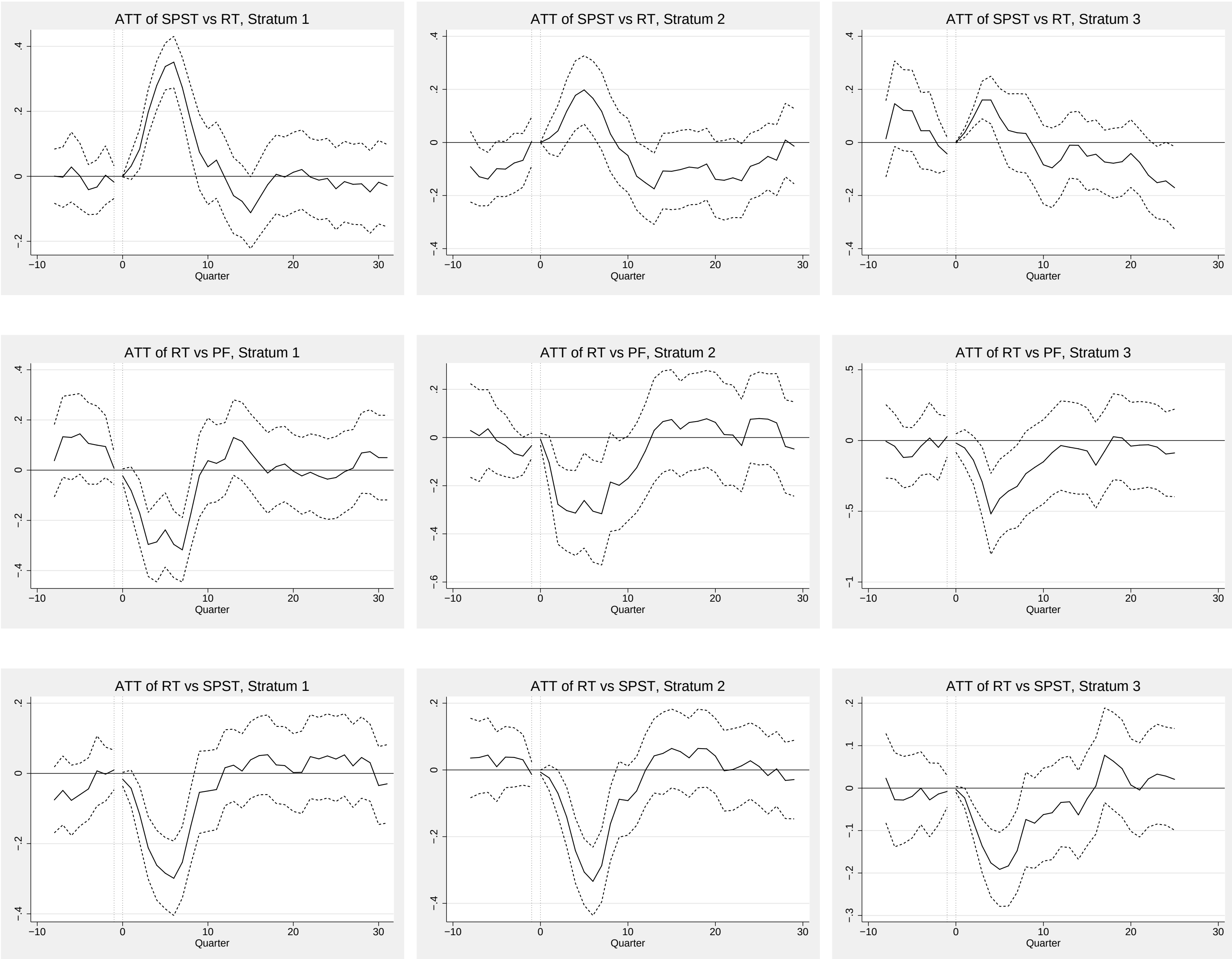
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) quarters on the abscissa.

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



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



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

Table 7: Cumulated differences in employment rates – sum of quarter specific average treatment effects on the treated since beginning of treatment – Training versus Waiting

	Cumulated Treatment Effects		
	8 quarters	16 quarters	24 quarters
	PF vs Waiting, Cohort 86/87		
Stratum 1	-0.159 (0.382)	0.586 (0.706)	1.817 (1.018)*
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)
	SPST vs Waiting, Cohort 86/87		
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)***
	RT vs Waiting, Cohort 86/87		
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)***
	PF vs Waiting, Cohort 93/94		
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)***
	SPST vs Waiting, Cohort 93/94		
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)***
	RT vs Waiting, Cohort 93/94		
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	-0.878 (0.260)***	-0.152 (0.563)	1.258 (0.904)

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

Table 8: Averages of quarter specific average treatment effects on the treated

	1st year	2nd year	3rd year	year 4 onwards
Average ATT, PF vs Waiting, Cohort 86/87				
Stratum 1	-0.086 (0.044)*	0.046 (0.061)	0.076 (0.057)	0.145 (0.048)***
Stratum 2	-0.039 (0.031)	0.080 (0.054)	0.115 (0.056)**	0.103 (0.052)**
Stratum 3	-0.041 (0.032)	0.110 (0.055)**	0.078 (0.059)	0.056 (0.059)
Average ATT, SPST vs Waiting, Cohort 86/87				
Stratum 1	-0.074 (0.013)***	0.118 (0.020)***	0.145 (0.020)***	0.143 (0.019)***
Stratum 2	-0.026 (0.018)	0.184 (0.030)***	0.179 (0.031)***	0.114 (0.027)***
Stratum 3	-0.003 (0.019)	0.178 (0.030)***	0.257 (0.034)***	0.198 (0.034)***
Average ATT, RT vs Waiting, Cohort 86/87				
Stratum 1	-0.221 (0.020)***	-0.117 (0.029)***	0.120 (0.029)***	0.143 (0.029)***
Stratum 2	-0.167 (0.024)***	-0.002 (0.046)	0.198 (0.043)***	0.229 (0.041)***
Stratum 3	-0.097 (0.020)***	0.010 (0.043)	0.206 (0.053)***	0.199 (0.049)***

	1st year	2nd year	3rd year	year 4 onwards
Average ATT, PF vs Waiting, Cohort 93/94				
Stratum 1	-0.066 (0.032)**	0.066(0.047)	0.063 (0.047)	0.050 (0.044)
Stratum 2	-0.015 (0.026)	0.100 (0.044)**	0.148 (0.045)***	0.163 (0.040)***
Stratum 3	0.016 (0.028)	0.120 (0.047)**	0.113 (0.048)**	0.131 (0.045)***
Average ATT, SPST vs Waiting, Cohort 93/94				
Stratum 1	-0.104 (0.012)***	0.101 (0.020)***	0.156 (0.020)***	0.146 (0.017)***
Stratum 2	-0.047 (0.013)***	0.142 (0.022)***	0.177 (0.021)***	0.164 (0.022)***
Stratum 3	-0.015 (0.009)	0.124 (0.018)***	0.141 (0.019)***	0.127 (0.017)***
Average ATT, RT vs Waiting, Cohort 93/94				
Stratum 1	-0.255 (0.014)***	-0.241 (0.026)***	0.019 (0.033)	0.069 (0.030)**
Stratum 2	-0.162 (0.017)***	-0.143 (0.035)***	0.097 (0.036)***	0.179 (0.037)***
Stratum 3	-0.122 (0.031)***	-0.097 (0.044)**	0.072 (0.052)	0.154 (0.050)***

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

Table 9: Cumulated differences in employment rates – sum of quarter specific average treatment effects on the treated since beginning of treatment – Pairwise comparisons of training programs

	Cumulated Treatment Effects		
	8 quarters	16 quarters	24 quarters
	PF vs SPST, Cohort 86/87		
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	0.053 (0.331)	-1.016 (0.866)	-1.065 (1.455)
	PF vs RT, Cohort 86/87		
Stratum 1	0.853 (0.395)**	0.348 (0.736)	0.259 (1.117)
Stratum 2	-0.650 (0.579)	-1.896 (1.101)*	-2.819 (1.726)
Stratum 3	1.237 (0.350)***	0.907 (0.836)	0.140 (1.402)
	SPST vs PF, Cohort 86/87		
Stratum 1	-0.125 (0.339)	-0.848 (0.714)	-2.114 (1.041)**
Stratum 2	0 (0.437)	-0.380 (0.888)	-1.275 (1.375)
Stratum 3	0.798 (0.406)**	1.837 (1.022)*	1.768 (1.601)
	SPST vs RT, Cohort 86/87		
Stratum 1	1.246 (0.354)***	1.072 (0.599)*	0.199 (0.789)
Stratum 2	0.380 (0.394)	-0.308 (0.761)	-1.177 (1.147)
Stratum 3	1.310 (0.286)***	1.625 (0.771)**	1.575 (1.240)
	RT vs PF, Cohort 86/87		
Stratum 1	-1.121 (0.509)**	-0.309 (1.154)	0.523 (1.857)
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)
	RT vs SPST, Cohort 86/87		
Stratum 1	-1.173 (0.227)***	-1.024 (0.440)**	-0.774 (0.698)
Stratum 2	-0.868 (0.347)**	-0.142 (0.824)	1.225 (1.326)
Stratum 3	-0.430 (0.269)	-0.207 (0.691)	-0.066 (1.098)

Cumulated Treatment Effects			
	8 quarters	16 quarters	24 quarters
PF vs SPST, Cohort 93/94			
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)
PF vs RT, Cohort 93/94			
	8 quarters	16 quarters	24 quarters
Stratum 1	2.002 (0.376)***	1.723 (0.763)**	1.534 (1.234)
Stratum 2	0.274 (0.521)	0.295 (0.881)	-0.166 (1.285)
Stratum 3	1.084 (0.327)***	1.119 (0.971)	0.707 (1.616)
SPST vs PF, Cohort 93/94			
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)
SPST vs RT, Cohort 93/94			
Stratum 1	1.554 (0.201)***	1.623 (0.531)***	1.552 (0.905)*
Stratum 2	0.836 (0.339)**	0.126 (0.659)	-0.805 (1.002)
Stratum 3	0.629 (0.221)***	0.323 (0.570)	-0.337 (0.864)
RT vs PF, Cohort 93/94			
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)
RT vs SPST, Cohort 93/94			
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.

Additional Appendix to “Get Training or Wait? Long–Run Employment Effects of Training Programs for the Unemployed in West Germany” by B. Fitzenberger, A. Osikominu, and R. Völter

Estimation Results for the Propensity Score

Sample Sizes

Cohort 86/87			
	Stratum 1	Stratum 2	Stratum 3
Waiting	20153	9440	6364
PF	74	60	69
SPST	503	257	176
RT	172	101	71

Cohort 93/94			
	Stratum 1	Stratum 2	Stratum 3
Waiting	24223	13751	9244
PF	102	102	86
SPST	528	481	669
RT	198	138	106

Variable Definitions

Table 10: Variable Definitions

Label	Definition
Personal Attributes	
aXXYY	Age at start of unemployment ≥XX and ≤ YY
age	Age at start of unemployment
female	Female
foreign	No German citizenship
kids	Has dependent children
married	Married
qual_u	No vocational training degree
qual_l	No vocational training degree or education information missing
qual_m	Vocational training degree
qual_h	University/College degree
Last Employment	
BER1	Apprentice
BER2	Blue Collar Worker
BER3	White Collar Worker
BER4	Worker at home with low hours or BER missing
BER5	Part–time working
pearn	Daily earnings ≥ 15 Euro per day in 1995 Euro
earncens	Earnings censored at social security taxation threshold
earn	Daily earnings if pearn=1 and earncens=0, otherwise zero
logearn	log(earn) if pearn=1 and earncens=0, otherwise zero
logearnsq	logearn squared
earnp90	Daily earnings above 90th percentile
Last Employer	
industry1	Agriculture
industry2	Basic materials
industry3	Metal, vehicles, electronics
industry4	Light industry
industry5	Construction
industry6	Production oriented services, trade, banking
<continued on next page>	

Table 10: Variable Definitions <continued>

Label	Definition
industry7	Consumer oriented services, organization and social services
frmsize1	Firm Size (employment) missing or $\leq$ 10
frmsize2	Firm Size (employment) $>$ 10 and $\leq$ 200
frmsize3	Firm Size (employment) $>$ 200 and $\leq$ 500
frmsize4	Firm Size (employment) $>$ 500
Employment and Program History	
preexM	Employed M (M=6, 12, 24) month before unemployment starts
preex60cumst	Number of months employed in the last 60 months before unemployment starts, standardized
preex60sq	preex60cumst squared
pretxY	Participation in any ALMP program reported in our data in year(s) Y (Y=1, 2, 3–5) before unemployment starts
Regional Information	
state6	Schleswig-Holstein/Hamburg
state7	Niedersachsen-Bremen
state8	Nordrhein-Westfalen
state9	Hessen
state10	Rheinland-Pfalz/ Saarland
state11	Baden-Württemberg
state12	Bayern
denst	population density (standardized)
densq	denst squared
R1	Population density $<$100 inhabitants per square kilometer, Rural area
R2	Population density $\geq$100 and $<$ 150, Medium population density
R3	Population density $\geq$150 and $<$ 400, Dense area
R4	Population density $\geq$400, Metropolitan area
ur	Unemployment rate at district level (Kreis), 80s
ursq	ur squared
urtb	Unemployment rate at district level (Kreis), 90s

<continued on next page>

Table 10: Variable Definitions <continued>

Label	Definition
urtbsq	urtb squared
urtb100	urtb/100
Calendar Time of Entry into Unemployment	
tnull	First unemployment month (months counted from January 1960)
uentry	First unemployment month (months counted from January 1986 (1993) in the 80s (90s))
uentry2	uentry squared
yYY	Unemployment begins in year YY
qQ	Unemployment begins in quarter Q of the year
yYYqQ	Unemployment begins in quarter Q of year YY
Interaction of Variables	
f_	female
for_	foreign

All variables are defined at the time of entry into unemployment and constant during the unemployment spell.

Results of Propensity Score Estimations and Balancing Tests

Remark: The propensity score tables show the estimated coefficients of the probit regressions of the conditional probability to participate in the first of the two treatments mentioned in the header. The estimations are carried out separately for each time window of elapsed unemployment duration (Stratum 1, 2, and 3). Standard errors are in parentheses. *, **, *** means significant at the 10%, 5%, 1% level, respectively, in a two–sided test. Each probit table is followed by two tables indicating how many regressors pass the Smith/Todd (2005) balancing test at different significance levels using a cubic and a quartic of the propensity score, respectively. Graphs with the densities of the propensity scores are in the next subsection.

Treatment PF vs Waiting, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
state10	0.327 (0.140)**		
state79	0.479 (0.090)***		
a2529			0.362 (0.315)
a2534		0.104 (0.197)	
a3034	-0.073 (0.124)		0.829 (0.307)***
a3539	-0.020 (0.136)	0.050 (0.233)	
a3544			0.734 (0.333)**
a4044	-0.129 (0.154)	0.107 (0.239)	
a4549	-0.070 (0.141)	0.240 (0.223)	0.487 (0.267)*
a5055	-0.308 (0.177)*		
ur	0.094 (0.067)	0.149 (0.085)*	0.265 (0.095)***
ursq	-0.005 (0.003)*	-0.007 (0.004)*	-0.011 (0.004)***
densq			0.094 (0.042)**
denst			-0.183 (0.081)**
earn			-0.021 (0.009)**
f_BER3		0.191 (0.198)	
f_a3034		0.546 (0.218)**	
f_a3539		0.439 (0.306)	
f_a3544	0.355 (0.200)*		
f_preex60cumst			-0.089 (0.090)
female	-0.388 (0.126)***	-0.709 (0.202)***	-0.468 (0.117)***
frmsize23	0.211 (0.089)**	0.148 (0.116)	
frmsize4		0.273 (0.154)*	
logearn	0.018 (0.071)	0.038 (0.066)	
logearnsq			0.099 (0.048)**
married			-0.218 (0.107)**
pearn	-0.374 (0.384)		
preex12			-0.305 (0.130)**
preex24	0.205 (0.117)*	-0.181 (0.098)*	
preex60cumst	-0.063 (0.057)		0.189 (0.073)***
preex60sq	0.084 (0.037)**		0.113 (0.043)***
pretx1			-0.423 (0.281)
pretx2			0.652 (0.208)***
pretx35			-0.232 (0.206)
qual_l		-0.528 (0.334)	
qual_l_a2539		0.866 (0.356)**	
qual_m_a3544			0.088 (0.189)
qual_m_a4555			0.036 (0.263)
uentry		0.036 (0.028)	0.010 (0.007)
uentry2		-0.002 (0.001)	
y86q2	0.705 (0.317)**		
y86q34	0.779 (0.290)***		
y87q1	0.812 (0.292)***		
y87q2	0.926 (0.303)***		
y87q3	1.050 (0.295)***		
y87q4	0.848 (0.296)***		
_cons	-3.935 (0.535)***	-3.496 (0.555)***	-4.605 (0.677)***
N	20227	9500	6433

Treatment PF vs Waiting, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	21	22	23	23
Stratum 2	16	17	17	18
Stratum 3	16	21	21	22

Treatment PF vs Waiting, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	15	17	17	23
Stratum 2	15	15	16	18
Stratum 3	12	14	17	22

Treatment SPST vs Waiting, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER1	-0.003 (0.194)		
BER2		-0.067 (0.126)	-0.117 (0.121)
BER3	0.302 (0.054)***	0.275 (0.170)	0.202 (0.112)*
BER3_a2539		-0.120 (0.127)	
industry3	0.317 (0.065)***		0.163 (0.146)
industry4	0.098 (0.080)		0.081 (0.160)
industry5	-0.161 (0.076)**		
industry6	0.230 (0.050)***		0.112 (0.129)
industry7			0.014 (0.132)
a2529	0.404 (0.091)***		1.075 (0.235)***
a3034	0.382 (0.094)***	-0.030 (0.102)	1.192 (0.236)***
a3539	0.445 (0.096)***	0.110 (0.226)	
a3544			0.814 (0.256)***
a4044	0.213 (0.118)*	-0.096 (0.231)	
a4549	0.233 (0.103)**	-0.023 (0.238)	0.727 (0.191)***
a5055		-0.504 (0.248)**	
ur			0.134 (0.059)**
ursq			-0.006 (0.003)**
denst			0.038 (0.034)
earncens	0.372 (0.232)	0.624 (0.240)***	
f_BER2	-0.292 (0.094)***		
f_BER3		0.063 (0.134)	
f_industry7		-0.224 (0.095)**	
f_a2529		-0.069 (0.126)	
f_a3544			0.101 (0.168)
f_a4044	0.355 (0.125)***		
f_a4555			0.626 (0.205)***
female	0.105 (0.056)*	0.012 (0.113)	-0.217 (0.100)**

SPST vs Waiting, Cohort 86/87 West Germany – continued			
	Stratum 1	Stratum 2	Stratum 3
for_age		-0.009 (0.003)***	
foreign	-0.109 (0.083)		-0.327 (0.142)**
logearn	0.107 (0.044)**		
logearnsq		0.032 (0.010)***	
m_industry5			-0.755 (0.293)***
married	-0.120 (0.042)***	-0.191 (0.059)***	-0.238 (0.073)***
preex12	0.129 (0.049)***		
preex60cumst	-0.045 (0.022)**	-0.033 (0.029)	
preex60sq			0.002 (0.033)
preex60sq_a3544		0.125 (0.039)***	0.145 (0.056)***
pretx1	0.204 (0.096)**	0.235 (0.122)*	0.372 (0.150)**
pretx2	0.071 (0.095)		
pretx35	0.066 (0.072)		
qual_h	0.277 (0.108)**	0.387 (0.196)**	
qual_h_a3544		-0.224 (0.299)	
qual_h_a4555		0.252 (0.349)	
qual_m	0.261 (0.072)***	0.316 (0.143)**	
qual_m_a3544		-0.224 (0.208)	
qual_m_a4555		0.077 (0.240)	
uentry	0.024 (0.010)**		
uentry2	-0.001 (0.000)*	-0 (0.000)***	
y86q2			-0.267 (0.141)*
y86q3			-0.046 (0.127)
y86q4			-0.120 (0.131)
y87q1			-0.245 (0.134)*
y87q2			-0.188 (0.134)
y87q3			-0.399 (0.146)***
y87q4			-0.016 (0.124)
_cons	-3.345 (0.213)***	-2.508 (0.233)***	-3.316 (0.419)***
N	20656	9697	6540

Treatment SPST vs Waiting, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	24	26	26	27
Stratum 2	20	25	26	26
Stratum 3	24	27	27	29

Treatment SPST vs Waiting, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	19	23	25	27
Stratum 2	19	21	24	26
Stratum 3	20	22	23	29

Treatment RT vs Waiting, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER3			-0.099 (0.121)
state10	-0.074 (0.128)	-0.275 (0.186)	
state11	0.081 (0.101)	-0.088 (0.140)	
state12		-0.245 (0.149)*	
state6		0.026 (0.132)	-0.030 (0.183)
state612	-0.175 (0.086)**		
state7	-0.089 (0.099)	-0.141 (0.128)	
state710			0.347 (0.108)***
state9	0.200 (0.100)**	0.077 (0.139)	
a2529	0.837 (0.146)***	0.994 (0.301)***	
a2534			0.760 (0.356)**
a3034	0.848 (0.150)***	1.062 (0.303)***	
a3539		0.666 (0.320)**	0.669 (0.425)
a3544	0.658 (0.151)***		
a4044		0.682 (0.326)**	0.400 (0.431)
a4549		0.449 (0.339)	
densq	-0.036 (0.029)	-0.170 (0.077)**	
denst	0.110 (0.049)**	0.119 (0.072)*	0.061 (0.047)
f_densq		0.072 (0.080)	
f_preex60cumst		-0.137 (0.076)*	
f_qual_h	0.341 (0.144)**		
f_uentry			0.009 (0.014)
female	-0.112 (0.070)	-0.585 (0.219)***	-0.196 (0.196)
foreign	-0.340 (0.130)***	-0.577 (0.205)***	-0.209 (0.188)
logearn	0.074 (0.061)	0.061 (0.055)	
logearnsq			0.028 (0.017)*
m_BER2		-0.420 (0.207)**	
m_BER3		-0.318 (0.228)	
pearn	-0.290 (0.329)		-0.633 (0.355)*
preex12	0.147 (0.073)**		
preex60cumst	-0.150 (0.035)***	-0.114 (0.064)*	
preex60cumst_a2534			-0.072 (0.055)
preex60cumst_a3544	0.164 (0.066)**	0.233 (0.104)**	-0.018 (0.079)
preex60sq		-0.061 (0.039)	
pretx1	0.267 (0.118)**		
qual_h			0.156 (0.263)
qual_h_a3544			0.587 (0.422)
qual_m			-0.138 (0.453)
qual_m_a2534			0.316 (0.472)
qual_m_a3544			0.423 (0.526)
uentry		-0.003 (0.006)	0.001 (0.010)
y86q2			0.362 (0.147)**
y86q23	0.155 (0.120)		
y86q4	0.208 (0.125)*		
y87q1	0.192 (0.119)		
y87q2	0.235 (0.136)*		
y87q3	0.371 (0.124)***		
y87q4	0.342 (0.119)***		
_cons	-3.308 (0.292)***	-2.609 (0.411)***	-2.950 (0.441)***
N	20325	9541	6435

Treatment RT vs Waiting, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	24	24	24	25
Stratum 2	19	22	23	24
Stratum 3	18	20	20	21

Treatment RT vs Waiting, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	15	19	22	25
Stratum 2	11	12	15	24
Stratum 3	13	14	17	21

Treatment PF vs SPST, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
R1	0.528 (0.226)**		
a2534	-0.197 (0.147)		
densq			0.134 (0.078)*
denst			-0.304 (0.139)**
f_BER3	-0.621 (0.275)**		
f_preex6		-1.006 (0.390)***	
f_preex60sq		-0.261 (0.129)**	
f_tnull	0.037 (0.011)***		
female		1.510 (0.741)**	-0.868 (0.200)***
foreign	0.437 (0.256)*	0.691 (0.318)**	
logearn	0.729 (0.330)**		1.170 (0.446)***
logearnsq	-0.181 (0.069)***	-0.066 (0.034)**	-0.232 (0.086)***
m_logearn		0.301 (0.189)	
m_tnull	0.039 (0.012)***		
preex6	-0.342 (0.167)**		
qual_mh		-0.753 (0.235)***	
qual_u			0.651 (0.227)***
tnull		0.014 (0.013)	
_cons	-13.225 (3.797)***	-4.685 (4.442)	-1.449 (0.647)**
N	577	317	245

Treatment PF vs SPST, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	8	9	9	9
Stratum 2	8	8	8	8
Stratum 3	5	5	6	6

Treatment PF vs SPST, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	8	9	9	9
Stratum 2	8	8	8	8
Stratum 3	6	6	6	6

Treatment PF vs RT, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
state10	0.831 (0.335)**		
state12	0.580 (0.299)*		
state7	0.868 (0.266)***		
state9	0.634 (0.270)**		
industry7		-0.707 (0.275)**	
a2529		-1.428 (0.418)***	-1.199 (0.326)***
a3034		-1.033 (0.410)**	-0.616 (0.318)*
a3539		-0.592 (0.473)	-0.663 (0.349)*
a4044		-0.822 (0.543)	
ur100			30.948 (21.463)
ursq			-141.761 (95.828)
densq	0.103 (0.099)	0.352 (0.147)**	
denst	-0.221 (0.157)	-0.524 (0.187)***	
f_a2534	-1.197 (0.386)***		
f_preex12	-0.317 (0.489)		
f_preex24	0.924 (0.546)*		
female	-1.041 (0.705)		-0.544 (0.239)**
foreign		1.455 (0.437)***	
m_a2534	-1.121 (0.376)***		
m_a3544	-0.868 (0.403)**		
married		-0.356 (0.237)	
tnull	0.007 (0.015)		
_cons	-2.379 (4.870)	0.632 (0.411)	-0.713 (1.168)
N	246	161	140

Treatment PF vs RT, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	12	12	13	13
Stratum 2	8	8	9	9
Stratum 3	6	6	6	6

Treatment PF vs RT, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	12	13	13	13
Stratum 2	8	8	9	9
Stratum 3	6	6	6	6

Treatment RT vs SPST, Cohort 86/87 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER2			0.327 (0.175)*
BER3	-0.474 (0.118)***		
state1012		-0.450 (0.209)**	
a2529			1.051 (0.382)***
a2534	1.004 (0.251)***	1.033 (0.263)***	
a3034			0.914 (0.393)**
a3539			1.132 (0.406)***
a3544	0.701 (0.265)***	0.392 (0.293)	
a4044			0.759 (0.451)*
densq	-0.111 (0.053)**	-0.119 (0.068)*	
denst	0.210 (0.086)**		
earnp90		-0.411 (0.265)	
f_preex60cumst		-0.465 (0.162)***	
f_preex60sq		-0.269 (0.093)***	
preex60cumst	-0.170 (0.070)**		
preex60sq	-0.031 (0.053)		
qual_h		-0.776 (0.330)**	
qual_m		-0.710 (0.229)***	
qual_u	0.463 (0.185)**		
y87	0.299 (0.112)***		
_cons	-1.453 (0.267)***	-0.372 (0.316)	-1.624 (0.359)***
N	675	358	247

Treatment RT vs SPST, Cohort 86/87 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	9	9	9	9
Stratum 2	8	9	9	9
Stratum 3	5	5	5	5

Treatment RT vs SPST, Cohort 86/87 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	9	9	9	9
Stratum 2	8	9	9	9
Stratum 3	5	5	5	5

Treatment PF vs Waiting, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER2	0.225 (0.154)	0.192 (0.152)	-0.157 (0.197)
BER3	0.275 (0.153)*	0.211 (0.156)	-0.492 (0.241)**
state10	0.216 (0.127)*		0.078 (0.186)
state1112			0.068 (0.143)
state67			0.367 (0.120)***
state7	0.263 (0.101)***	0.259 (0.091)***	
state9	0.207 (0.108)*		0.085 (0.181)
industry6			-0.120 (0.126)
a3034	0.130 (0.104)	-0.276 (0.122)**	
a3539	0.124 (0.114)	0.046 (0.110)	0.373 (0.129)***
a4044		-0.075 (0.132)	0.368 (0.134)***
a4049	0.274 (0.097)***		
a4549		-0.086 (0.139)	0.311 (0.147)**
a5055	-0.304 (0.173)*	-0.276 (0.138)**	0.071 (0.163)
urtb	-0.015 (0.016)		0.275 (0.130)**
urtbsq			-0.013 (0.007)**
densq		0.012 (0.034)	0.087 (0.040)**
denst	-0.013 (0.038)	-0.074 (0.060)	-0.211 (0.080)***
f_BER3			1.107 (0.260)***
f_industry6			0.338 (0.204)*
f_a4055		0.236 (0.156)	
f_a5055			0.518 (0.223)**
f_logearn	0.141 (0.128)	0.109 (0.111)	
f_qual_m	0.670 (0.221)***		
f_uentry	0.018 (0.010)*		
female	-1.233 (0.559)**	0.014 (0.399)	-0.977 (0.219)***
foreign		0.070 (0.100)	-0.202 (0.138)
logearnsq	-0.009 (0.010)		0.004 (0.011)
m_logearn		0.152 (0.089)*	
m_pretx35			0.362 (0.137)***
married			-0.215 (0.097)**
pearn		-0.794 (0.420)*	
preex12			-0.171 (0.117)
preex24	0.119 (0.079)		
preex60cumst		-0.014 (0.041)	0.119 (0.058)**
preex60sq	-0.028 (0.046)		
pretx35	0.240 (0.091)***		
qual_m		0.205 (0.093)**	
uentry	-0.007 (0.007)		-0.012 (0.007)*
y93q2		-0.251 (0.148)*	
y93q3		-0.065 (0.127)	
y93q4		-0.162 (0.132)	
y94q1		-0.117 (0.132)	
y94q2		-0.175 (0.150)	
y94q3		-0.125 (0.139)	
y94q4		-0.319 (0.155)**	
_cons	-2.824 (0.264)***	-2.391 (0.380)***	-3.505 (0.686)***
N	24325	13853	9330

Treatment PF vs Waiting, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	19	19	19	20
Stratum 2	23	24	25	25
Stratum 3	23	25	25	26

Treatment PF vs Waiting, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	14	18	19	20
Stratum 2	17	17	19	25
Stratum 3	20	22	24	26

Treatment SPST vs Waiting, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER1			0.481 (0.169)***
BER2	-0.124 (0.073)*	-0.088 (0.082)	0.039 (0.077)
BER3	0.270 (0.071)***	0.110 (0.083)	0.151 (0.080)*
state10	0.199 (0.070)***	0.222 (0.085)***	
state11	-0.123 (0.063)*	0.031 (0.069)	
state12	-0.027 (0.055)	0.096 (0.069)	
state6	-0.047 (0.075)	-0.003 (0.088)	
state7	-0.100 (0.065)	0.021 (0.073)	
state9	-0.118 (0.073)	-0.128 (0.085)	
industry3	0.147 (0.076)*	-0.058 (0.086)	0.038 (0.080)
industry4	-0.010 (0.090)	0.076 (0.094)	0.007 (0.091)
industry5	-0.057 (0.094)	-0.377 (0.120)***	-0.223 (0.106)**
industry6	0.073 (0.072)	-0.003 (0.080)	0.026 (0.076)
industry7	-0.127 (0.079)	-0.195 (0.088)**	-0.085 (0.081)
a3034	0.019 (0.054)	0.097 (0.063)	0.155 (0.062)**
a3539	-0.065 (0.069)	0.162 (0.068)**	0.189 (0.075)**
a4044		-0.483 (0.157)***	-0.283 (0.119)**
a4049	-0.085 (0.064)		
a4549		-0.654 (0.162)***	-0.482 (0.128)***
a5055	-0.460 (0.082)***	-0.891 (0.161)***	-0.914 (0.128)***
urtb			-0.014 (0.008)*
densq		0.036 (0.018)**	
denst		-0.033 (0.037)	
earncens	0.023 (0.190)	0.362 (0.335)	-0.298 (0.306)
f_industry5		0.851 (0.235)***	
f_a2534	-0.192 (0.078)**		-0.402 (0.107)***
f_a3544			-0.185 (0.109)*
f_for_a2539		-0.181 (0.093)*	

SPST vs Waiting, Cohort 93/94 West Germany – continued			
	Stratum 1	Stratum 2	Stratum 3
f_married			-0.183 (0.066)***
f_qual_h			-0.623 (0.177)***
f_qual_m			-0.249 (0.086)***
female	0.034 (0.058)	0.009 (0.081)	0.428 (0.118)***
for_a2534	-0.301 (0.098)***		
for_a2539		-0.205 (0.084)**	
for_a3544	-0.209 (0.119)*		
foreign			-0.331 (0.062)***
frmsize2	0.103 (0.046)**		
frmsize3	0.225 (0.068)***		
frmsize4	0.192 (0.064)***		
logearn	-0.010 (0.036)	0.104 (0.070)	-0.001 (0.063)
married		-0.111 (0.047)**	
pearn		-0.355 (0.319)	0.484 (0.306)
preex12	0.138 (0.049)***	0.087 (0.060)	
preex60cumst		0.023 (0.030)	0.020 (0.025)
preex60sq	0.050 (0.025)**	-0.028 (0.030)	
pretx1		-0.033 (0.125)	0.105 (0.122)
pretx2		0.252 (0.097)***	0.021 (0.101)
pretx35		0.103 (0.068)	0.239 (0.063)***
qual_h		0.133 (0.114)	
qual_h_a4055		0.578 (0.203)***	0.571 (0.169)***
qual_m		-0.026 (0.073)	
qual_m_a4055		0.571 (0.145)***	0.487 (0.104)***
y93q2		0.109 (0.110)	-0.011 (0.082)
y93q3		0.225 (0.103)**	0.058 (0.079)
y93q4		0.400 (0.098)***	0.104 (0.079)
y94q1		0.512 (0.096)***	0.124 (0.079)
y94q2	0.390 (0.055)***	0.507 (0.100)***	0.097 (0.084)
y94q3	0.311 (0.056)***	0.567 (0.097)***	0.160 (0.081)**
y94q4	0.409 (0.052)***	0.554 (0.097)***	0.021 (0.084)
_cons	-2.308 (0.168)***	-2.229 (0.229)***	-1.913 (0.241)***
N	24751	14232	9913

Treatment SPST vs Waiting, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	29	30	30	31
Stratum 2	42	44	44	45
Stratum 3	33	37	37	37

Treatment SPST vs Waiting, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	29	30	30	31
Stratum 2	37	40	43	45
Stratum 3	34	36	36	37

Treatment RT vs Waiting, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER2	0.312 (0.103)***	0.048 (0.124)	0.215 (0.145)
BER3	0.017 (0.107)	0.111 (0.127)	0.234 (0.152)
state11		-0.211 (0.106)**	
state1112	-0.174 (0.072)**		
state12		-0.210 (0.108)*	
industry3	-0.089 (0.108)		
industry4	-0.096 (0.122)		
industry5	-0.253 (0.129)*		
industry67	0.050 (0.092)		
a2529		1.028 (0.188)***	0.424 (0.114)***
a3034	-0.042 (0.069)	1.021 (0.181)***	0.605 (0.109)***
a3539		0.811 (0.181)***	0.559 (0.119)***
a3544	-0.556 (0.156)***		
a4044		0.612 (0.184)***	
a4549	-0.497 (0.135)***		
a5055	-0.822 (0.155)***		
urtb	0.019 (0.013)		
denst	-0.013 (0.030)	-0.083 (0.035)**	
f_age		0.016 (0.011)	
f_preex12		0.594 (0.197)***	
f_preex60cumst	-0.078 (0.055)	-0.285 (0.081)***	
f_qual_m	0.331 (0.133)**		
female	-0.216 (0.116)*	-1.137 (0.397)***	
for_age	0.015 (0.012)	-0.009 (0.003)**	
foreign	-0.951 (0.447)**		-0.227 (0.115)**
frmsize2	0.121 (0.068)*		
frmsize34	0.267 (0.078)***		
logearnsq			0.012 (0.010)
m_preex60cumst	-0.091 (0.047)*		
preex12	0.116 (0.072)		
preex24	0.181 (0.078)**		
preex60cumst		0.071 (0.046)	-0.043 (0.043)
pretx35			0.310 (0.103)***
qual_h		-0.462 (0.209)**	
qual_m	-0.251 (0.086)***		
qual_m_a3544	0.462 (0.166)***		
uentry	0.015 (0.014)	0.012 (0.018)	
uentry2	-0.001 (0.001)	-0.001 (0.001)	
y94q34			-0.260 (0.108)**
_cons	-2.722 (0.216)***	-3.049 (0.224)***	-2.976 (0.196)***
N	24421	13889	9350

Treatment RT vs Waiting, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	25	26	26	27
Stratum 2	15	17	18	18
Stratum 3	8	10	10	10

Treatment RT vs Waiting, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	23	23	25	27
Stratum 2	12	12	12	18
Stratum 3	7	9	10	10

Treatment PF vs SPST, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER2	0.634 (0.204)***		
BER3			-0.708 (0.207)***
state10	0.185 (0.248)	-0.552 (0.329)*	
state11	-0.101 (0.274)	-0.585 (0.260)**	
state12	0.232 (0.215)	-0.033 (0.207)	
state6	0.062 (0.283)	0.163 (0.267)	0.277 (0.207)
state7	0.743 (0.216)***	0.343 (0.212)	0.627 (0.162)***
state9	0.665 (0.244)***	0.439 (0.252)*	
a2529	-0.356 (0.162)**		
a3034		-0.535 (0.245)**	
a3555		0.056 (0.165)	
a4044			0.188 (0.174)
a4549			0.327 (0.203)
a5055			0.718 (0.194)***
urtb100	-5.813 (2.990)*		
denst		-0.222 (0.068)***	
f_BER2	-0.303 (0.330)		
f_BER3			1.665 (0.340)***
f_a3034		-0.272 (0.411)	
f_logearn		-2.979 (1.379)**	
f_logearnsq		0.547 (0.195)***	
f_preex12	-0.295 (0.312)		
f_preex24	0.196 (0.242)		
f_preex6	-0.063 (0.310)		
f_qual_u	-1.233 (0.455)***		
f_tnull			-0.006 (0.018)
female	0.671 (0.331)**	2.948 (2.640)	-7.643 (8.579)
foreign		0.602 (0.201)***	
logearn		1.923 (1.307)	
logearnsq		-0.320 (0.170)*	
m_tnull			-0.023 (0.012)*
preex60sq	-0.175 (0.084)**		
qual_u	0.523 (0.233)**		
y93q2	0.196 (0.271)	-0.561 (0.317)*	
y93q3	0.208 (0.263)		
y93q34		-0.458 (0.241)*	
y93q4	-0.118 (0.271)		
y94q1	0.206 (0.269)	-0.797 (0.267)***	
y94q2	-0.297 (0.255)	-0.989 (0.292)***	
y94q3	-0.205 (0.249)	-0.847 (0.269)***	
y94q4	-0.681 (0.268)**	-1.246 (0.293)***	
_cons	-0.853 (0.399)**	-2.614 (2.603)	7.908 (4.722)*
N	630	583	755

Treatment PF vs SPST, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	24	24	24	24
Stratum 2	22	22	22	22
Stratum 3	10	10	10	10

Treatment PF vs SPST, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	23	24	24	24
Stratum 2	21	22	22	22
Stratum 3	7	9	10	10

Treatment PF vs RT, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER3	0.476 (0.197)**	-0.540 (0.308)*	
state10		0.224 (0.440)	
state11	0.323 (0.330)	0.061 (0.361)	
state12	0.748 (0.246)***	0.698 (0.310)**	
state6	0.480 (0.340)	0.623 (0.360)*	
state7	0.615 (0.242)**	0.810 (0.275)***	
state9		0.844 (0.347)**	
state910	0.602 (0.247)**		
industry7		-0.724 (0.246)***	
a3034		-0.228 (0.257)	
a3539	0.281 (0.228)		
a3544		0.608 (0.239)**	
a4044	0.654 (0.224)***		
a4549	0.710 (0.408)*	1.763 (0.450)***	
a5055	0.412 (0.453)	2.046 (0.631)***	
f_BER3		1.427 (0.481)***	
f_BER34			1.923 (0.524)***
f_state10			-0.236 (1.087)
f_state11			-0.768 (1.137)
f_state12			0.781 (0.754)
f_state6			0.054 (0.959)
f_state7			0.448 (0.744)
f_state9			0.206 (0.898)
f_a3539			-0.410 (0.592)
f_a4044		0.976 (0.536)*	
f_a4055			1.568 (0.546)***
f_a4549	0.818 (0.652)		
f_married			-0.503 (0.466)
f_qual_m	1.091 (0.487)**		
female	-1.111 (0.481)**	-0.804 (0.330)**	-0.353 (0.926)
foreign	0.502 (0.307)	0.602 (0.276)**	
m_BER25			1.370 (0.408)***
m_state10			0.971 (0.659)
m_state11			0.165 (0.421)
m_state12			-1.058 (0.642)*
m_state6			0.460 (0.501)
m_state7			0.984 (0.407)**
m_state9			-0.066 (0.510)
m_a3539			0.955 (0.351)***
m_a4055			2.325 (0.485)***
m_married			-1.142 (0.384)***
tnull		-0.018 (0.015)	
_cons	-1.212 (0.191)***	6.714 (5.956)	-1.844 (0.511)***
N	300	240	192

Treatment PF vs RT, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	13	13	13	14
Stratum 2	14	16	17	17
Stratum 3	20	20	21	21

Treatment PF vs RT, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	13	14	14	14
Stratum 2	15	17	17	17
Stratum 3	17	20	20	21

Treatment RT vs SPST, Cohort 93/94 West Germany			
	Stratum 1	Stratum 2	Stratum 3
BER2	0.808 (0.122)***		
state10	-0.512 (0.199)**	-0.384 (0.234)	
state11	-0.312 (0.186)*	-0.338 (0.194)*	
state12	-0.404 (0.151)***	-0.523 (0.185)***	
state79			0.278 (0.133)**
a2529		1.612 (0.505)***	
a3034		1.511 (0.507)***	
a3539	-0.241 (0.154)		
a3544			-0.323 (0.131)**
a3549		1.046 (0.502)**	
a4044	-0.286 (0.169)*		
a4549	-0.725 (0.233)***		
a4555			-0.799 (0.196)***
a5055	-0.896 (0.283)***		
f_BER2		0.743 (0.242)***	0.473 (0.215)**
f_married	-0.137 (0.179)		
f_preex60cumst		-0.180 (0.117)	
f_qual_m		0.382 (0.260)	
f_qual_u	-0.490 (0.297)*		
female	0.502 (0.163)***	-0.893 (0.303)***	-0.097 (0.164)
logearn			0.199 (0.127)
m_married	0.360 (0.148)**		
m_qual_h		-1.020 (0.359)***	
m_qual_m		-0.303 (0.196)	
married			0.354 (0.174)**
marriedBER2			-0.302 (0.217)
pearn			-0.658 (0.797)
qual_h			-0.458 (0.292)
qual_u	0.521 (0.183)***		
y93q2	-0.187 (0.250)	-0.427 (0.286)	
y93q3	0.237 (0.218)	-0.480 (0.265)*	
y93q4	-0.111 (0.224)	-0.630 (0.250)**	
y94q1	0.084 (0.220)	-0.567 (0.245)**	
y94q2	-0.215 (0.206)	-0.602 (0.270)**	
y94q3	-0.289 (0.214)	-0.933 (0.268)***	
y94q34			-0.340 (0.155)**
y94q4	-0.481 (0.209)**	-1.123 (0.270)***	
_cons	-0.900 (0.199)***	-0.949 (0.545)*	-1.088 (0.628)*
N	726	619	775

Treatment RT vs SPST, Cohort 93/94 West Germany, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	20	20	20	20
Stratum 2	17	18	18	19
Stratum 3	10	11	11	11

Treatment RT vs SPST, Cohort 93/94 West Germany, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	17	18	20	20
Stratum 2	17	18	19	19
Stratum 3	11	11	11	11

Common Support

Figure 8: Densities of Propensity Scores for Cohort 86/87

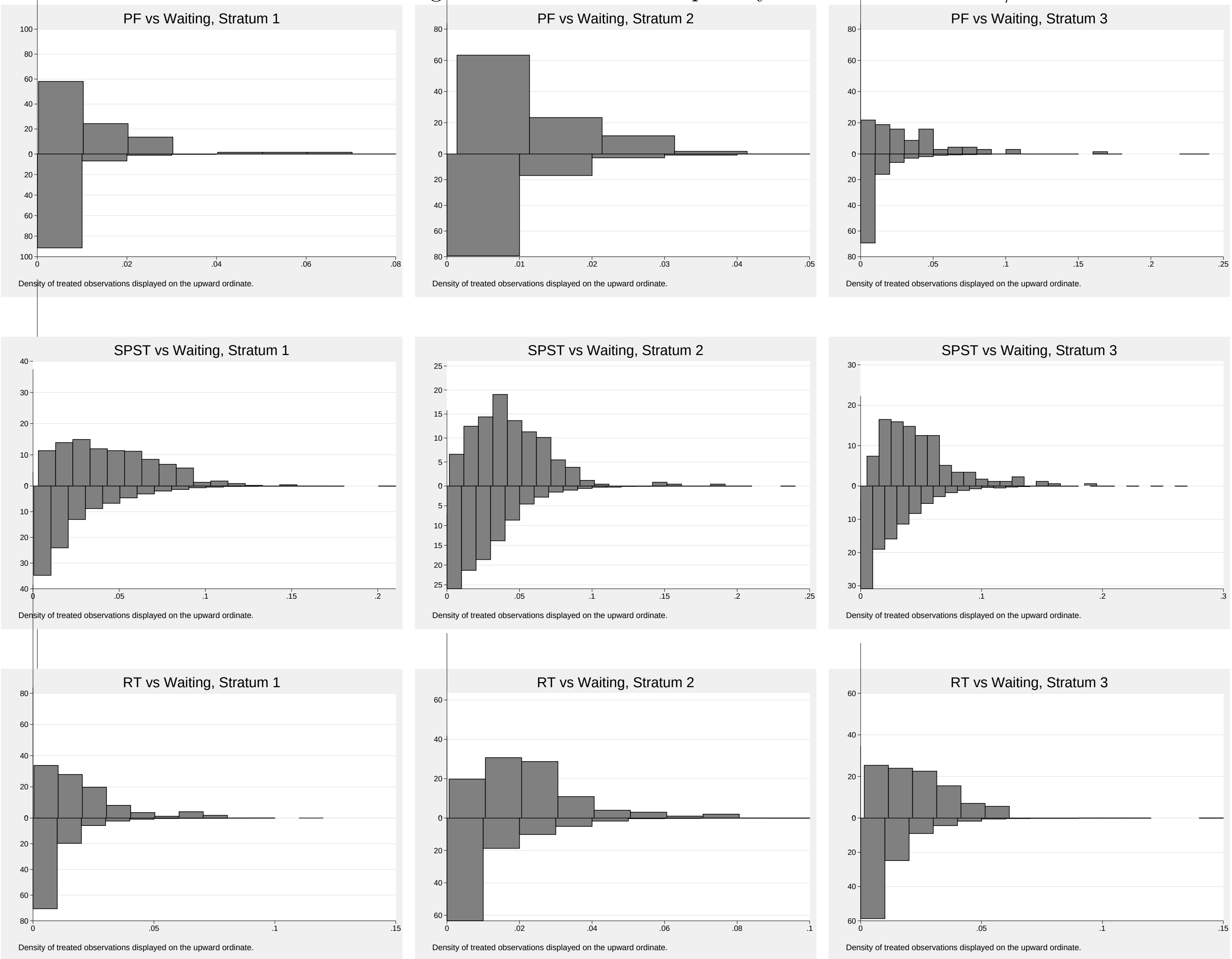


Figure 9: Densities of Propensity Scores for Cohort 86/87

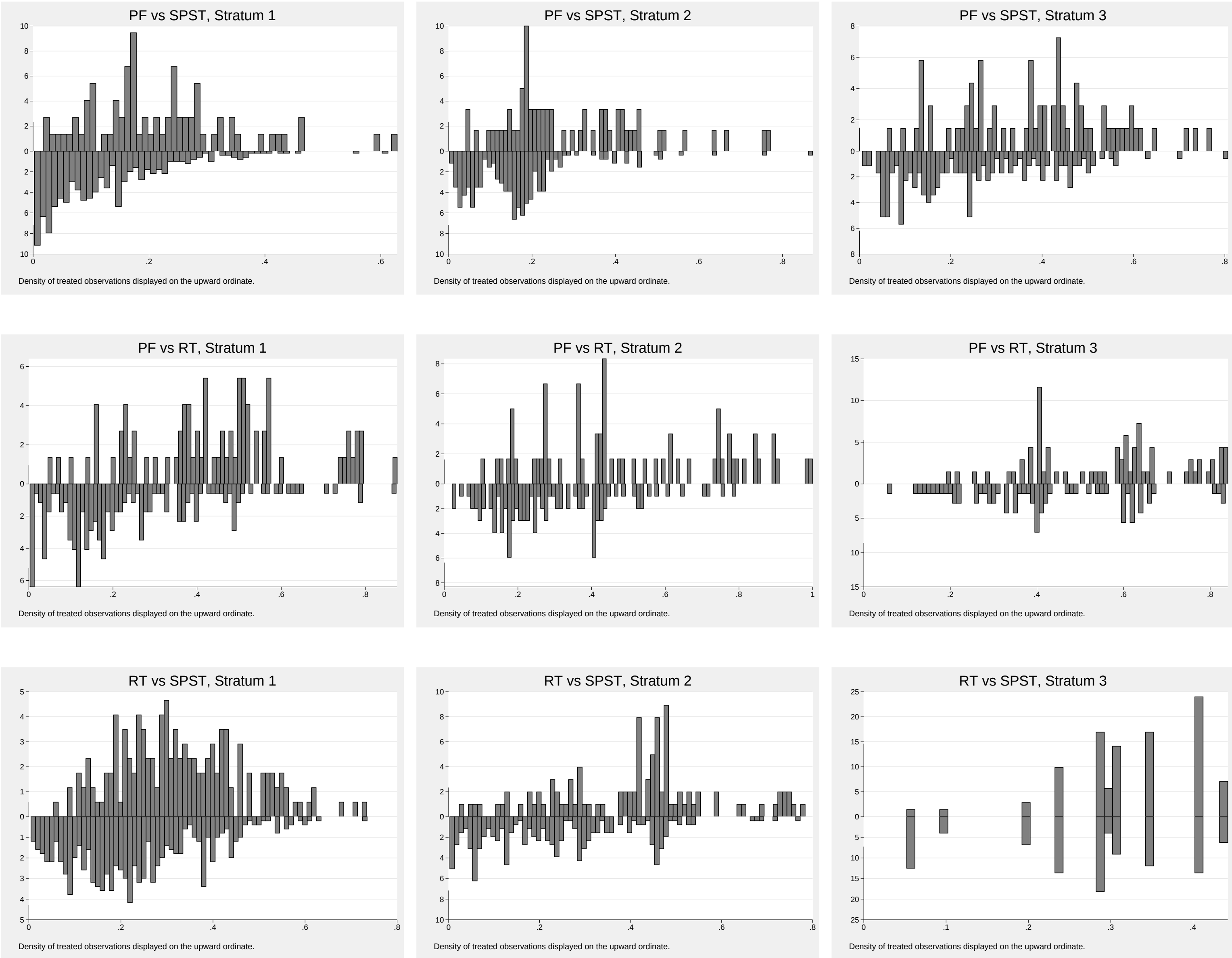


Figure 10: Densities of Propensity Scores for Cohort 93/94

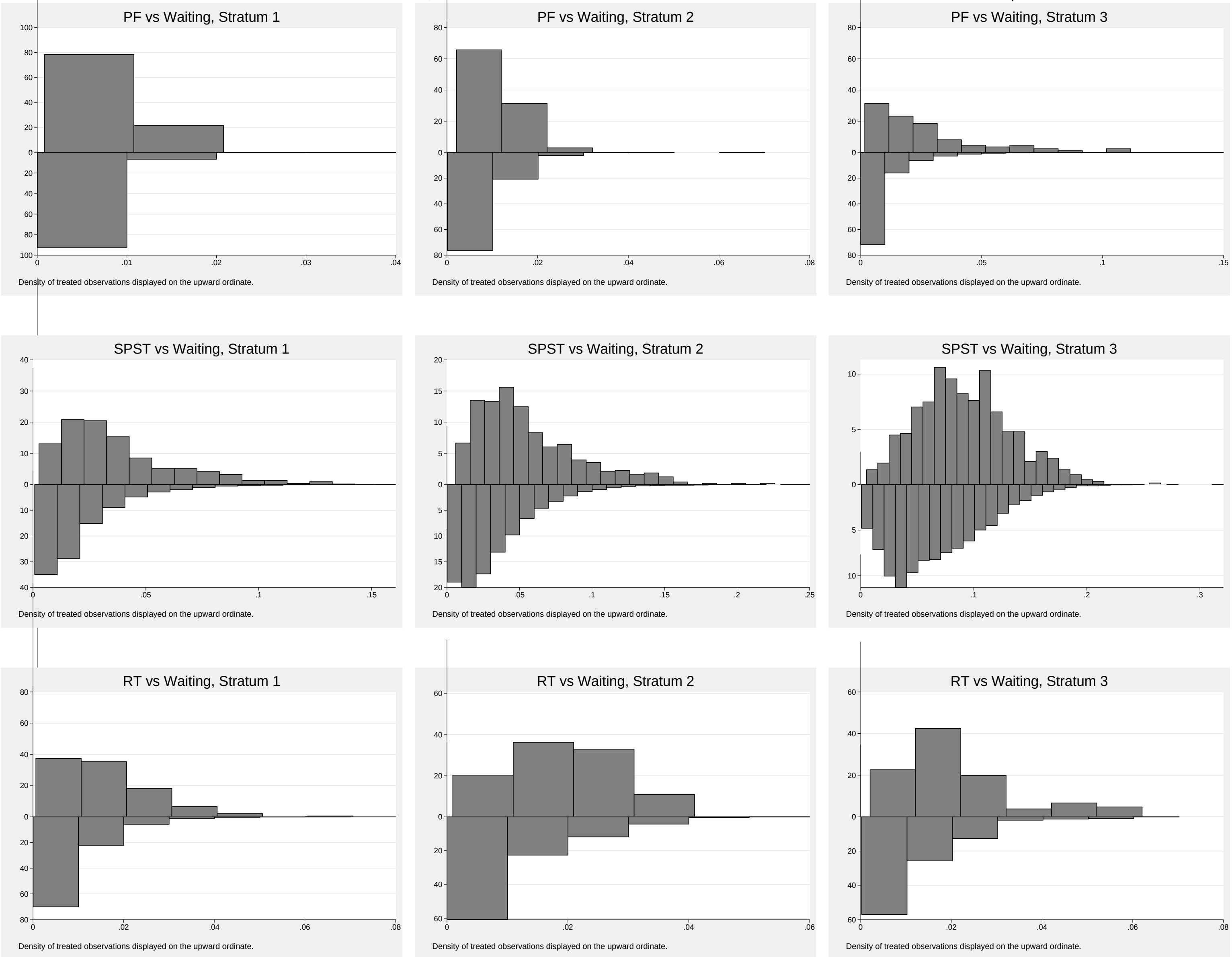
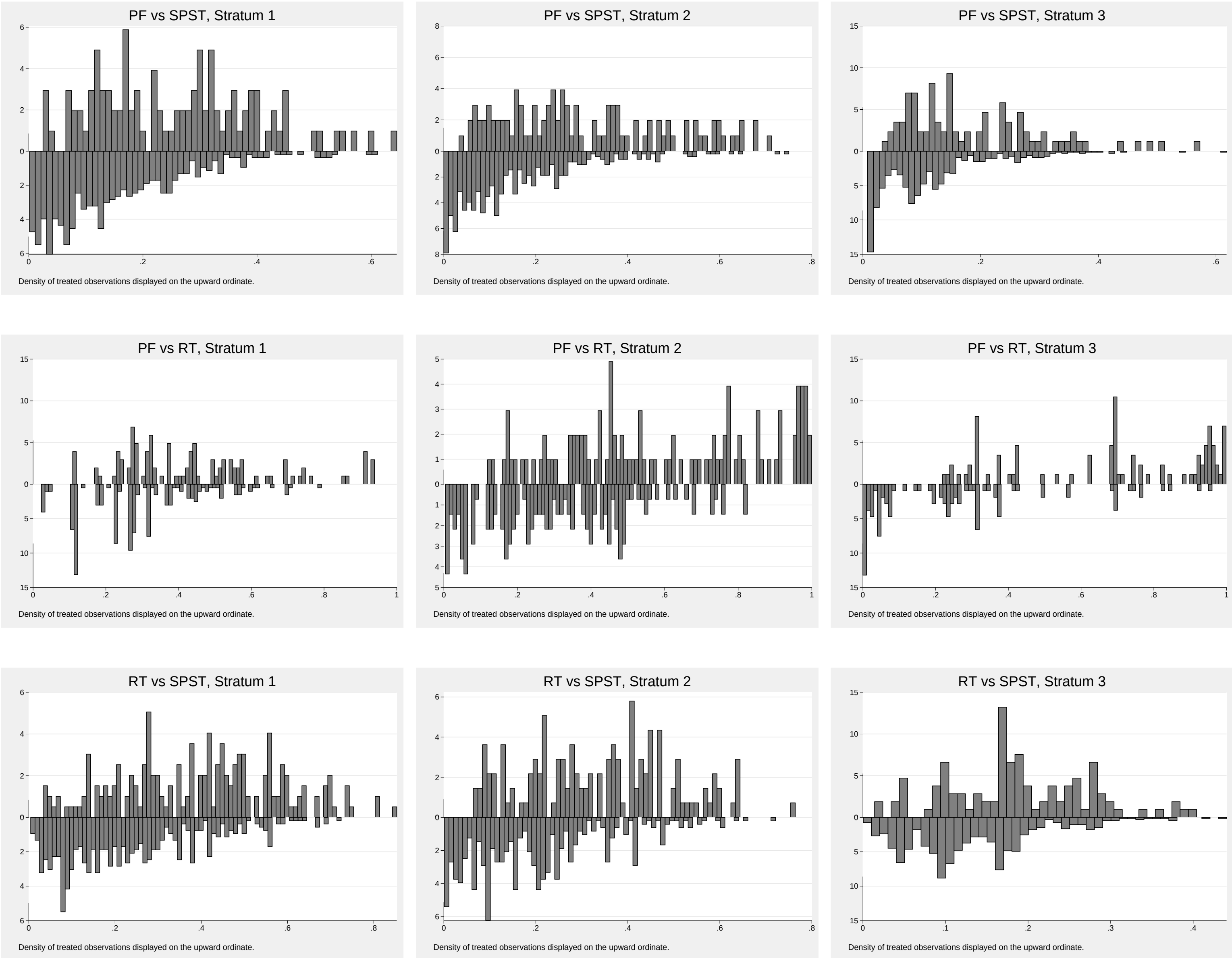


Figure 11: Densities of Propensity Scores for Cohort 93/94



Background Information about the Data

Types of Further Training under the Labor Promotion Act

In this study we are interested in active labor market programs for unemployed who have previously been employed and who have not already found a new job. Here however, we want to give a short overview over the full set of training schemes administered under the Labor Promotion Act (Arbeitsförderungsgesetz, AFG).

Further Vocational Training (Berufliche Fortbildung)
A bunch of different training courses is subsumed under this heading. It includes theoretical as well as practical training schemes within the occupation of the participant.

Retraining (Umschulung)
This training scheme provides a complete, new vocational degree according to the German apprenticeship system.

Short–term Training according to §41a AFG
These programs last only about four weeks and were offered from 1979 until 1992. They are mainly intended to evaluate the participant's problems in finding regular employment. Starting 1993 such programs are no longer recorded as independent programs but as part of the regular counseling for unemployed. Hence we can not identify them in the inflow sample 1993/94. In order to make the samples comparable we treat the programs according to §41a in the 1986/87 inflow sample also as open unemployment. Thus if an unemployed first takes part in a §41a program and later in the same unemployment spell in Retraining we would consider the retraining the first program and evaluate it.

German Language Course (Deutschsprachlehrgang)
The German Language Courses are intended for newly arrived immigrants. So the participants typically have not been employed in Germany before the German Course and hence are not part of the focus group of this study, the previously employed unemployed.

Career Advancement (Aufstiegsförderung)
These programs are typically targeted at the employed and have been more important when the Labor Promotion Act was introduced in 1969. By providing

additional human capital the participant's risk of becoming unemployed should be lowered. Prime examples are courses in which the participants with a vocational training degree obtain additional certificates which allow them to independently run craftsman's establishments and to train trainees in the dual system of vocational training.

Integration Subsidy (Einarbeitungszuschuss)
Wage subsidies are paid for the employment of formerly long–term unemployed and are intended to decrease the competitive disadvantage of these recruits for the period of familiarization with the skill requirement of the job. Even if the target group of wage subsidies are also unemployed we do not evaluate them because they require a job for which the wage subsidy is paid. This means provision of wage subsidies is already conditional on employment which is the success criteria for the other programs.

Construction of the Monthly Panel

The IABS employment and LED benefit payment data are daily register data whereas the FuU training data gives monthly information about program participation. This study uses the merged data as described in Bender et al. (2005). From the merged data we construct a monthly panel. If the original daily data contain more than one spell overlapping a specific month we take the information from the spell with the largest overlap as the spell defining the monthly information.

The defining condition to be part of our inflow sample into unemployment is a transition from an employment month to a nonemployment month, in which the last employment month was between December 1985 (1992) and November 1986 (1993) and thus the first unemployment month was between January 1986 (1993) and December 1987 (1994). In order to divide nonemployment (to be precise: not employed subject to social security contributions) into unemployment and other states (like labor market leavers, transition into self employment, employment as civil servant) we additionally require a month with benefit payments from the employment office within the first twelve month of nonemployment or indication of participation in any labor market program in one of our data to be part of the inflow sample in unemployment.

Later on we aggregate the information further from monthly to quarterly information. Whereas the monthly employment information is binary the quarterly employment information can take the values 0, 1/3, 2/3 or 1.

We identify program participation if a person starts a program while being in the defining unemployment spell. The participant must not be employed in in the first month of the program. Otherwise we would consider such a program as a program which starts together with a job which we do not evaluate. In this case we would treat such a person as being employed. The exact identification of the program types will be explained in the following.

Identifying Program Participation from the Data

We identify participation in a further training program from a combination of FuU training data information, the benefit payment information and the employment status information. In principle, every participant in a further training program should be recorded in the FuU training data and we would not need the benefit payment data for identification of participation. There are two reasons to use the benefit payment data as well. First, we find the training data to be incomplete, many recipients of training related benefits are not contained in the training data.[26] Only using the benefit payment data identifies these participants. Second, quite often the type of training in the training data is given very unspecific as "Other adjustment of working skills". The benefit payment data can give more information about these programs. Finally, we need the employment status to identify participation because we only evaluate programs which start while being unemployed. In particular we do not consider integration subsidies which are associated with a regular job. We exclude programs starting together with a job because our outcome variable is employment and program starts that are conditional on having found a job are (partly) endogenous.

In the remaining part of this section we describe how we aggregated the benefit payment information and the training data information. The next section contains the exact coding plan. We disclose in detail which combination of information from benefit payment and training data we identify as PF, SPST or RT.[27]

[26] Remember the purpose of the training data was only internal documentation. This might explain its incompleteness.

[27] More details about the benefit payment data and training data can be found in Speckesser (2004), Fitzenberger and Speckesser (2007) and Bender et al. (2005).

Benefit Payment Information in the LED–Data

The merged data we use contain three variables with benefit payment information from the original LED data, ("parallel original benefit information 1–3" [*Leistungsart im Original 1–3*] L1LA1, L2LA1, L3LA1). The main variable is L1LA1. If there are two parallel payment informations in the original data L1LA2 also contains information and only if there is a third parallel payment spell L3LA1 is also filled. In general we use L1LA1. Only if L1LA1 is not informative about program participation and L2LA1 is we use L2LA1 and only if L1LA1 and L2LA1 are not informative but L3LA1 we use L3LA1. The benefit payment information is given in time varying three–digit codes (for the coding plan see Bender et al. 2005). We extracted the program related information from the benefit payment information as given in table 13. The main distinction regarding program participation is the distinction between no benefits at all or unemployment benefits/assistance on the one hand and program related maintenance benefits on the other hand. There are five types of program related benefits. Most important for us are the more general maintenance benefits while in further training and the more specific maintenance benefits while in retraining.

Table 13: Aggregated types of benefit payment

German Abbreviation	Description
ALG	Unemployment benefits
ALHi	Unemployment assistance
UHG §41a	Income maintenance while in specific short term training program
UHG Fortbildung	Income maintenance while in further training
UHG Umschulung	Income maintenance while in retraining
UHG Darlehen	Income maintenance as a loan
UHG Deutsch	Income maintenance while in a German course

The original benefit payment information is given in three variables L1LA1, L2LA1 and L3LA1 with time varying three–digit codes.

Training Types in the FuU–Data

In this evaluation study one of the most important advantages compared to survey data is the information about the precise type of training. It allows us to identify homogeneous treatments for the evaluation. In the merging process, up to two parallel FuU–spells were merged to one spell of the IABS data because in many cases the FuU–data provided more than one parallel spell. These two parallel spells provide two variables indicating the type of course (*Maßnahmeart* [FMASART1, FMASART2]).

Correcting Type of Training for 1986 The annual frequency for the type of training *14* in 1986 looks very different than in the years before and after. Additionally the distributions of the planned durations and the types of examination completing the program *14* in 1986 are different than in the adjacent years. We think this is due to a lacking recoding of *14* to *12*, which was necessary for the years until 1985 because the coding of FMASART changed over the years. Hence we recode *14* to *12* in 1986 if the planned duration is less than 10 month.

Aggregating the training type information Since type of treatment (*Maßnahmeart*) is often coded as "other adjustment"
(FMASART1=12 [*Sonstige Anpassungen*]) in the FuU–data, we increase the precision of information about the type of treatment by relying on the second parallel information about the type of training: The second FuU–spell is used if the first FuU–spell is coded as "other adjustment" (*"Sonstige Anpassungen"*) and a second spell includes a code different from 12. Such combined information of FMASART1 and FMASART2 is referred to as FMASART* in the following.

Combining the Information

When using information from different sources, the sources may give differing information. If the training data indicated training participation and the benefit payment data did not or vice versa we relied on the source which indicated training for the following reasons. If somebody receives training related benefits it is more likely that the employment agency forgot to fill in the training data record than the agency wrongly induced payment of benefits. And if somebody is contained in the training data but does not receive maintenance benefits he either receives no benefits, which is possible while being in training, or receives unemployment benefits/assistance and the payment is just wrongly labeled.

If both training and benefit payment data indicate program participation but differ in the type of program we generally use the training data information. An example: the benefit payment indicates maintenance payments for further training and the training data indicates Retraining. We use Retraining from the training data.

The only exception is unspecific program information from the training data "other adjustment". If in such cases the benefit payment data give specific information like Retraining we use the information from benefit payment data. All possible combinations of training and benefit payment information which we use to identify participation in one of the three programs are given in the following section.

Coding Plan for the Treatment Information

This section gives the exact coding plans for identification of Practice Firm, SPST and Retraining. In general we identify program participation as start of a program in an unemployment spell before another employment begins. This means that we only identify a start of a program if the employment status in the first month of the program indicates no employment (BTYP≠1).

Practice Firm

Practice Firm is a consolidation of the program types Practice enterprise and Practice studio from the FuU training data. There is no specific benefit payment type related to Practice Firms, rather the participants shall receive the general maintenance payment for further training. Since the training data are more reliable than the benefit payment data regarding type of the program we identify Practice Firm whenever FMASART shows the codes 11 or 12 independently of the payment information.

Program code	Label	Label in German
10	Practice enterprise	Übungsfirma
11	Practice studio	Übungswerkstatt

In table 14 we show how often which combination of benefit payment information and program type information identifies *Practice Firm* in the two inflow samples.

Provision of Specific Professional Skills and Techniques

We identify SPST in the following cases.

Table 14: Identification of *Practice Firm* with program type and benefit payment type: Frequencies

Program	Type of payment No benefits	UB/UA	Income Maintenance for Short–term Training	Further Training	Retraining	Total
Practice enterprise	4	5	1	198	2	210
Practice studio	11	19	0	311	20	361
Total	15	24	1	509	22	571

Both inflow samples together. BTYP≠1 as an additional requirement.

(a) Identification from training data and benefit payment data

We identify SPST if the training data indicates the general program "Other adjustment" and the benefit payment information is no benefit payments, unemployment benefits, unemployment assistance or maintenance payments while in retraining.

Program code	Label	Label in German
12	Other adjustment of working skills	sonst. Anpassung der berufl. Kenntnisse

(b) Reliance on benefit payment data

We identify SPST if the program information from the training data is missing and the benefit payment information is maintenance payments while in further training.

Program code	Label	Label in German
-9	missing	fehlende Angabe

(c) Additional program from training data

We also identify SPST when another program of little quantitative importance but comparable content is recorded in the training data independent of the benefit payment information.

Program code	Label	Label in German
31	Further education of trainers and multidisciplinary qualification	Heran-/Fortbildung v. Ausbildungskräften/ berufsfeldübergreifende Qualifikation

(d) Additional combinatioin

Finally we identify SPST if the training data indicate the unspecific "other career advancement" and the benefit payment information indicates further training.

Program code	Label	Label in German
28	Other promotion	sonstiger Aufstieg (< 97)

In table 15 we show how often which combination of benefit payment information and program type information identifies *SPST* in the two inflow samples.

Table 15: Identification of *SPST* with program type and benefit payment type: Frequencies

Program	Type of payment No benefits	UB/UA	Income Maintenance for further training	Total
Missing	0	0	644	644
Other adjustment of working skills	57	89	2095	2241
Other promotion	0	0	150	150
Further education of trainers and multidisciplinary qualification	0	1	1	2
Total	57	90	2890	3037

Both inflow samples together. BTYP$\neq$1 as an additional requirement.

Retraining

Retraining or longer "Qualification for the first labor market via the education system" is taking part in a new vocational training and obtaining a new vocational training degree according to the German dual education system. Additionally, but quantitatively of little importance we see the make up of a missed examination "Certification" as comparable to retraining because the result is the same. Furthermore and also only of marginal importance we see participation in the programs "Technican" or "Master of Business administration (not comparable to an american style MBA)" while not receiving maintenance benefits as a loan as Retraining. Conventionally these two programs are considered as career advancement programs which we do not evaluate. Benefits as a loan would underline their character as career advancements.

(a) Identification from training data

We identify the following two programs as Retraining independent of the benefit payment information.

Program code	Label	Label in German
29	Certification	berufl. Abschlussprüfung
32	Retraining	Umschulung

(b) Reliance on benefit payment data

If the training data is uninformative and maintenance benefits for Retraining are paid we identify Retraining.

Program code	Label	Label in German
-9	missing	fehlende Angabe
12	Other adjustment of working skills	sonst. Anpassung der berufl. Kenntnisse

(c) Other programs from training data

Two other programs are identified from the training data. They typically also take two years full time and require an existing vocational training degree, hence are somewhat comparable to retraining in a narrower definition. Not identified if maintenance benefits are paid as a loan.

Program code	Label	Label in German
26	Technician	Techniker (<97)
27	Master of business administration	Betriebswirt (<97)

In table 16 we show how often which combination of benefit payment information and program type information identifies *Retraining* in the two inflow samples.

Table 16: Identification of *Retraining* with program type and benefit payment type: Frequencies

	Type of payment					
	No benefits	UB/UA	Income Maintenance			
Program			Further Training	Retraining	Loan	Total
missing	0	0	0	110	0	110
Other adjustment of working skills	0	0	0	65	0	65
Technician	2	1	5	2	0	10
Master of business administration	0	2	1	1	0	4
Certification	4	1	20	7	0	32
Retraining	11	13	231	355	2	612
Total	17	17	257	540	2	833

Both inflow samples together. BTYP≠1 as an additional requirement.