The Comparative Effectiveness of Public Sponsored Training Revisited: The Merits of Using Rich Administrative Data[1]

MARTIN BIEWEN,[♭] BERND FITZENBERGER,[♮] ADERONKE OSIKOMINU,[♯] and MARIE PAUL[§]

[♭]University of Tübingen, IZA, DIW, martin.biewen@uni-tuebingen.de

[♮]Albert-Ludwigs-University Freiburg, ZEW, IZA, IFS, bernd.fitzenberger@vwl.uni-freiburg.de

[♯]Albert-Ludwigs-University Freiburg, IZA, osikominu@vwl.uni-freiburg.de

[§]née Waller, Albert-Ludwigs-University Freiburg, marie.paul@vwl.uni-freiburg.de

Correspondence:

Dept. of Economics, Albert-Ludwigs-University Freiburg, 79085 Freiburg, Germany

This version: June 26, 2010

Abstract: We use a new and exceptionally rich administrative data set for Germany to evaluate the employment and earnings effects of three different public sponsored training programs in the early 2000s. Building on the work of Sianesi (2004, 2008), we employ propensity score matching methods in a dynamic, multiple treatment framework to address dynamic selection into heterogeneous programs. We carefully asses to what extent various aspects of our analysis such as conditioning on employment and benefit histories in a flexible way, the availability of rich personal information, the availability of data about other programs, and the amount of specification search affect the estimates. Our results imply that training programs lead to employment and earnings gains for individuals who have been unemployed for some time. Data and specification issues can have a large effect on impact estimates. There is no evidence for effect heterogeneity.

Keywords: multiple treatments, dynamic treatment effects, local linear matching, active labor market programs, administrative data

JEL: C 14, J 68, H 43

[1]This study is part of the project "Employment effects of further training programs 2000-2002 – An evaluation based on register data provided by the Institute of Employment Research, IAB (Die Beschäftigungswirkungen der FbW-Maßnahmen 2000-2002 auf individueller Ebene – Eine Evaluation auf Basis der prozessproduzierten Daten des IAB)" (IAB project number 6-531.1A). The project is joint 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 and material support by the IAB. We thank Peter Mueser and participants in numerous seminars for helpful comments and we are especially grateful to Stefan Bender for his efforts to make available the data on which this study is based. We are indebted to Jeffrey Smith and two referees for many useful comments and suggestions. The usual caveat applies.

1 Introduction

There has been an enormous interest in the evaluation of active labor market policies, both in the US and Europe.[2] While, due to methodological and data limitations, earlier studies typically focused on the evaluation of a single program, recent developments in evaluation methodology and data access allow researchers to gain deeper insights into the possibly very heterogeneous effects of different types of programs and their comparative effectiveness.[3] This progress has been made possible both by methodological developments, in particular the extension of propensity score matching methods to the case of multiple treatments (Imbens (2000), Lechner (2001)), and by the increasing availability of large, administrative data sets that provide the necessary sample sizes and the program information to carry out in-depth evaluations of narrowly defined sub-programs. Given these rich data sources, it is possible not only to evaluate the differential effects of the classical instruments of active labor market policy such as public employment services, job creation in the public sector, or public training programs, but also to evaluate different sub-programs within these categories, for example different forms of employment subsidies or different forms of public training programs as well as the heterogeneity of program effects for different subgroups. Such evidence is crucial for policy purposes, as it allows policy makers to improve the allocation of scarce public resources. Furthermore, the availability of rich administrative data over a long time period allows one to control in a very detailed way for the labor market history of the program participants, an issue which has not fully been explored in the literature so far (see Dolton et al. (2008) as a recent exception for the UK). This is important to justify matching estimators for treatment effects which rely on a selection on observables assumption (Heckman et al. (1998), Heckman et al. (1999), Heckman and Smith (1999)).

This paper makes two contributions. First, in view of a major policy shift that took part in the German labor market policy during the last decade and that massively shifted resources from long-term and medium-term public sponsored training programs to short-term programs and to more practically oriented programs (see Bundesagentur für Arbeit (2005b), and figure 1 below), we investigate how shorter and more practically oriented training programs compare in effectiveness to the longer and more theoretically oriented programs that were used to a large extent before

[2] For comprehensive overviews see Heckman et al. (1999), Martin (2000), Martin and Grubb (2001), Kluve and Schmidt (2002), Carcillo and Grubb (2006) and Kluve (2010).

[3] Examples of evaluations involving multiple comparisons of different programs are Lechner (2002), Gerfin and Lechner (2002), Larsson (2003), Hardoy (2005), Dyke et al. (2006), Dorsett (2006), Frölich (2008), and Sianesi (2008).

the policy shift. We thus contribute to the growing literature on the differential effects of training programs (see next section for an overview). Given the wide range of training programs offered by the German Federal Employment Office, we believe that the comparison of different forms of training used in Germany also holds lessons for other countries that employ similar programs. The specific form of short-term training is particularly interesting because, in addition to the provision of skills, it comprises elements of job search assistance as well as profiling and monitoring of the unemployed in addition to the provision of skills. We therefore contribute to the literature that has focused on these specific forms of active labor market policy (see e.g. Martin (2000), Dolton and O'Neill (2002) and OECD (2005), Dyke et al. (2006), Hotz et al. (2006), Heinrich et al. (2010)). Whether more practically oriented training programs are more effective in providing the skills needed to improve the labor market chances of the unemployed is a question of considerable policy interest. Our results are consistent with the literature (Martin and Grubb (2001) and OECD (2005)) suggesting that practically oriented training is more effective than pure classroom training. However, these results are in contrast to earlier findings for Germany during the 1990's (Fitzenberger and Völter (2007), Fitzenberger et al. (2010), and Lechner et al. (2010)).

Our evaluation study builds on the work of Sianesi (2004, 2008) on dynamic treatments and on the work of Imbens (2000) and Lechner (2001) on pairwise comparison of multiple treatments. In order to take account of dynamic sorting processes, we stratify treatment effects by elapsed duration of unemployment. Our results show that average effects for broad populations may hide statistically and economically significant treatment effects for individual subgroups and therefore help to understand why previous evaluation studies often yielded inconclusive results. A key advantage of our methodology is that it allows us to directly *compare* training programs, i.e. to ask the question of what would have happened if participants in short-term programs had participated in longer-term programs, or if participants in classroom training had taken part in more practically oriented training. This leads to more informative results than if one compares the effectiveness of different types of training when compared to not taking part in training at all. In addition, we provide formal tests of effect heterogeneity.

Our second contribution is a methodological one. We base our evaluation on a new and particularly rich administrative data set that was made available for research recently. This data set, which merges information from different administrative sources, contains not only precise and extensive information on individual employment and transfer receipt histories covering more than ten years but also comprehensive and detailed information on participation in all active labor market programs. Furthermore, it includes all the information on personal characteristics and professional goals of a particular job-seeker that is collected by the employment officer in the course of a counseling interview and that serves as a basis for future services and assignments offered to the job-seeker. The large sample sizes involved make it possible to address aspects that have been difficult or impossible to address such as the heterogeneity of programs, the heterogeneity of effects across different groups of participants and the dynamic selection into different programs.

Given the richness of our data set and the implied wide range of possible specification choices, it is a highly relevant question to what extent features of the data and specific specification choices influence evaluation outcomes. In order to address this question, we provide a detailed sensitivity analysis examining to what extent evaluation results hinge on various data and specification issues. Using different balancing tests, we first show that the information on personal characteristics and individual employment and benefit histories in our data suffice to balance individual characteristics and pre-treatment outcomes in our treatment and comparison groups. We then carry out a careful sensitivity analysis that compares our benchmark evaluation results to different variations in which we sequentially omit important aspects of our data or specification. We believe that such an analysis provides important information for the evaluation of labor market programs in other countries. For example, many data sets used for evaluation purposes lack the information about what other programs control group members might participate in (see e.g. Mueser et al., 2007). We investigate the sensitivity of the results if we proceed as if we did not have this information. Similarly, in the spirit of Card and Sullivan (1988), Heckman and Smith (1999), and Dolton et al. (2008) we examine what difference flexible conditioning on various aspects of the employment and benefit history makes. In order to assess to what extent evaluation results hinge on the availability of rich personal information, we compare our results to several specifications in which we omit important features of our data such as information about the motivation of participants based on case-worker assessments, or more detailed information on health and other characteristics which are often not available in data sets used for evaluation purposes. We also investigate to what extent the amount of specification search contributes to evaluation outcomes.

Finally, we compare our estimates, which are based on a dynamic stratification of treatment and control groups in the spirit of Fredriksson and Johansson (2008) and Sianesi (2004, 2008), to results based on a more traditional static evaluation setup in which the control group is defined as those individuals who do not receive treatment over a fairly wide observation window. Our results imply that such a static evaluation approach leads to smaller positive and larger negative estimated treatment effects. We also demonstrate that positive treatment effects do not just reflect the lock–in effect of future program participation in our control group. Our results show that both data and specification issues may have a large impact on evaluation results and that aspects such as information about other programs and stratifying individuals according to the elapsed unemployment duration are important.

The remainder of the paper proceeds as follows. Section 2 discusses the literature. A description of the programs is given in section 3. Section 4 describes the data and section 5 presents the econometric approach. The results are presented in section 6. Section 7 concludes. The appendix provides details on the data and estimation results. Further details can be found in the additional Online-Appendix, which is available upon request.

2 Literature Review

We focus on studies for training programs, in particular those using a multiple treatment framework. Studies for the US, the UK, or Canada often distinguish between alternative training programs. Barnow (1987) reviews evaluations of the US Comprehensive Employment and Training Act (CETA), distinguishing classroom training, on-the-job training, and work experience. The results of different studies were quite ambiguous. Programs were generally found to be more effective for women than for men and on-the-job training was generally found to be more effective than classroom training and work experience. Evaluation studies for the subsequent Job Training Partnership Act (JTPA) (Bloom et al., 1997) yielded more conclusive results and confirmed the earlier mentioned findings. More recent evaluations for the US investigate the differential effects of different labor market and training programs (Dyke et al. (2006), Hotz et al. (2006), Heinrich et al. (2010)) using more modern econometric methods. These studies address in particular the evaluation of programs including job search assistance/job readiness as well as training components and the comparison of longer to shorter programs. The results of Dyke et al. (2006) and Hotz et al. (2006) are not directly comparable to our study because they are obtained for a very specific group of welfare recipients, whereas our study concerns the somewhat broader group of recent inflows into unemployment. Heinrich et al. (2010) find positive medium-run to long-run effects for the broader group of participants in programs under the Workforce Investment Act. Their results suggest higher gains for participants of general programs for adults than for participants of the programs more specifically targeted to dislocated workers. For Canada, Park et al. (1996) distinguish training programs for the long-term unemployed, re-entry programs for women and programs aimed at eliminating specific skill shortages. Their results suggest that the latter two but not the first one involve positive earnings effects. Hui and Smith (2003a,b) distinguish different types of training mainly according to the source of financing (private, public, employer). Using different evaluation methods, the authors conclude that the quality of their data is not sufficient to provide credible estimates. For the UK, Dorsett (2006) analyzes the relative effectiveness of the five options offered under the New Deal for Young People. The five options are remaining on the gateway of intensive job search, subsidized full time employment, full-time education or training, work placement in a voluntary sector, and placement in the so-called Environmental Task Force. In most cases, entering one of the other options was not found to be more effective than intensive job search.

As one of the first studies analyzing the differential effects of heterogeneous training programs, Gerfin and Lechner (2002) analyze five types of public sponsored training programs in Switzerland with durations ranging between 5 and 13 weeks. The study finds that, one year after program start, the employment rate of participants was lower than that of comparable non-participants. Longer, more involved training courses show less negative effects than shorter ones.

For the Nordic countries, a number of studies investigate the effects of public sponsored training programs, often based on high quality data from administrative sources. Larsson (2003) evaluates the comparative effectiveness of two youth labor market programs in Sweden (youth practice vs. labor market training). She finds zero or even negative effects compared to no treatment. Sianesi (2008) analyzes the comparative effectiveness of different labor market programs in Sweden, among them vocational class-room training, work practice, and trainee replacement (on-the-job practice). She finds moderate positive effects for trainee placement but no positive effects for other forms of training. Based on between-program comparisons, Hardoy (2005) evaluates a number of youth labor market programs in Norway, among them employment programs, vocational programs and training programs. She finds that training programs and vocational programs are counterproductive, reducing significantly the employment prospects of their participants. A detailed comparative evaluation of different training programs in Norway is presented by Jespersen et al. (2008), who find that both private and public job training programs yield positive earnings effects, while classroom-training does not significantly improve the employment or earnings prospects of their participants.

Based on administrative data made recently available, a number of studies evaluate the differential effects of training programs in Germany for the 1990s. A disadvantage of these data compared to the data used in this study is that they only contain information about benefit receipt but not about registered unemployment and that less rich information on personal characteristics is available. Moreover, information about other programs is incomplete, which is a drawback because we show below that such information can be important. For instance, based on the data for the 1990s, it is not possible to distinguish participation in employment programs from ordinary employment. An advantage of the data for the 1990s is that they allow researchers to estimate long-term effects. Lechner et al. (2007, 2010) evaluate the effects of a variety of training programs in the 1990s. They distinguish between medium-term programs (mean duration 4 months), longer programs (mean duration 9 to 12 months) and long programs with specific contents such as retraining or training in a practice firm. Most of the programs show positive effects in the long run, even in East Germany. An important finding is that medium-term programs outperform longer programs as they exhibit a much shorter lock-in period with otherwise similar employment effects after the end of the program. Similar findings are obtained by Fitzenberger and Speckesser (2007), Fitzenberger and Völter (2007), and Fitzenberger et al. (2010) based on the same data but using different econometric methods. These studies for Germany do not find that practical training as implemented in the 1990s dominates other kinds of training, a finding in contrast to Martin and Grubb (2001) or OECD (2005). Based on recent data similar to those used in this study, Schneider et al. (2006) analyze the effects of the *Hartz-Reforms* on training programs. The study provides some results on the comparative effectiveness of a number of medium-term and long-term training programs, implying that shorter programs may be more effective than longer ones.

A drawback of the above mentioned studies for Germany is that they omit short-term training ('Trainingsmaßnahmen'), the by now most important type of public sponsored training. Short-term training courses typically last only 2 to 12 weeks and often combine elements of job search assistance with the provision of specific skills (see more detailed description below). In light of the policy debate (Martin and Grubb (2001), OECD (2005)), short-term training seems attractive since it may serve the purpose of activating the unemployed without locking them in lengthy training programs. Furthermore, a number of recent studies argue that increased job search assistance may be an inexpensive way to increase employment chances of unemployed individuals (Blundell et al. (2004), Weber and Hofer (2004), Fougère et al. (2005), Hujer et al. (2005), Crépon et al. (2005), and Van den Berg and Van der Klaauw (2006)). Our companion paper, Fitzenberger et al. (2008), estimates the effects of different types of short-term training for the 1980s, the early 1990s, and the early 2000s and finds modest positive employment effects. Using a duration approach, Hujer et al. (2006) find that participation in short-term training measures reduces the unemployment duration of West German job-seekers. The studies mentioned so far do not compare short-term training to other programs.

Most closely related to our study, Wunsch and Lechner (2008) and Lechner and Wunsch (2009) use the same data to estimate differential effects of various programs including short-term training. Wunsch and Lechner (2008) find for West Germany that none of the programs have any positive effect on employment or earnings. This is in sharp contrast to our findings. We find significantly positive employment and earnings effects in a considerable number of cases. Our study differs in the following important respects from Wunsch and Lechner (2008) (and Lechner and Wunsch (2009)). First, we stratify treatment and control groups according to elapsed unemployment duration, thereby accounting for the dynamic selection into programs as suggested by Sianesi (2004). Such a stratification is important for detecting the cases for which treatment effects are positive. Wunsch and Lechner (2008) define the control group based on nonparticipation during the observation window. Second, we do not have to use the method of randomly drawing artificial start dates for nonparticipants as Wunsch and Lechner (2008). Third, we provide very detailed balancing tests that show that our treatment and comparison groups are well balanced. Fourth, we carry out formal tests for heterogeneity of treatment effects. Fifth, we present detailed further sensitivity analyses demonstrating the value added of using rich administrative data. Also, the last three points mostly apply to our own previous work. Our analysis allows us to draw a more positive picture of the comparative effectiveness of training programs in West Germany than the study of Wunsch and Lechner (2008). In line with Lechner and Wunsch (2009), our earlier version of this paper (Biewen et al. (2007)) does not find significantly positive employment effects for East Germany. These results are omitted here because of space constraints.

3 Training as Part of Active Labor Market Policy

The main goal of German active labor market policy is to permanently reintegrate unemployed individuals (and individuals who are at risk of becoming unemployed) back into employment. The policy instruments cover a wide range of different programs such as employment subsidies, job creation in the public sector, measures directed at youth unemployment, measures to promote self-employment, and public training programs. For an overview of the different programs and their importance, see figure 1 and table 3 in the appendix. This paper focuses on training programs attended between 2000 and 2002, before the *Hartz-Reforms* were enacted. The reforms changed some of the rules on public training programs (Biewen and Fitzenberger (2004), Schneider et al. (2006)).

— Figure 1 about here —

Training programs have traditionally been the most important part of German active labor market policy (figure 1). There are three main categories: *short-term training* ('Trainingsmaßnahmen'), *further training* ('Berufliche Weiterbildung'), and *retraining* ('Umschulung').[4] Apart from the fact that all three types of training require full-time participation, they differ considerably in length and contents. Recently, *short-term training* has become the largest training program regarding the number of participants – for the following, see Kurtz (2003). Short-term training programs last only two to twelve weeks (the mean duration is slightly over four weeks, see table 1) and typically pursue one or several of the following three aims. A first potential aim is aptitude and qualification testing, i.e. the program is used to assess job seekers' labor market opportunities and their suitability for different jobs. This may also entail profiling activities on the side of the Federal Employment Office and preparation of more detailed work plans to reintegrate the job seeker into the labor market. A second aim is to test the job seeker's willingness to work and to improve job search skills. This may be achieved through activities such as job-application training, simulation of job interviews or general counseling on job search methods. The third and final aim of short-term training measures is the provision of specific skills that are necessary to improve the job seeker's labor market prospects. Typical examples for this type of measures are computer courses or courses providing commercial training. In 2001, 22 percent of short-term training measures belonged to the first type, 19 percent to the second type, and some 28 percent to the third type. About 31 percent were combinations of the different types. In most cases, these were combinations of job search assistance and the provision of specific skills, or aptitude testing and the provision of specific skills (Kurtz, 2003, tables A3 and A6).

[4]In addition, there are specific training schemes for youth unemployed and disabled persons, as well as German language courses for asylum seekers and ethnic Germans returning from former German settlements in Eastern Europe. Such training programs are not considered here.

In comparison to short-term training, the more substantial *further training* programs typically take much longer and are more involved. With durations ranging between several months and one year, further training measures can be classified as medium-term programs. Their aim is to maintain, update, adjust, and extend professional skills and qualifications. Further training programs cover a wide range of courses in a variety of fields and may also comprise practical elements such as on-the-job training, internships or working in practice firms. In our evaluation we will distinguish between practically-oriented further training programs (which are typically of shorter duration) and pure classroom training. Apart from short-term and further training, employment offices also offer *retraining*. Retraining programs last two to three years and typically lead to a new vocational education degree within the German apprenticeship system. In view of the long duration of retraining programs when compared to the observational window of our data set, we refrain from reporting estimates for retraining in the following.[5]

To become eligible for participation in one of the training programs, job seekers have to register personally at the local labor office. This involves a counseling interview with the caseworker. Besides being registered as unemployed or as a job seeker at risk of becoming unemployed, candidates for short-term training do not have to fulfil any additional eligibility criteria. In the case of medium- and long-term training, individuals are typically eligible only if they also fulfil a minimum work requirement of one year and if they are entitled to unemployment compensation. However, there are several exceptions to these requirements. The binding criterium is that the training program has to be considered necessary in order for the job seeker to find a new job. This is, for example, the case if the employment chances in the target occupation of a job seeker are good but require an additional adjustment of skills. Training programs are usually assigned by the caseworker. Depending on regional and local circumstances, caseworkers may exercise a great deal of discretion when allocating the different programs. Suitable programs are chosen from a pool of certified public or private institutions or firms.

If a person is admitted to one of the training measures, the employment office pays all direct training costs. In addition, the participants of short-term training may continue to receive unemployment benefits or means-tested unemployment assistance, if eligible. Participants of short-term training are still registered unemployed during the program. In contrast, participants of further training or retraining do not remain registered unemployed during the program. Participants of further training

[5]We reported estimates for retraining in an earlier version, see Biewen et al. (2007). In general, retraining measures are associated with a long and pronounced lock-in period.

(and retraining) usually also receive a subsistence allowance provided they fulfill a minimum work requirement of twelve months during the last three years. This subsistence allowance usually equals unemployment benefits or unemployment assistance. Overall, there are no sizeable financial incentives for unemployed individuals to participate in a training program, in contrast to the situation in Germany before 1998 (Fitzenberger et al. (2010)).

— Table 1 about here —

Table 1 shows that the average monthly training costs per participant are lower for short-term training courses (about 570 Euros in 2001) than for the longer-term measures (664 Euros). Given the that average length of short-term measures is only 1.1 months while that of longer-term measures is some 9.3 months, this results in training costs for short-term measures (627 Euros) of about one tenth that amount to only about one tenth of those for medium- and long-term programs (6175 Euros). Since 2002, in light of huge differences in costs, the Federal Employment Office has drastically increased the share of short-term training measures at the expense of longer-term measures (see figure 1). Of course, the higher training costs may be justified if the medium- to long-term measures lead to correspondingly higher gains in employment probabilities. This question is part of the motivation for our analysis.

4 Data

4.1 Integrated Employment Biographies Sample

Our study uses a new and exceptionally rich administrative data base, the so-called Integrated Biographies Sample (IEBS). This data base has only recently been made available by the Federal Employment Office of Germany.[6] The IEBS is a merged 2.2% random sample of individual data drawn from the universe of data records collected through four different administrative processes. The IEBS contains detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different programs of active labor market policy. In addition, the IEBS comprises a large variety of covariates including socio-economic characteristics (information on family,

[6]Detailed information on the IEBS is available on the website of the Research Data Center (FDZ) of the Federal Employment Office (http://fdz.iab.de/en).

health, and educational attainment), occupation and job characteristics, extensive firm and sectoral information, as well as details on individual job search histories and assessments of case workers. Our version of the IEBS has been supplemented with additional information which is not available in the standard version (e.g. information on health). For evaluation purposes, a rich set of covariates is essential as it can be used to reconstruct the circumstances that did or did not lead to the participation in a particular program.

We give a brief description of the IEBS in order to underscore its value for evaluation purposes. The IEBS is based on four different administrative data sources the so-called Employment History ('Beschäftigten-Historik'), the Benefit Recipient History ('Leistungsempfänger-Historik'), the Supply of Applicants ('Bewerberangebot'), and the Data Base of Program Participants ('Massnahme-Teilnehmer-Gesamtdatenbank'). The data is organized in spell format involving the start date and the end date at a daily frequency.

The *Employment History* involves register data comprising employment information for all employees subject to contributions to the public social security system. It covers the time period 1990 to 2004. The detailed information on the employment status allows us to account for the labor market history of individuals and to measure employment outcomes. For each employment spell, there is information on personal as well as job and firm characteristics such as wage, industry, or occupation.

The second data source, the *Benefit Recipient History*, includes spells of all unemployment benefit, unemployment assistance and subsistence allowance payments between January 1990 and June 2005. It also contains information on personal characteristics. The Benefit Recipient History is important as it provides information on the periods in which individuals were out of employment and therefore not covered by the Employment History. In particular, the Benefit Recipient History includes information about the exact start and end dates of periods of transfer receipt. The information on benefit payments allows us to construct individual benefit histories dating back several years. Moreover, we use additional information contained in the Benefit Recipients History involving periods of benefit withdrawal imposed as sanctions by the employment office as proxy for lack of motivation.

The third data source is the *Supply of Applicants*, which contains data on individuals searching for jobs. The Supply of Applicants data cover the period January 1997 to June 2005. In our study they are used in two ways. First, they provide additional information about the labor market status of a person, in particular whether the person in question searches for a job but is not (yet) registered as unemployed or whether he or she is sick while registered unemployed. Second, the job search episodes include additional information about personal characteristics, in particular about educational attainment, nationality, and marital status. They also provide information about whether the applicant wishes to change occupations and about health problems that might influence employment chances as assessed by the caseworker. Finally, the data on applicants include regional and local identifiers, which we use to link regional and local information, for example unemployment rates at the district level.

The fourth data source is the *Data Base of Program Participants*, which contains detailed information on participation in all public sector sponsored labor market programs covering the period January 2000 to July 2005. The data involves the type of the program as well as additional information on the program such as the planned end date, whether the participant entered the program with a delay, and whether the program was successfully completed. The Data Base of Program Participants not only contains information on the set of training measures evaluated in this paper, but also on all other possible programs such as employment subsidies. This is important, as it enables us to distinguish between regular employment and subsidized employment when evaluating employment outcomes.[7]

Being among the first to use the IEBS, we were involved in necessary comprehensive data checks because of the non-trivial task of merging four large scale administrative data sources of very different designs. We ran extensive consistency checks of the records coming from the different sources, making use of additional information on the data generating process provided to us by the Institute for Employment Research. This work is documented in Bender et al. (2004, 2005). Our conclusion is that on the one hand the employment and benefit data are highly reliable concerning employment status, wage and transfer payments, and the start and end dates of spells. The likely reason for this is that contribution rates and benefit entitlements are directly based on this information. On the other hand, information not needed for these administrative purposes can be less reliable. For example, in the employment data base the educational variable appears to be affected by non-negligible measurement error as it is not directly relevant for social security entitlements (see Fitzenberger et al. (2006) for imputation methods to correct the education variable).

[7]A disadvantage of the data covering labor market training in Germany in the 1990s used in studies such as Fitzenberger and Speckesser (2007), Fitzenberger and Völter (2007), Fitzenberger et al. (2010), and Lechner et al. (2007, 2010) is that it is not possible to distinguish whether participants found employment in the regular labor market or whether they obtained employment through a job creation program. Note that for the time period from the year 2000 onwards, Lechner et al. (2007, 2010) make use of the information on subsized employment provided by the IEBS.

Personal characteristics exhibit a higher degree of reliability in the program participation and job seeker data, because they are relevant for the purpose of assigning job offers or programs to the unemployed. In our evaluation, we exploit the available information as efficiently as possible by choosing the data source that is most reliable for a given purpose. Although the data in the IEBS generally seem very reliable, there is some need for data corrections. We corrected the end dates of program spells recorded in the data base of program participation using the procedure described in Waller (2008), which proved necessary for a small share of spells.

4.2 Evaluation Sample and Training Programs

We follow an evaluation strategy (see below) that is based on comparisons with (multiple) control groups. A common feature of control group approaches is that they partition the group of potential participants into a group of participants and a group of non-participants. As a consequence, one has to determine first as to who is a potential program participant.

For several reasons, we decide to focus on individuals who become unemployed after having been continuously employed for at least three months, instead of individuals who are *observed* unemployed at a given point of time. This is to avoid the case of individuals registering as unemployed from being out of labor force because they want to participate in a training program. In interviews, caseworkers told us that especially women returning from maternity leave, divorcees, or university graduates who have difficulty finding a job may contact the local employment office inquiring about the possibility of participating in public training programs. However, these individuals often only register as unemployed if the chances of actual participation are high enough. An evaluation sample based on observed unemployment status (instead of an inflow sample into unemployment) would therefore suffer from the problem of an incompletely observed control group for such participants: Because they are not registered as unemployed, non-participating counterparts are not recorded in the data. Analyzing an inflow sample into unemployment, we focus on individuals who have been attached to the labor market, which enables us to construct a suitable control group based on the relevant information in the data. Furthermore, the beginning of unemployment defines a natural time scale to align treated and nontreated individuals.

In the following, we focus on an inflow sample into unemployment consisting of individuals who became unemployed between the beginning of February 2000 and the end of January 2002, after having been continuously employed for at least three months. Entering unemployment is defined as ending regular, non-subsidized full-time or part-time employment and subsequently being in contact with the employment office (not necessarily immediately), either through benefit receipt, program participation or a job search spell.[8] In order to exclude individuals eligible for specific labor market programs for the youth and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 53 years at the beginning of their unemployment spell. Our evaluation focusses on the first training program that is attended in the course of an unemployment spell.

Based on the description of program types in section 3, we distinguish three different types of training, which closely follows the legal grouping of program types: *Short-term Training (STT)*, *Classroom further training (CFT)*, and *Practical further training (PFT)*. In some cases, we grouped programs whose planned duration and contents did not really fit into the category defined by the law into the category that was most appropriate from an economic point of view. According to the same criteria, we also grouped measures of 'discretionary support' (Freie Förderung) and measures financed through the European Social Fund (Europäischer Sozialfond, ESF) into one of the four program categories. For the reasons explained above, we do not consider retraining programs here, but use the information about whether a person participated in retraining when evaluating the other forms of training. That is, when defining the treatment group, participants in retraining as well as participants in other programs (e.g. employment subsidies) are not counted as in open unemployment. Also for the reasons explained above, we do not report estimates for East Germany (estimates for retraining and for East Germany can be found in the earlier version of this paper Biewen et al. (2007)).

We estimate program effects separately for men and women and for different durations of elapsed unemployment (as explained below). This results in a total number of six evaluation samples, the sample sizes of which are shown in table 4 in the appendix. Table 7 and figure 21 provide descriptive information on the duration of different program types. STT is the shortest program. Durations for CFT are fairly uniformly distributed between 1 and 11 months and they show a strong spike at 12 months. PFT is shorter than CFT and shows a strong spike at 7 months.

[8] This implies that the same individual may appear more than once in our evaluation sample. About ten percent of the individuals in our sample are represented by more than one unemployment spell according to the above definition. We take account of multiple inclusion of the same individual in the sample when calculating standard errors, see section 5.

5 Econometric Approach

5.1 Evaluating 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, exclusive treatments. Our analysis distinguishes three different treatments, STT, CFT, and PFT, as well as a nontreatment state. Let Y^k represent the potential outcome associated with training program k ($k = 1, 2, 3$) and Y^0 the potential outcome when not participating in any program. For each individual, only one of the 4 potential outcomes is observed and the remaining 3 outcomes are counterfactual. Our goal is to estimate the average treatment effect on the treated (ATT) of receiving treatment $k = 1, 2, 3$ both against nonparticipation $k = 0$ (treatment versus 'searching') and against treatment l, with $k \neq l$ and $k, l \neq 0$ (differential effect of two programs).

In our context, participation in training is possible at any point in time during the unemployment spell. A job-seeker who has not enrolled until a certain point in time may join at some later point in time provided he remains unemployed up to that time. In such a setting, persons with longer completed unemployment durations are more likely to end up receiving treatment than persons with shorter completed unemployment durations. Fredriksson and Johansson (2008) show that in this case a static evaluation analysis that assumes that the treatment is administered only once yields biased treatment effects because the definition of treatment and control groups implicitly conditions on future outcomes. Thus, we think that a conventional static evaluation approach is not warranted in our case. We follow the approach suggested in Sianesi (2004, 2008) to extend the static multiple treatment approach to a dynamic setting. Sianesi focuses on the effect of participating after a given unemployment experience (i.e. period of eligibility) versus not participating up to that point in time. Thus, treatment effects are defined conditional on a given starting date during the unemployment spell and treatment and control group are aligned in their unemployment duration up to then. The nontreatment group comprises all individuals who have been in open unemployment for the same period of time as the treated individuals but who do not enrol in a program at the starting date under consideration. Nonparticipation up to a given point in time entails the possibility of participation at some later point in time. Hence, individuals in the nontreatment group may enrol in training at some later point during the unemployment spell. This treatment parameter (treatment versus searching) mirrors the decision problem of the caseworker and the unemployed who recurrently during the unemployment spell decide whether to join any of the training programs now or to postpone participation to the future to see whether job search is successful without training. When discussing the policy implications of our estimates, one has to be careful because those nontreated individuals in one stratum who receive treatment later experience the impact of the treatment.[9]

We distinguish between treatment starting during months 0 to 3 of the unemployment spell (stratum 1), treatment starting during months 4 to 6 (stratum 2), and treatment starting during months 7 to 12 (stratum 3). We choose the widths of the three time windows based on sample size considerations. We experimented with smaller strata but found that the number of treated and comparison individuals became too small to implement our matching approach that relies on both exact matching and propensity score matching (see below). It is clear that one faces a bias-variance trade-off. On the one hand, reducing the interval length reduces the bias due to a static evaluation design but increases the variance as a consequence of smaller sample sizes. On the other hand, by defining wider intervals one ignores more of the variation in the timing of treatment that takes place within a given stratum. Due to even smaller sample sizes, we do not evaluate training programs starting more than 12 months after the beginning of unemployment.

We evaluate treatments conditional on the unemployment spell lasting at least until the start of treatment k and this being the first treatment during the unemployment spell considered. Therefore, the ATT parameter for treatment k against the alternative l ($k \in \{1,2,3\}$, $l \in \{0,1,2,3\}$, $k \neq l$) is given by

$$
\begin{aligned}
(1) \qquad \theta(k,l;s,\tau) = \quad & E(Y^k(s,\tau)|T_s = k, U^o = u-1, \underline{u}_s \leq u \leq \bar{u}_s) \\
& - E(Y^l(s,\tau)|T_s = k, U^o \geq u-1, \underline{u}_s \leq u \leq \bar{u}_s)\,,
\end{aligned}
$$

where $s \in \{1,2,3\}$ denotes the stratum of unemployment. $\tau = 0,1,2,\ldots$ counts the elapsed months since the beginning of treatment k. T_s is a variable indicating treatment status in stratum s. U^o denotes the random time spent in open (i.e. untreated) unemployment and $u \in \{\underline{u}_s,\ldots,\bar{u}_s\}$ the month in which treatment k

[9] As an example, when there is a cut in treatment during one stratum, our estimates assume that the incidence of later treatment conditional on the covariates considered does not change. When the likelihood of receiving treatment later is particularly high among those treated in an early stratum (had they not received treatment in the early stratum), then our estimates implicitly assume that the treatment cases grow during later strata (unless the lack of early treatment results in sufficiently quicker exits to employment). Thus, our estimates are not applicable for a situation where a cut in an earlier stratum is associated with overall treatment cases being unchanged in later strata.

starts, with $\underline{u}_s$ ($\bar{u}_s$) the start (end) of stratum s. Thus, for treatment k starting in period u we require that possible comparison individuals receiving treatment l have spent the same amount of time in open unemployment as of period $u-1$ and receive treatment l no earlier during the stratum considered than the treated individual, i.e. treatment of the comparison individuals starts during the time interval $[u, \bar{u}_s]$.[10] While we align treated and control persons exactly by elapsed time in open unemployment, we display separate treatment effects by stratum only. This means we average the period specific treatment effects with respect to the distribution of starting dates within a stratum.

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

$$\begin{aligned} & E(Y^l(s,\tau)|T_s = k, U^o \geq u-1, \underline{u}_s \leq u \leq \bar{u}_s, \boldsymbol{X}) \\ = & E(Y^l(s,\tau)|T_s = l, U^o \geq u-1, \underline{u}_s \leq u \leq \bar{u}_s, \boldsymbol{X}) \ , \end{aligned} \tag{2}$$

where $\boldsymbol{X}$ denotes a vector of covariates that are time-invariant within a stratum. We effectively assume that, conditional on $\boldsymbol{X}$ and on remaining in open unemployment at least until period $u-1$, individuals are comparable in their outcome for treatment l occurring between u and $\bar{u}_s$. As usual, implicit is the assumption that potential outcomes are independent across individuals, ruling out general equilibrium effects.

5.2 Combining Exact and Propensity Score Matching

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

To account for the dynamic nature of treatment assignment, we estimate the proba-

[10] For $l = 0$, the comparison group comprises all individuals whose unemployment spell lasts at least $u-1$ periods and who do not participate in any program during the stratum considered. This entails the possibility of exit to employment before the end of the stratum.

bility of receiving treatment k versus l in a given stratum using all the individuals in group k and l who are still unemployed at the beginning of the stratum considered. For treatment during months 0 to 3, we take the total sample of unemployed, who participate in k or l during months 0 to 3 (stratum 1), and estimate a Probit model for participation in k. For $l = 0$, the control group includes those unemployed who either never participate in any program or who start some treatment after month 3. Similarly, for treatment during strata 2 and 3, the basic sample consists of those individuals in group k and l who are still unemployed at the beginning of the stratum.

We then combine kernel matching involving the propensity score with cell matching. In particular, we implement a stratified kernel matching approach by imposing additional criteria along which treated and control observations are aligned exactly. First, we require that the matching partners for an individual receiving treatment k in a given month are still unemployed in the month before treatment k starts, i.e. we align treated and nontreated individuals exactly with respect to the elapsed unemployment duration in months. Second, we match exactly on a set of ten different employment histories covering the four years before the beginning of the unemployment spell (see below for further details). Then, we estimate the counterfactual employment and earnings outcomes by means of a Nadaraya-Watson kernel regression. We use a product kernel in the estimated propensity score and the calendar month of entry into unemployment

$$KK_j(p_d, c_d) = K\left(\frac{p_d - p_j}{h_p}\right) \cdot h_c^{|c_d - c_j|}, \tag{3}$$

where $K(z)$ is the Gaussian kernel function, p_d and c_d are the propensity score and the calendar month of entry into unemployment of a particular treated individual d, p_j and c_j are the estimated propensity score and the calendar month of entry into unemployment of an individual j belonging to the comparison group of individuals treated with l. The parameters h_p and h_c are the bandwidths. Taken together, we impose the following four matching requirements: (i) similarity in the pairwise propensity score, (ii) similarity in the calendar date of the beginning of unemployment, (iii) equality in the elapsed time in open unemployment, and (iv) equality in the previous employment history.

We use a bivariate crossvalidation procedure to obtain the bandwidths h_p and h_c by minimizing the squared prediction error of the average l-outcome for the individuals in the l-group who are most similar to the participants in program k.[11] We obtain standard errors and pointwise confidence bands for our estimated treatment parameters through bootstrapping based on 250 resamples. We resample individuals with their entire observation vector to obtain standard errors clustered at the individual level, thus also accounting for the multiple appearance of individuals as participants or nonparticipants, respectively. The propensity score is reestimated in each resample to take account of its estimation error.

5.3 Empirical Content of the DCIA

At this point, we need to discuss the plausibility of the DCIA (see equation (2)) in our application. In our analysis, we focus on the average treatment effect on the treated (ATT) conditional on some given unemployment experience. The DCIA states that conditional on a given unemployment experience and a vector of observable covariates, the sequence of potential outcomes associated with alternative treatment l in the same time window, $\{Y^l(s,\tau)\}^\tau$, $\tau = 0, 1, 2, \ldots$, is mean independent of the treatment status in this time window, T_s. The index l also comprises the option of nonparticipation. In a dynamic context, nonparticipation in the current stratum entails the possibility of participation in later strata. Thus, the DCIA implies that the treatment status in subsequent strata and the associated potential outcomes are conditionally mean independent of the treatment status in the current stratum.

Our matching approach will produce valid estimates if we consider all the determinants that jointly influence treatment status and potential outcomes in the current stratum as well as in future strata. Conditional on these determinants individuals are randomly allocated to one of the four treatment options that are available in a given stratum and there is no anticipation of future treatment and employment options. In the following, we explain the process of program assignment in Germany and describe the information that we have in our data. We think that conditional on the information in our data there remain no anticipatory effects with respect to future program participation and individual re-employment prospects.

During the time period we consider, training programs were assigned by the caseworker. This was often done in agreement with the job-seeker, considering his or her willingness to receive training and to work in a specific field. However, the final

[11] This method is also used in Fitzenberger et al. (2010) and is an extension of the crossvalidation procedure suggested in Bergemann et al. (2009).

decision was subject to the discretion of the caseworker. Blaschke and Plath (2000) report that indicators available to the caseworker like the composition of a group of participants in a particular course or his/her assessment of the motivation of the unemployed played an important role. Placement in employment had priority over program participation. Therefore, the unemployed were encouraged to continue job search at any time, even while participating in a training program.

The allocation of a particular job-seeker to a particular training program was driven strongly by the supply of courses (Schneider et al., 2006). Anecdotal evidence in Schneider et al. (2006) suggests that belated assignments and referrals on very short notice were commonly used in order to assure a high capacity utilization of the courses that were generally booked in advance for a calendar year by the employment agencies and to keep up job search incentives. This suggests that program assignment was not very targeted but driven by regional specific variation in the supply of courses throughout the year. From the perspective of the unemployed, the assignment to a particular program operated in a rather idiosyncratic way which he/she could not perfectly anticipate.

Moreover, with the Supply of Applicants database, we have at our disposal the complete set of information that is collected by the caseworker in the course of a counseling interview. This database builds the basis for the caseworker's decisions on the mix of services he/she will offer to a particular job-seeker. It comprises rich information on the demographic characteristics of a job-seeker, his/her qualifications (degrees held as well as subjective assessments by the caseworker) as well as his/her needs and preferences with regard to a suitable job opportunity (health constraints, target occupation, looking for a part-time job).

To be specific, we consider the following variables that reflect the caseworker's assessment of the motivation, plans and labor market prospects of a particular unemployed: the caseworker's assessment of the job-seeker's current health status, information on his/her previous health status (i.e. in the last two years before the start of the current unemployment spell), a dummy variable indicating whether the unemployed person showed signs of lacking motivation in the last three years (e.g. failed to show up at regular meetings), dummies indicating whether the job-seeker dropped out of a program in the last three years, whether he/she was suspended from benefits within the last three years, whether the person participated in a program providing psychosocial support in the last three years. In addition, we include variables indicating whether the job-seeker would like to work in a different occupation, whether he/she is looking for a part-time job, and the number of job proposals he/she received from the employment office.

In order to eliminate potential anticipatory effects related to the individual re-employment prospects of a job-seeker we exploit the detailed and complete longitudinal information in our administrative data sources covering employment and benefit receipt and impose several restrictions to ensure that our sample as a whole as well as the different treatment and comparison groups are sufficiently homogeneous. Our analysis sample is composed of individuals who, between 2000 and 2002, exit regular employment lasting three months or longer. We stratify treatment and comparison individuals based on ten groups of employment history sequences as well as elapsed unemployment duration and then match within each group by the propensity score and the start of the unemployment spell in calendar time. Following Card and Sullivan (1988), we construct ten different employment history groups that indicate different combinations of the annual employment status of an individual during the four years preceding the unemployment spell.[12] We exactly match on the nine sequences (1000), (1001), (1010), (1011), (1100), (1101), (1110), (1111), and (0000). E.g. the sequence (1010) represents employment during the first and the third year before entry into unemployment and nonemployment during the second and the fourth year. The nine sequences considered so far cover over 95 percent of the cases. The remaining sequences are subsumed in a residual category.

Comparative analyses and simulation studies (Heckman et al. (1997), Heckman et al. (1998), Heckman et al. (1999), Heckman and Smith (1999)) stress the importance of conditioning flexibly on lagged employment and wages, benefit receipt history, and local labor market conditions. We include in the propensity scores a rich set of variables depicting past employment and nonemployment histories as well as local labor market conditions. In particular, we include dummy variables referring to each of the ten employment sequences described above in the propensity score. Further, we use several different variables recording wages prior to unemployment as well as occupation and industry of the last job, indicators of whether the last job was part-time, whether it was a white-collar or blue-collar position, the reason why the job ended, and the quarter of the beginning of the unemployment spell. We also include indicators of whether the person was employed in month 6, 12, or 24 before the beginning of the unemployment spell, the number of days employed during the preceding seven years as well as the average annual earnings and the annual employment status in each of the seven years before the beginning of the unemployment spell.

[12] An individual is considered as employed in a given year if he/she is employed for at least 50 percent of the days.

Turning to nonemployment, we include the remaining (nonexhausted) period of entitlement for unemployment benefits at the beginning of the current unemployment stratum as well as variables counting the number of days during which someone received unemployment insurance payments (i.e. unemployment benefits, unemployment assistance, or subsistence allowance) in the last three years, the number of days without any information in the data set (e.g. periods of self-employment or out of the labor force), indicators for whether somebody was disabled or unable to work at some point in time during the preceding three years, and a variable recording previous contacts with the employment office during the last three years. Finally, to model local labor market conditions we use several different variables on the unemployment rate in the county of residence, indicators for the federal state and indicators for the economic situation of the local labor market.[13]

5.4 Specification of the Propensity Scores

Using all of the variables described in section 5.3 as possible regressors, we fit the propensity scores separately for each of the six evaluation subsamples (men/women, stratum 1/2/3), and each treatment comparison pair. In each case, we run an extensive specification search. The final specification was chosen based on which variables (according to institutional and economic knowledge) may drive the selection into programs, based on statistical significance of the variables included, and based on the balancing tests described below. We usually start with a fairly general specification and drop variables that are grossly insignificant. This leads to satisfactory specifications in most cases. In the few cases in which the balancing condition fails, we modify the specification and reinclude covariates or add interactions in order to eliminate the covariate imbalance. We take particular care in ensuring that pre-treatment employment and earnings outcomes are well balanced between treatment and control group members.[14] The final specifications typically include between 20 and 35 covariates.

[13] Descriptive statistics for selected covariates are given in the appendix. More detailed information on the sensitivity of results to including the variables reflecting caseworker assessments can be found in the Online-Appendix.

[14] Detailed test results are available in the Online-Appendix. Further estimation results are available from the authors upon request.

5.5 Testing for Covariate Balance

We employ two different balancing tests to check whether our treatment and control groups are sufficiently balanced. First, we test whether the means of important covariates differ significantly between treatment and matched control groups using bootstrap t-tests and χ^2-tests (for groups of variables that belong together). In each subsample, the comparison of means is based on the same set of 90 regressors that we consider particularly important (cf. the discussion in section 5.3). We calculate the matched mean of a given regressor by applying the matching procedure described above to this regressor in exactly the same way as we do to predict the counterfactual employment and earnings outcomes.[15]

Second, we investigate whether treated and matched nontreated individuals differ significantly in their employment and earnings outcomes over a period of seven years before the beginning of unemployment. 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 as well as the same type of four-year employment history before the beginning of the unemployment spell.

5.6 Estimating Effect Heterogeneity

The estimation of the ATT provides a semiparametric, aggregate impact measure of possibly heterogeneous treatment effects depending upon observable characteristics of the participating individuals. In general, treated and control groups may not be perfectly aligned with respect to the distribution of conditioning variables after matching. Then, the estimated ATT may be biased if treated and matched comparison individuals are not matched very well by observable characteristics affecting the treatment effect (Abadie and Imbens, 2002). In addition, a given value of the aggregate ATT may hide different positive and negative treatment effects for different subgroups of the treated. One strategy to estimate effect heterogeneity across observed characteristics consists in applying the dynamic matching estimator as described in section 5.1 separately to each subgroup of interest. However, the implementation of such an approach is limited by the curse of dimensionality. The sample sizes of the subgroups need to be sufficiently large to do so with reasonable precision. Therefore, we pursue a different strategy involving an ex post outcome

[15]Detailed test results can be found in the Online-Appendix to this paper.

regression in the matched sample. An additional advantage of our approach is that it allows us to test for mismatch between treated and comparison individuals.

Following Abadie and Imbens (2002)[16] and Mueser et al. (2007), we apply an regression based adjustment for remaining mismatch between treated and comparison individuals after matching. At the same time, we account for effect heterogeneity by running weighted linear regressions of the individual outcomes on those covariates, which could cause effect heterogeneity, and interaction terms with the treatment dummy.[17,18] We pool over all possible starting dates u within a stratum. Estimation is based on

$$\min \sum_{i} \sum_{u=\underline{u}_s}^{\bar{u}_s} g_i(u)[\bar{Y}_i(u) - \alpha - \boldsymbol{X}_i\boldsymbol{\beta} - \gamma D_i - D_i(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{\delta}]^2 \,, \tag{4}$$

where the summation is over all individuals i who either receive treatment k or l in the stratum considered $[\underline{u}_s, \bar{u}_s]$. $g_i(u)$ is the set of weights for possible treatment starts in month u. $\bar{Y}_i(u)$ is the average of the individual period specific outcomes over 24 months from month u onwards, including the potential lock-in period of the treatment. D_i is a dummy variable for treatment status in the respective time window, i.e. $D_i = 1$ ($D_i = 0$) if individual i receives treatment k (l). $\boldsymbol{X}_i$ denotes a vector of observable characteristics and $\bar{\boldsymbol{X}}$ the average of the characteristics in the sample of individuals receiving treatment k. Let $U^s \in \{\underline{u}_s, \dots, \bar{u}_s\}$ denote the month of the beginning of treatment for the treated receiving treatment k ($D_i = 1$). Then, the weights for the treated individuals ($D_i = 1$) pick the observed start month, i.e. $g_i(u) = 1$ for persons starting treatment k in month u ($U^s = u$) and $g_i(u) = 0$ if

[16]Note that the Econometrica publication drops the mismatch corrected estimator, see footnote 7 in Abadie and Imbens (2006, p. 241). Abadie and Imbens (2002) show that such an ex post regression adjustment improves the asymptotic properties of nearest neighbor matching based on a fixed number of matches because it controls for the leading bias term of the ATT estimate, see also Mueser et al. (2007, p. 765) for a discussion. However, this point is not relevant in our case since we apply kernel matching. Nevertheless, the ex post regression adjustment is a useful specification device to investigate whether our kernel matching approach suffers from mismatch bias, although the discussion in Abadie and Imbens (2002) suggests that bias adjustment is likely to be less important in large samples like ours.

[17]Recent evaluation studies rarely estimate such ex post outcome regressions after matching to study effect heterogeneity. In the literature on statistical treatment choice, Frölich (2008) uses a similar approach to predict potential outcomes as a function of individual characteristics.

[18]Mueser et al. (2007) adjust each set of comparison cases to the particular treated case using a linear regression in the weighted dataset of comparison cases (this approach is also implemented by Heinrich et al. (2010)). In contrast to our approach, their analysis does not account for interaction effects because their goal is not to estimate how treatment effects vary by covariates. However, the estimated adjusted average treatment effect, i.e. our estimated γ in equation (4), should be very similar to what we would have obtained had we applied the approach taken by Mueser et al. (2007).

treatment starts not in month u ($U^s \neq u$). For a nontreated individual j ($D_j = 0$) serving as a comparison for treatment k starting in month u, the weight $g_j(u)$ equals

$$\sum_{d \in \{T_s = k \wedge U^s = u\}} K\left(\frac{p_d - p_j}{h_p}\right) \cdot h_c^{|c_d - c_j|},$$

where d ($d \neq i$) indexes those individuals starting treatment k in month u ($T_s = k \wedge U^s = u$). The function inside the sum corresponds to the kernel weight as given in equation (3). Thus, the control observation j is used repeatedly for each possible treatment starting date u in the respective stratum for which j serves as a comparison unit, i.e. observation j is used up to ($\bar{u}_s - \underline{u}_s + 1$) times in the pooled regression (4). This way, we obtain a proper temporal alignment of treated and control observations.

We use regression (4) to analyze heterogeneity of the treatment effects across observed characteristics. If $\boldsymbol{\delta} = 0$ in the regression (4), then there is no linear effect heterogeneity by the level of covariates.

Furthermore, regression (4) can be used to adjust for possible remaining mismatch between treated individuals and matched controls. The term $\boldsymbol{X}_i\boldsymbol{\beta}$ in regression (4) controls for the effect of the characteristics on the average outcome variable. The coefficient on the treatment dummy, γ, gives the mismatch corrected estimate of the aggregate ATT. This mismatch corrected estimate can be compared to the estimate obtained without the additional regression adjustment. If the two estimates are very similar then this is evidence that the matching procedure works well in aligning the distributions of the conditioning variables in the comparison group to the one in the treatment group. The standard errors of the estimated coefficients in (4) are obtained through the full bootstrap procedure for the matching estimator by rerunning the regression for all resamples.

6 Empirical Results

6.1 Employment Effects

6.1.1 Training vs. Searching

The evaluation results for training vs. not participating in any active labor market program (i.e. 'searching') in a given time window are shown in figures 2 to 4. Each graph displays the average treatment effect on the treated (ATT), i.e. the difference between the actual and the counterfactual employment outcome averaged over those individuals who participate in the program under consideration. More precisely, we compare the actual employment outcome of the treated to the employment outcome these individuals would have had, *had they not taken part in any program in the respective time window of their unemployment spell.* As already mentioned, we distinguish between programs starting in three different time windows (strata) of elapsed unemployment: 0 to 3 months (stratum 1), 4 to 6 months (stratum 2), and 7 to 12 months (stratum 3). Due to the smaller number of treated individuals, we only consider one time window ranging from month 0 to 12 for participants in practical further training (PFT).

We evaluate treatment effects at different points in time. On the time axis in our graphs, positive values denote months since the program start, while negative values represent pre-unemployment months and years. We omit the period between the start of unemployment and the start of the program where both control and treatment group are unemployed. The dashed lines around the estimated ATT are bootstrapped 95 % pointwise confidence bands. Treatment effects for a particular time period are statistically significant if zero is not contained in the confidence band.

First of all note that, in all the cases depicted in figures 2 to 4, pre-treatment employment outcomes of treatment and control group do not differ in any significant way in the seven years before the start of the current unemployment spell. We interpret this as evidence for the high quality of our matching approach.

— Figure 2 about here —

Figure 2 shows estimated treatment effects for short-term training programs (STT). The results for men (women) are in the left (right) column. We find short and not very pronounced lock-in effects of STT programs in the order of minus five percentage points (ppoints), i.e. during the program participants have a five ppoints lower monthly employment rate than they would have if they did not participate in the program. These lock-in effects do not last for more than two to three months, which is not surprising given the short duration of STT programs. After the short lock-in period, the treatment effects, i.e. the difference between actual and counterfactual employment outcomes of participants, turn positive in general, but they are not always statistically significant. The results seem to depend strongly on elapsed unemployment duration at program start. While there is no evidence for statistically significant treatment effects for individuals participating in the first three months of their unemployment spell (stratum 1), treatment effects for men starting an STT program in months 7 to 12 (stratum 3) of their unemployment spell, and women starting one in month 4 or later are significantly positive. Accordingly, the monthly employment rate of West German men (women) participating in STT is increased by about 5 ppoints (about 10 ppoints) after the end of the lock-in period.

— Figure 3 about here —

Results for the more substantive classroom further training programs (CFT) are given in figure 3. The most conspicuous difference between these results and those for STT programs is the long and pronounced lock-in effect. During the first months of their participation in the program, participants have an employment rate that is up to 25 ppoints lower than it would have been if they had not taken part in the program. The lock-in period lasts up to 12 months for individuals who take up their treatment during the first 6 months of their unemployment spell. Interestingly, lock-in effects are less deep and shorter for individuals that have been unemployed for more than 6 months (stratum 3).

Since the populations of treated and control individuals change over time, there are several possible reasons for this finding. First, it might be that individuals with a longer elapsed unemployment duration are assigned to shorter CFT programs. Second, it is possible that such individuals drop out of the program more often or earlier. A third reason may be that a large number of those just having become unemployed easily find new jobs if they do not take part in a training program. If these individuals are assigned to a CFT program anyway, they will be 'locked-in', while many of their counterparts in the control group quickly exit to employment. This could mean that some of the short-term unemployed receive training even though they do not need it to overcome unemployment. In addition, there may be a tendency towards finding less pronounced lock-in effects for the late program starts if many of the long-term unemployed in the control group abandon their job search and move out of labor force. Hence, an additional channel through which training programs work may consist in keeping the long-term unemployed in the labor force.

While there is little evidence for statistically significant employment effects for West German men starting a CFT program in months 0 to 6 of their unemployment spell (strata 1 and 2) or West German women starting it in the first 3 months of unemployment (stratum 1), treatment effects for longer-term unemployed men (stratum 3) and medium to longer-term unemployed women (strata 2 and 3) are large and statistically significant. After the initial lock-in phase, they amount to some 10 ppoints for men and for women.

— Figure 4 about here —

In contrast to pure classroom further training, practical further training (PFT) also includes practical elements such as internships or working in a practice firm. Evaluation results for these measures are given in figure 4. There are considerable positive employment effects of about 10 ppoints for women after a lock-in period of up to 8 months. There are no such effects for men. A reason for this finding could be that particularly in practice-related jobs, men and women select themselves into different occupations. If women more often participate in training for occupations in the service sector, where employment chances are generally better than in manufacturing or construction jobs, this will lead to more positive effects of practical training programs for women.[19]

6.1.2 Pairwise Comparisons of Training Programs

Given that, in many cases, training programs show considerable employment effects when compared to attending no program, the question arises which of the different training programs is the most effective for a given subpopulation. The results presented in the previous section already suggest that STT may have similarly positive effects as CFT or PFT when each of the programs is compared to attending no program. However, this does not necessarily mean that participants in STT could not have improved their employment chances by attending CFT or PFT programs instead, or that participants in CFT or PFT would not have lost from taking part in STT instead. This is the question we address next. For each comparison pair in figures 5 to 8, the treatment named first corresponds to the treatment considered (whose participants define the subpopulation considered) and the second one to the comparison treatment.

— Figure 5 about here —

Figure 5 shows that, in the first two and a half years since program start, participants in STT generally would *not* have improved their employment chances by attending

[19]Lechner et al. (2007) and Fitzenberger and Völter (2007) consider the role of gender specific target professions for explaining gender differences in treatment effects of public sector sponsored training in East Germany in the 1990s.

CFT. On the contrary, the much shorter and less pronounced lock-in effect of STT programs make this form of training seem more effective than the longer-term classroom training. Furthermore, figure 6 shows that participants in CFT would *not* have lost in general from attending STT instead.[20] This is remarkable since it means that these individuals could have been assigned to the much less expensive STT programs without reducing their employment chances. Again, taking the shorter lock-in effect at program start into account, STT would have even been preferable for these individuals. Taken together, CFT does not lead in general to higher employment gains than STT within the first two and a half years after program start.

— Figure 6 about here —

How does PFT compare to STT? Before discussing these results it should be noted that matching STT and CFT against PFT turns out to be difficult. First, we pool the three strata for treatment versus PFT (as in the PFT against Searching or Treatment case) because of the small number of participants in PFT. But severe balancing problems due to a small number of controls (PFT participants) for the later participants in STT or CFT, respectively, remain. Therefore, for these particular cases, we use only participants in STT (or CFT, respectively) of the first half year, while for the control group PFT participant of the first year are used. This improved the balancing results. The ATT results turned out to be robust and the graphical evidence stayed almost the same. Nevertheless, the results involving comparisons with PFT should be interpreted with some caution, because – due to small sample sizes – we can not always balance our covariates perfectly in these cases, even though we undertake an extensive specification search.[21] Figure 7 shows that individuals who took part in STT would not have gained from attending PFT instead. More importantly, there is no evidence that participants in PFT would have lost significantly within the first two years after program start if they had participated in STT instead.

— Figure 7 about here —

How do the PFT programs compare to the more theoretical CFT programs? Providing evidence on this question, figure 8 implies that participants in theoretical CFT

[20]The situation is less clear for men who have been unemployed for at least 6 months.

[21]The results of the balancing tests for our final specification are given in the additional Online-Appendix.

would neither have lost nor gained significantly, had they been assigned to PFT instead. Moreover, PFT is better for participants of PFT than CFT would have been. This holds especially for female participants in PFT whose employment probability remains significantly higher than it would have been with CFT even after the end of the program. Note that PFT programs also exhibit significantly smaller lock-in effects, a finding which is not surprising given their shorter average length.

— Figure 8 about here —

6.1.3 Cumulated Effects

Next, we discuss the cumulated effects of the different ATTs that are displayed in table 2. They are calculated by summing the monthly treatment effects over the first 24 months counting from the program start. They provide a simple measure of whether, in total after a period of two years, participation in the program considered leads to employment gains or losses compared to participation in the comparison treatment for the individuals receiving the program considered. The higher effectiveness of shorter programs compared to longer programs is confirmed in table 2. For example, for men participating in short-term training after having been unemployed for at least 7 months (stratum 3), the net gain of the program versus searching is 0.9 employment months during the first 24 months after the program start. The effects are even stronger for women who gain on average 1.8 (stratum 2) and 2.1 (stratum 3) months of employment. The results suggest that STT is the only form of training that has statistically significant positive cumulated employment effects during the first two years, and this only holds for those individuals who are not treated too soon after entering unemployment.

— Table 2 about here —

Table 2 summarizes the comparative effectiveness of the different programs. Rows 2 and 3 show that STT was better in terms of its cumulated employment effects for those participating in it than CFT or PFT would have been (STT vs. CFT and STT vs. PFT). On the other hand, row 5 (CFT vs. STT) suggest that for participants in CFT, STT would be preferable over a period of 24 months, particularly for women. Row 8 (PFT vs. STT) indicates that participants in practical further training programs would not lose if they were assigned to STT instead. Row 9 (PFT vs. CFT) shows that participants in practical training would fare worse if they were assigned to CFT instead. On the other hand, participants in classroom further training would *not* fare worse if they were assigned to practically oriented programs instead (CFT vs. PFT, row 6). This indicates that, at least as long as one focuses on employment effects alone, practically oriented programs have some advantage over pure classroom training.

6.1.4 Effect Heterogeneity

The results presented so far reveal considerable heterogeneity in the effects for different subpopulations. A consistent finding seems to be that programs are only effective for those individuals who start a program after having been unemployed for some time. Another finding is that training effects are generally larger for women than for men (figures 2 and 3). Women also seem to benefit from practical training, while men do not (figure 4). Now, we further investigate whether treatment effects averaged over broad subpopulations hide statistically and economically significantly different effects for particular subpopulations. For this purpose, we run ex post outcome regressions using treated and matched control observations as described in section 5.6 above. The dependent variable, the average observed outcome for the first 24 months after the start of the program, is regressed on a number of personal characteristics. We investigate how the treatment effects vary with the personal characteristics included in the regression. We also investigate whether there is mismatch between treated and control individuals with respect to particular individual characteristics. We use the ex post outcome regression to compute 'mismatch corrected' treatment effects (by omitting the mismatch when calculating average treatment effects based on the regression of individual treatment effects on personal characteristics). If our matching approach works well, the mismatch corrected average treatment effects should coincide with uncorrected average treatment effects.

The results of the Wald-tests for effect heterogeneity and the comparison between the mismatch-corrected and the uncorrected ATT estimates are given in section 7.2 in the appendix. The detailed coefficient estimates are reported in the additional Online-Appendix. The mismatch-corrected estimates of average treatment effects reported in table 9 in the appendix coincide sufficiently well with the uncorrected estimates. In no case does the uncorrected ATT lie outside of the 95%-confidence interval of the missmatch-corrected ATT. These findings support the validity of our matching procedure. Furthermore, there is little evidence for effect heterogeneity as the vast majority of the interaction effects proves insignificant. According to the results reported in table 8, the hypothesis that all interactions between the treatment dummy D_i and the characteristics have zero coefficients ($H_0 : \boldsymbol{\delta} = 0$, see equation 4) can only be rejected in the case of females participating in PFT at the 5%-level. The significant effect heterogeneity in this case is caused by the differences across occupations as well as by the variables child and health. Unemployed whose previous job involved a technical occupation show worse effects. Having a child reduces the effectiveness and having health problems (health2), which are considered to be without impact on placement, increases the effectiveness. For the other treatment cases, a very small number of single coefficients for interaction effects prove significant but there is no generalizable pattern of effect heterogeneity across the different cases. Some of the single significant coefficient estimates take implausibly high values and are very difficult to rationalize, which may be due to small samples or multicollinearity problems. The overall insignificance for all cases, except for females participating in PFT, suggests that we should not interpret these effects.

The general finding for the comparisons of training vs. searching seems to be that – in addition to the heterogeneity found between the subgroups defined by gender and unemployment duration – there is very little heterogeneity along observed characteristics.

6.2 Earnings Effects

Although the main purpose of training programs for the unemployed is to reintegrate them into the labor market, there is also a lot of interest in the impact of different programs on earnings. If two programs are equally effective in increasing employment probabilities, the program with the greater earnings effects should be preferred (holding other things constant). Similarly, a program that does not increase employment probabilities may nevertheless be effective if it has a positive effect on earnings. Next, we carry out the same evaluations as before but with earnings as the outcome of interest. We focus on half–yearly earnings.[22] The results for earnings are shown in figures 9 to 15.

[22] As information on earnings comes from the social security records, it is collected only once in a given calendar year if a person is employed with the same employer throughout and no changes in relevant information, such as a change of the health insurance, occur within the year. As a compromise between the administrative reporting rule and short-run earnings dynamics arising from take-up of a new job within a given year we decided to focus on half–yearly earnings.

6.2.1 Training vs. Searching

Figure 9 shows the results for the case 'STT vs. searching'. There are statistically significant earnings gains of the order of 500 - 1,000 Euros for female participants in STT in stratum 2 and 3, and for male participants in stratum 3. This corresponds to the positive employments effects for these groups presented above. However, there are also statistically significant earnings gains for female participants in stratum 1 although there were no clear employment gains for this group. This is a case where evaluating earnings effects provides valuable extra information.

— Figure 9 here —

Figure 10 shows a similar pattern for the case 'CFT vs. searching'. After an initial lock-in effect, there are significant earnings gains for male participants in CFT in stratum 3 (and marginally in stratum 2), and in all strata for women. The earnings gains after lock-in tend to be larger than those for STT, i.e. 1,000 Euros per half-year or larger in many cases. This means that although CFT and STT provide comparable employment gains (after lock-in), CFT participants benefit on average more from their program than participants in STT.

— Figure 10 here —

The evidence for 'PFT vs. searching' shown in figure 11 suggests that there are large and significant earnings gains for women, amounting to up to 1,500 Euros per half-year two years after program start, but no significant effects for men. The latter is consistent with the finding that there are no employment gains for this group of participants either (see figure 4).

— Figure 11 here —

The figures presented so far also show that the earnings differences between treatment and control group members are generally not significantly different from zero in the seven years before the start of the unemployment spell considered.[23]

[23] The only exception is the case 'STT vs. searching for women in stratum 1' where treated individuals seem to have significantly *lower* earnings before treatment. A possible consequence is that the earnings gain induced by the program may be even larger than the estimates suggest.

6.2.2 Pairwise Comparisons of Training Programs

As shown in figure 12, participants in STT did not have any significant earnings disadvantage compared to taking part in CFT instead. In the short term, participating in STT is even associated with earnings gains compared to CFT because of the longer lock-in period for CFT.

— Figure 12 here —

The comparisons 'STT vs. PFT' and 'PFT vs. STT' presented in figure 13 suggest that male participants in STT have significantly *higher* earnings than they would have had, had they taken part in PFT (however, as shown in figures 2 and 9 participating in STT was not better than not taking part in any program). For female participants in STT, taking part in PFT instead would not have been more advantageous. The graphs in the lower part of figure 13 show that male participants in PFT would also not have incurred earnings losses if they participated in STT instead. By contrast, female participants in PFT would have experienced lower earnings if they had been assigned to STT instead. This is remarkable as there is no clear gain in employment probability for female participants in PFT compared to STT.

— Figure 13 here —

As figure 14 shows, there is also no clear earnings gain for female participants in CFT compared to STT (right column). For men, the situation is slightly different (left column): although the effects are only marginally significant, male participants in CFT would have incurred earnings losses (after the lock-in period) if they had been assigned to STT instead.

— Figure 14 here —

Finally, the comparison between CFT and PFT in figure 15 suggests that male participants in CFT would have experienced significant earnings losses if they had been assigned to PFT. This is not true for female participants. On the other hand, male and female participants in PFT would not have benefited from switching to CFT.

— Figure 15 here —

6.3 How Sensitive Are Impact Estimates to Features of the Data and to Specification Choices?

Given the richness of our data and the wide range of possible specification choices, it is a highly relevant question to what extent features of our data and specification choices influence the outcome of our evaluation. The following analysis is therefore meant to contribute to the understanding of what kind of data and specification issues are important when choosing the econometric evaluation strategy. In the following, we start from our most general specification (whose results we reported in the previous sections) as a benchmark. We then sequentially drop specification features and/or information in the data in order to see to what extent results hinge on these aspects. As an outcome variable we consider the probability to be employed in a given time period. The complete sensitivity analysis is available in the Online-Appendix; here we only report results for one benchmark case (CFT vs. searching for men in stratum 2) along with one selected other case in order to illustrate the range of variations that may result as a consequence of data or specification changes.

6.3.1 Sensitivity Analysis 1: Employment History

In our first sensitivity analysis, following Card and Sullivan (1988), Heckman and Smith (1999), and Dolton et al. (2008), we investigate how much difference it makes to condition flexibly on past labor market outcomes. Starting from our benchmark specification, we first drop the requirement of exact matching on four-year history sequences. In a second step, we remove in addition the history sequence dummies from the propensity score. In the third step, the remaining seven-year history variables are dropped. The fourth step further removes all information on the benefit history, and in the fifth step, we finally drop the requirement to exactly match on the unemployment duration prior to treatment.

— Figures 16 and 17 here —

The results of this exercise are shown in figures 16 and 17. The left column always shows the case 'CFT vs. searching for men in stratum 2', while the right column presents one further case illustrating what else may happen.[24] The results show that omitting the requirement of exact matching on employment histories in general

[24] The full range of other cases is given in the Online-Appendix.

changes impact estimates only little. In the case of 'CFT vs. searching for men in stratum 2' treatment effects are pulled downwards by a moderately small amount suggesting that not controlling for employment sequences leaves too many relatively successful individuals in the control group. Looking at all other cases, this result is not generalizable however. More typical is what is shown for the case 'SST vs. searching for men in stratum 2' (first row, right column), where not conditioning on employment sequences makes no discernable difference whatsoever. The same is true if in addition the employment history dummies are dropped from the propensity score and the window of pre-unemployment information is reduced from seven to three years (steps 2 and 3). If, in addition, benefit-related information is dropped, estimates also change very little (step 4). It turns out however, that dropping the requirement of exact matching on unemployment duration (step 5) considerably reduces estimated treatment effects in most cases, especially for individuals with long unemployment durations. The case 'SST vs. searching for men in stratum 3' involves the strongest change: positive treatment effects are completely eliminated. This suggests that treated individuals tend to have longer unemployment durations, biasing downward impact estimates if unemployment durations are not accounted for.

The finding that controlling for past employment and unemployment information is important is in line with the results reported by Card and Sullivan (1988), Heckman and Smith (1999), and Dolton et al (2008). We find however, that doing so *flexibly* and using seven years instead of three years does not add much to increase the balance in the employment history between treated and controls. It should be noted however, that in contrast to some of the above cited studies, our evaluation sample is already constructed in a way that strongly conditions on past labor market transitions. This is because we consider individuals who became unemployed after a continuous spell of employment.

6.3.2 Sensitivity Analysis 2: Rich Personal Information and Specification Search

In our second sensitivity analysis, we vary the amount of personal information used for matching and the way in which it is used in specifying the propensity score. First, we investigate what happens if we omit the personality variables described above (information on whether the person took part in a program providing psychosocial support in the past, whether there were signs of a lack of motivation during previous unemployment spells, information on past penalties or past dropout out of program, the number of job proposals received, and the desire to change the occupation). In a second step, we omit in addition rich personal information in case it was included in the propensity score in the benchmark specification (information on household type, number and age of dependent children, marital status, information on health, disability, previous part-time employment and reasons why the last job was ended). In a related but separate analysis, we compare the benchmark specification with a specification in which all variables are only used in a fixed, basic way without investing in the specification search for each group as described above.

— Figure 18 here —

The results in the first row of figure 18 show that omitting the personality variables changes impact estimates surprisingly little as long as all other variables of the benchmark specification are controlled for. By contrast, omitting other personal information (row 2) may have a considerable effect, not so much in the benchmark case of CFT, but in practically all other cases (especially for STT after short unemployment). The direction of the changes is not systematic, but omitting rich personal information tends to overstate treatment effects (suggesting that treated individuals tend to have favorable personal characteristics). The fact that the inclusion of personality variables does not change much suggests that the kind of information they represent are already sufficiently included in the employment history variables and in the other personal characteristics. This is in contrast to the findings in Dolton et al. (2008), where related attitudinal variables added much to control for the selection into treatment.[25] Finally, we find that varying the amount of specification search leads to relatively minor changes in most cases. The changes also do not display any clear direction (row 3).

6.3.3 Sensitivity Analysis 3: Comparison to MTG and Information on Other Programs

Mueser et al. (2007) (henceforth MTG) present a recent state-of-the-art evaluation of job training programs based on administrative data from Missouri. Their setup is representative for what is possible using administrative data taken from different sources in the US and elsewhere. Our data is also based on administrative

[25] Note, however, that the attitudinal variables in Dolton et al. (2008) were directly elicited in a survey, while our variables are derived from information on program participation, receipt of unemployment benefits and caseworker assessments.

sources but seems to be richer in terms of individual characteristics and details of employment histories. It is also more comprehensive in another important respect: it contains full information on all possible programs a person might participate in. We investigate how our impact estimates would look if we only had available the kind of data used in MTG.

Our first sensitivity check compares our benchmark specification to a basic specification which is as close as possible to the one used in MTG (p. 768, basic demographic and educational information – i.e. variables on age, education, occupation, nationality and region – as well as four quarters of earnings history and four quarters of employment dummies based on whether earnings in the given quarter were positive). We stick to our requirement to exactly match on the unemployment duration prior to treatment even though MTG do not measure unemployment durations at a monthly frequency which implies that their treatment and control group are more heterogeneous compared to our sample. Dropping this requirement has a strong impact on our results as step 5 of the sensitivity analysis on the value of employment history information has shown, but it may be less important in the institutional environment of MTG in which program participation does not require unemployment. Our second sensitivity check addresses the point that we have complete knowledge on what kind of other programs someone is participating in, which is in contrast to the data used in MTG and many other data sets used in evaluation studies. For this check we compare our benchmark specification to one that ignores all information on programs other than the one in question (this mimics the situation in MTG, who acknowledge that an unknown number of controls may have participated in other programs).

— Figure 19 here —

Results for this exercise are shown in figure 19. The first row shows the comparison with a specification that uses the kind of information used in MTG. It turns out that in most cases treatment effects are overstated if matching is based on this kind of information (the graph on the left 'CFT vs. searching for men in stratum 2' is relatively untypical in that respect, most cases look more like the graph on the right 'CFT vs. searching for men in stratum 3'). This specification differs from our benchmark specification mainly in dropping detailed history information and rich personal information. Our detailed sensitivity analysis on history information has shown that dropping this kind of information leads only to small changes in the results, as long as the requirement of the same elapsed unemployment duration before program start is not dropped (we do not drop this requirement here). But dropping rich personal information, like for example variables on family status and health status, leads to an overstatement of the treatment effects. In fact, part of the figures of the second step of sensitivity analysis 2 (graphs in the middle row of figure 18 and the respective graphs in the additional Online-Appendix) look already close to the figures of the present exercise. The remaining difference must be due to dropping even more information from the propensity score (mainly industry variables, a whitecollar dummy and exact measurement of days in employment in the last three years) and to dropping detailed history information and rich personal information at the same time. The results for the comparison with a situation where no information on other programs is available are shown in the second row of figure 19. In many cases, effects are pulled downwards as in the case 'CFT vs. searching for men in stratum 2' (left graph). In other cases ignoring other programs makes no difference (right graph).[26] The reason why ignoring other program information tends to pull down impact estimates seems to be that, if information on other programs is available, participants in these other programs are only counted in evaluation windows in which they have not yet started the other program, i.e. when they are unemployed, whereas both their unemployment and their employment spells are counted if it is not known that they take part in some other program. When judging the value of this information one should consider that the unavailability of information on participation in other programs is likely to be less severe for countries with a relatively low program density, as the US, than for those continental European countries with a particular high program density. Even if not completely generalizable to every institutional environment, our results clearly show that limitations in data availability can have sizeable effects on evaluation results.

6.3.4 Sensitivity Analysis 4: Future Participation in Other Programs

Estimating the effect of treatment versus searching, the question arises to what extent our results are determined by the fact that we include in our control group individuals who may participate in some program in the future (i.e. after the end of the stratum considered). In order to address this question, we consider two variations. First, we completely exclude all individuals *from the beginning* who will eventually (within the next 35 months and within the same unemployment spell) participate in training or another intensive active labor market program. This corresponds to the setup used in many evaluation studies in which the control group

[26] Remember that the full set of graphs is available in the Online-Appendix.

is defined as the set of individuals who are *never* treated. Fredriksson and Johansson (2008) heavily criticize this approach because excluding future participants implicitly conditions on future outcomes and future treatment (i.e. future participants tend to be negatively selected because they would not have been treated if they had found employment before). As a second variation, we exclude future participants from the control group *from the month onwards when they enter a program* (similar to the approach suggested in Fredriksson and Johansson (2008) and DeLuna and Johansson (2009)).

— Figure 20 here —

Results for both variations are shown in figure 20. The first row shows the case in which future participants are completely excluded. This drastically reduces impact estimates in all cases, a finding which is in accordance with the view that excluding future participants results in a more positively selected control group. This might be a reason why previous evaluation studies did not find positive treatment effects in many cases (see Crépon et al. (2009) for related results).

The second row presents selected results for the case in which future participants are only excluded from the month onwards they enter a program, but not before. A concern related to the dynamic evaluation setup as suggested by Sianesi (2004) and used in this study is that the control group in a given stratum includes individuals who might be treated – and therefore affected by a lock-in effect – in future strata, making the counterfactual a compound of treatment and non-treatment outcome. We think that our estimated treatment effects are of interest because the control group for our results above is given by those individuals who are continuing their search, i.e. treatment now is contrasted to what would 'normally' happen from now onwards including possible participation in the future. However, when discussing the policy implications of our results, one has to be careful because the nontreated individuals in one stratum who receive treatment later experience the impact of the treatment in future outcomes (see footnote 9 above). As discussed above, our approach mirrors the decision problem of the caseworker who has to decide between treatment now and treatment at a later point of time. This issue is an important one as typically about 20 percent of the members of the searching group participate in programs in the future (see appendix for detailed numbers). In order to check whether our treatment effects are systematically driven up because control group members are locked in future programs, we censor the information provided by later participants from the month they enter the program. The lower row of figure 20 shows that estimated treatment effects are not higher because the control group includes individuals that are in locked-in in programs that begin in later periods. Rather, we find the opposite, i.e. excluding later participants from the months they enter a program may increase estimated treatment effects. This suggest in fact that excluding future post treatment outcomes in the control group leads on average to an overoptimistic assessment of treatment effects because of the positive employment dynamics after the end of the lock-in period for future participants.

7 Summary and Conclusions

This paper analyzes and compares the employment effects of the three important types of public sector sponsored training in West Germany in the early 2000s. These are short-term training, classroom further training, and practical further training. In light of recent policy reforms fostering shorter training programs, we are particularly interested in the question of how short-term training programs compare in terms of effectiveness to traditional longer-term further training schemes, and how practically-oriented programs compare to theoretically-oriented programs. Our econometric approach combines exact matching and kernel matching methods in a dynamic, multiple treatment framework. Having access to rich administrative data, we also investigate to what extent data features and specification choices influence the evaluation results.

Our results suggest that the effectiveness of the different programs strongly depends on the gender of the participants and the timing of program participation during unemployment. We find statistically significant positive employment effects for male and female participants in short-term training and classroom further training who started their training not too early during their unemployment spell. There are no positive employment effects for individuals who started training early in their unemployment spell. Moreover, women but not men benefit from practical training. A closer look reveals that, over the first two to two and a half years after program start, employment effects of short-term training were of a similar magnitude as those of traditional longer-term programs, but, due to the shorter length, positive effects materialize much earlier. According to our results, both short-term and longer-term training may increase the employment rate of their participants by some 5 to 10 percentage points for men, and by an even larger amount for women. The somewhat surprising effectiveness of short-term training in terms of employment probabilities is also confirmed in pairwise comparisons to longer-term further training. Also, our results imply that practical training may have advantages over pure classroom training in terms of employment probabilities, a finding that is consistent with the international evidence (see Martin and Grubb (2001) and OECD (2005)), but not with evidence for Germany in the 1990's (Fitzenberger and Völter (2007), Fitzenberger et al. (2010), Lechner et al. (2010)).

The comparison is not as clear cut regarding earnings effects. Similar to the comparison of classroom vs. practical further training, the relative effectiveness of short-term vs. the longer further training is reduced if in addition earnings effects are taken into account. Although all three types of training programs generate significant earnings gains in the cases where there are also positive employment effects, earnings gains induced by classroom and practical further training generally appear to be larger than those of short-term training. This is true both when participation in one of the training programs is compared to not participating in any of the programs, and when the different training programs are directly compared with each other. As to the comparison between practical and classroom further training the evidence is mixed. Practical training seems to lead to higher earnings gains for women but not so for men.

Given that short-term and longer-term training yield comparable employment gains but longer-term training is more effective in terms of earnings, it is not entirely clear whether the policy shift that took place in 2002 and that massively substituted longer-term training courses by inexpensive short-term courses was justified. One has to keep in mind however, that short-term training programs are (due to their much shorter durations) associated with much shorter lock-in periods and considerably lower total costs. Although we lack detailed information on training costs, the results in table 2 combined with the information on average costs in table 1 suggest that an average short-term training course costing only 627 Euros may lead to a long-term employment gain comparable to that of an average further training course costing 6,175 Euros. In the very short run, longer-term training is dominated by short-term training because of the longer lock-in period associated with longer-term programs. Although a general cost-benefit analysis seems difficult, simple calculations raise some doubts whether the more expensive longer-term programs can be cost-effective in a fiscal sense. Considering a period of five years and even ignoring the lock-in effect during the first year, providing longer-term further training to 100 unemployed individuals would cost some 617,500 Euros (compared to 62,700 Euros in the case of short-term training). Assuming a positive employment effect of plus 10 percentage points, this would bring on average 10 out of the 100 unemployed back into employment, saving some 10 $\times$ 1400 Euros $\times$ 12 (months) $\times$ 4 (years)

= 672,000 Euros in unemployment compensation the Federal Employment Office would have to pay otherwise (based on an assumed monthly unemployment benefit level of 1400 Euros). Note that in many cases, employment effects are smaller or even zero. Even if other aspects such as earnings gains are taken into account, it will take a rather long time before longer-term further training programs amortize their costs. By comparison, the picture looks more positive in the case of short-term training.

While it is a common finding shared by studies such as Hujer et al. (2006), Fitzenberger et al. (2008), and Lechner et al. (2007, 2010) that, especially if lock-in effects are taken into account, shorter training programs may outperform longer ones,[27] it seems surprising that short-term programs lasting only two to twelve weeks may have employment effects at all. Given that it is hard to believe that such short programs lead to substantial increases in human capital, other aspects may be more relevant. Looking at the particular contents of short-term training analyzed here, it seems more plausible that these programs help to activate their participants who may otherwise not look as intensively for new jobs as they do when they are assigned short-term training programs that often comprise elements of profiling, job search assistance, or monitoring. Our results are therefore in line with a number of recent studies that focus on the positive effects of increased job search assistance and activation, see e.g. Blundell et al. (2004), Weber and Hofer (2004), Fougère et al. (2005), Hujer et al. (2005), Crépon et al. (2005), and Van den Berg and Van der Klaauw (2006), and Fitzenberger et al. (2008). It should be emphasized that in a number of cases neither short-term nor medium-term training programs showed any significant positive effects at all, especially when they are assigned early during the unemployment spell.

Our results also show that the effects of training programs may be very different across different subgroups. One result is that employment effects are usually larger for individuals who start their program at a later point during their unemployment spell. In fact, in many cases we do not find significant employment effects for individuals who start their treatment very early in their unemployment spell. It would be wrong to conclude from this that treatment is the more effective the later it is provided to the participants as individuals who are long-term unemployed may differ in observed and unobserved characteristics from those who are short-term unemployed. However, the result is remarkable because it suggests that in cases of long-term unemployment, training programs, as we observe them, do help to restore

[27]However, note that neither of these studies consider the very short training programs analyzed here.

the employment chances of their participants. We also find that training effects are heterogeneous with respect to gender. In line with other results in the literature (see Bergemann and van den Berg (2010) for an overview), we find that employment effects of training are generally larger for women than for men. The finding that program effects are generally larger for women than for men has been questioned by Lechner and Wiehler (2009) whose analysis (using Austrian data) suggests that positive effects might be overstated if women respond to prolonged unemployment by adjusting their family planning behavior. While we cannot completely rule out such an effect because of missing information on pregnancies in our data, we also do not find heterogeneity in treatment effects for women below vs. above 40 years of age, a finding which may be interpreted as indirect evidence against such an fertility effect. Carrying out a further detailed analysis of effect heterogeneity, we do not find much further heterogeneity along observed characteristics.

Finally, this paper also presents a careful sensitivity analysis examining to what extent data and specification features influence our evaluation outcomes. We believe that such a sensitivity analysis conveys useful information about the adequate design of evaluation studies for similar settings as ours. Our results suggest that the most important feature of the employment history that should be controlled for is the unemployment duration prior to treatment. Matching in addition on past employment sequences and accounting for long histories of employment and benefit information seems to play a surprisingly small role. Our results further indicate that the availability of rich personal information, information on training programs other than the ones in focus as well as the amount of specification search may have a considerable impact on evaluation results. Surprisingly, personality variables and information on caseworker assessments of various aspects made little difference for our results as long as other personal characteristics and labor market histories were controlled for. Finally, our sensitivity analysis shows that variations in the dynamic definition of the control group may have strong consequences for evaluation results. In particular, in line with Fredriksson and Johansson (2008), we find that confining the control group to individuals who are never treated leads to considerably lower estimates of treatment effects compared to a dynamic stratification that aims to avoid conditioning on future outcomes.

References

Abadie, A. and G. Imbens (2002). "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects." *NBER Technical Working Paper* T0283.

Abadie, A. and G. Imbens (2006). "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74, 235-267.

Barnow, B.S. (1987). "The Impact of CETA Programs on Earnings. A Review of the Literature." *Journal of Human Resources* 22, 157-193.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, S. Lischke, R. Miquel, A. Osikominu, T. Wenzel, C. Wunsch (2004). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB – Vorläufiger Zwischenbericht Oktober 2004." Goethe University Frankfurt and SIAW St. Gallen.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, R. Miquel, A. Osikominu, M. Waller, and C. Wunsch (2005). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB – Zwischenbericht Oktober 2005." Goethe University Frankfurt and SIAW St. Gallen.

Bergemann, A., S. Speckesser, and B. Fitzenberger (2009). "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Differences-in-Differences." *Journal of Applied Econometrics* 24, 797-823.

Bergemann, A. and G.J. van den Berg (2010). "Active Labor Market Policy Effects for Women in Europe: A Survey." *Annales d'Economie et de Statistique*, forthcoming.

Biewen, M. and B. Fitzenberger (2004). "Neuausrichtung der Förderung beruflicher Weiterbildung." In: Hagen, T. and A. Spermann (Ed.): *Hartz-Gesetze – Methodische Ansätze zu einer Evaluierung*, ZEW Wirtschaftsanalysen, Bd. 74, Baden-Baden, 190-202.

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Waller (2007). "Which program for whom: Evidence on the comparative effectiveness of public sponsored training programs in Germany." *IZA Discussion Paper* No. 2885.

Blaschke, D. and H.-E. Plath (2000). "Möglichkeiten und Grenzen des Erkenntnisgewinns durch Evaluation aktiver Arbeitsmarktpolitik." *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung (MittAB)* 33, 462-482.

Bloom, H.S., L.L. Orr, S.H. Bell, G. Cave, F. Doolittle, W. Lin, and J.M. Bos (1997). "The Benefits and Costs of JTPA Title II-A Programs." *Journal of Human Resources* 32, 549-576.

Blundell, R., M. Costas Dias, C. Meghir, and J. Van Reenen (2004). "Evaluating the Employment Impact of a Mandatory Job Search Program." *Journal of the European Economic Association* 2, 569-606.

Bundesagentur für Arbeit (2001, 2002a, 2003a, 2004a). *Daten zu den Eingliederungsbilanzen 2000, 2001, 2002, 2003*, Nürnberg (various issues).

Bundesagentur für Arbeit (2002b, 2003b, 2004b, 2005a). *Arbeitsmarkt 2001, 2002, 2003, 2004*, Nürnberg (various issues).

Bundesagentur für Arbeit (2005b). *Geschäftsbericht 2004*, Nürnberg.

Carcillo, S. and D. Grubb (2006). "From Inactivity to Work: The Role of Active Labour Market Policies." *OECD Social Employment and Migration Working Papers* No. 36.

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

Crépon, B., M. Dejemeppe, and M. Gurgand (2005). "Counseling the unemployed: does it lower unemployment duration and recurrence?" *IZA Discussion Paper* 1796.

Crépon, B, M. Ferracci, G. Jolivet, and G.J van den Berg (2009). "Active Labor Market Policy Effects in a Dynamic Setting." *Journal of the European Economic Association* 7, 595-605.

DeLuna, X. and P. Johansson (2009). "Non-Parametric Inference for the Effect of a Treatment on Survival Times with Application in the Health and Social Sciences." *IZA Discussion Paper* No. 3966.

Dolton, P. and D. O'Neill (2002). "The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Experimental Evidence from the United Kingdom." *Journal of Labor Economics* 20, 381-403.

Dolton, P., J. Smith, and J.P. Azevedo (2008). "The Impact of the UK New Deal for Lone Parents on Benefit Receipt." Working Paper, London School of Economics.

Dorsett, R. (2006). "The new deal for young people: effect on the labour market status of young men Labour Economics." *Labour Economics* 13, 405-422.

Dyke, A., C.J. Heinrich, P.R. Mueser, K.R. Troske, and K.-S. Jeon (2006): "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes." *Journal of Labor Economics* 24, 567-607.

Fitzenberger, B., O. Orlyanskaya, A. Osikominu, and M. Waller (2008). "Déjà Vu? Short–Term Training in Germany 1980–1992 and 2000–2003." *IZA Discussion Paper* No. 3540.

Fitzenberger, B., A. Osikominu, and R. Völter (2010). "Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany." *Annales d'Economie et de Statistique* 91, 315-348.

Fitzenberger, B., A. Osikominu, and R. Völter (2006). "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample." *Journal of Applied Social Science Studies* 126, 405-436.

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

Fitzenberger, B. and R. Völter (2007). "Long Run Effects of Training Programs for the Unemployed in East Germany." *Labour Economics* 14, 730 - 755.

Fougère, D., J. Pradel, and M. Roger (2005). "Does Job Search Assistance Affect Search Effort and Outcomes?" *IZA Discussion Paper* 1825.

Fredriksson, P. and P. Johansson (2008). "Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Data." *Journal of Business and Economic Statistics* 26, 435-445.

Frölich, M. (2008). "Statistical Treatment Choice: An Application to Active Labour Market Programmes." *Journal of American Statistical Association* 103, 547-558.

Gerfin, M. and M. Lechner (2002). "Microeconometric Evaluation of the Active Labor Market Policy in Switzerland." *Economic Journal* 112, 854-893.

Hardoy, I. (2005). "Impact of Multiple Labour Market Programmes on Multiple Outcomes: The Case of Norwegian Youth Programmes." *LABOUR* 19, 425-467.

Heckman, J., H. Ichimura, and P. Todd (1997). "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64, 605-654.

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 J. Smith (1999). "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies." *Economic Journal* 109, 313-348.

Heinrich, C.J., P.R. Mueser, K. Troske, K.-S. Jeon, and D.C. Kahvecioglu (2010). "New Estimates of Public Employment and Training Program Net Impacts: A Nonexperimental Evaluation of the Workforce Investment Act Program", *Discussion Paper*, University of Missouri, available under `http://economics.missouri.edu/working-papers/2010/WP1003_Mueser.pdf` (downloaded on June 15, 2010).

Hotz, V.J., Imbens, G. and J.A. Klerman (2006). "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program." *Journal of Labor Economics* 24, 521-566.

Hujer, R., M. Caliendo and S. Thomsen (2005). "Mikroökonometrische Evaluation des Integrationserfolges." *Beiträge zur Arbeitsmarkt- und Berufsforschung (BeitrAB)* 291, Nürnberg.

Hujer, R., S.L. Thomsen, and C. Zeiss (2006). "The Effects of Short–Term Training Measures on the Individual Unemployment Duration in West Germany." Discussion Paper 06–065, ZEW Mannheim.

Hui, S. and J. Smith (2003a). "The labour market impacts of adult education and training in Canada." Research Report, Statistics Canada.

Hui, S. and J. Smith (2003b). "Issues in the design of Canada's Adult Education and Training Survey." Research Report, Statistics Canada.

Imbens, G. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* 87, 706-710.

Jespersen, S.T., J.R. Munch, and L. Skipper (2008). "Costs and benefits of Danish active labour market programmes." *Labour Economics* 15, 859-884.

Kluve, J. (2010). "The Effectiveness of European Active Labor Market Policy." *Labour Economics*, forthcoming.

Kluve, J. and C.M. Schmidt (2002). "Can training and employment subsidies combat European unemployment?" *Economic Policy* 35, 409-448.

Kurtz, B. (2003). "Trainingsmaßnahmen - Was verbirgt sich dahinter?" IAB Werkstattbericht Nr. 8, Nürnberg.

Larsson, L. (2003). "Evaluation of Swedish Youth Labor Market Programs." *Journal of Human Resources* 38, 891-927.

Lechner, M. (2001). "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption." In: M. Lechner and F. Pfeiffer (eds.): *Econometric Evaluation of Active Labor Market Politics in Europe*, Heidelberg: Physica-Verlag.

Lechner, M. (2002). "Program heterogeneity and propensity score matching: An Application to the evaluation of active labor market policies". *Review of Economics and Statistics* 84, 205-220.

Lechner, M., R. Miquel, and C. Wunsch (2007). "The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification." *German Economic Review* 8, 468-507.

Lechner, M., R. Miquel, and C. Wunsch (2010). "Long-Run Effects of Public Sector Sponsored Training in West Germany." *Journal of the European Economic Association*, forthcoming.

Lechner, M. and S. Wiehler (2009). "Kids or courses? Gender differences in the effects of active labor market policies." *Journal of Population Economics*, forthcoming.

Lechner, M. and C. Wunsch (2009). "Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off." *The Economics of Transition* 4, 661-702.

Martin, J.P. (2000). "What works among active labour market policies: Evidence from OECD countries' experiences." *OECD Economic Studies* No. 30, 79-113.

Martin, J.P. and D. Grubb (2001). "What works for whom: A review of OECD countries' experiences with active labour market policies." *Swedish Economic Policy Review* 8, 96.

Mueser, P.R., K.R. Troske, and A. Gorislavsky (2007). "Using state administrative data to measure program performacnce." *Review of Economics and Statistics* 89, 761-783.

OECD (2005). "Labour Market Programmes and Activation Strategies: Evaluating the Impacts." *OECD Employment Outlook*, Ch. 4.

Park, N., B. Power, W.C. Riddell, and G. Wong (1996). "An Assessment of the Impact of Government-Sponsored Training." *The Canadian Journal of Economics* 29, S93-S98.

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. Steinwedel, and 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. (2004). "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s." *Review of Economics and Statistics* 86, 133-155.

Sianesi, B. (2008). "Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults in the 1990s." *Labour Economics* 15, 370-399.

Van den Berg, G.J. and B. van der Klaauw (2006). “Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment”. *International Economic Review* 47, 895-936.

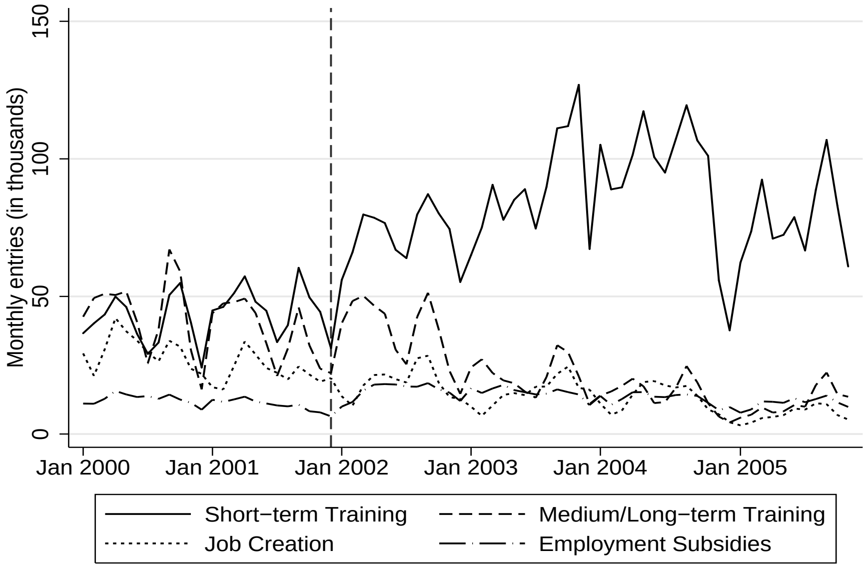
Waller, Marie (2008). “On the Importance of Correcting Reported End Dates of Labor Market Programs.” *Journal of Applied Social Science Studies* 128, 213-236.

Weber, A. and H. Hofer (2004). “Employment effects of early interventions on job search programs.” *IZA Discussion Paper* No. 1076.

Wunsch, C. and L. Lechner (2008). “What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes.” *Kyklos* 61, 134–174.

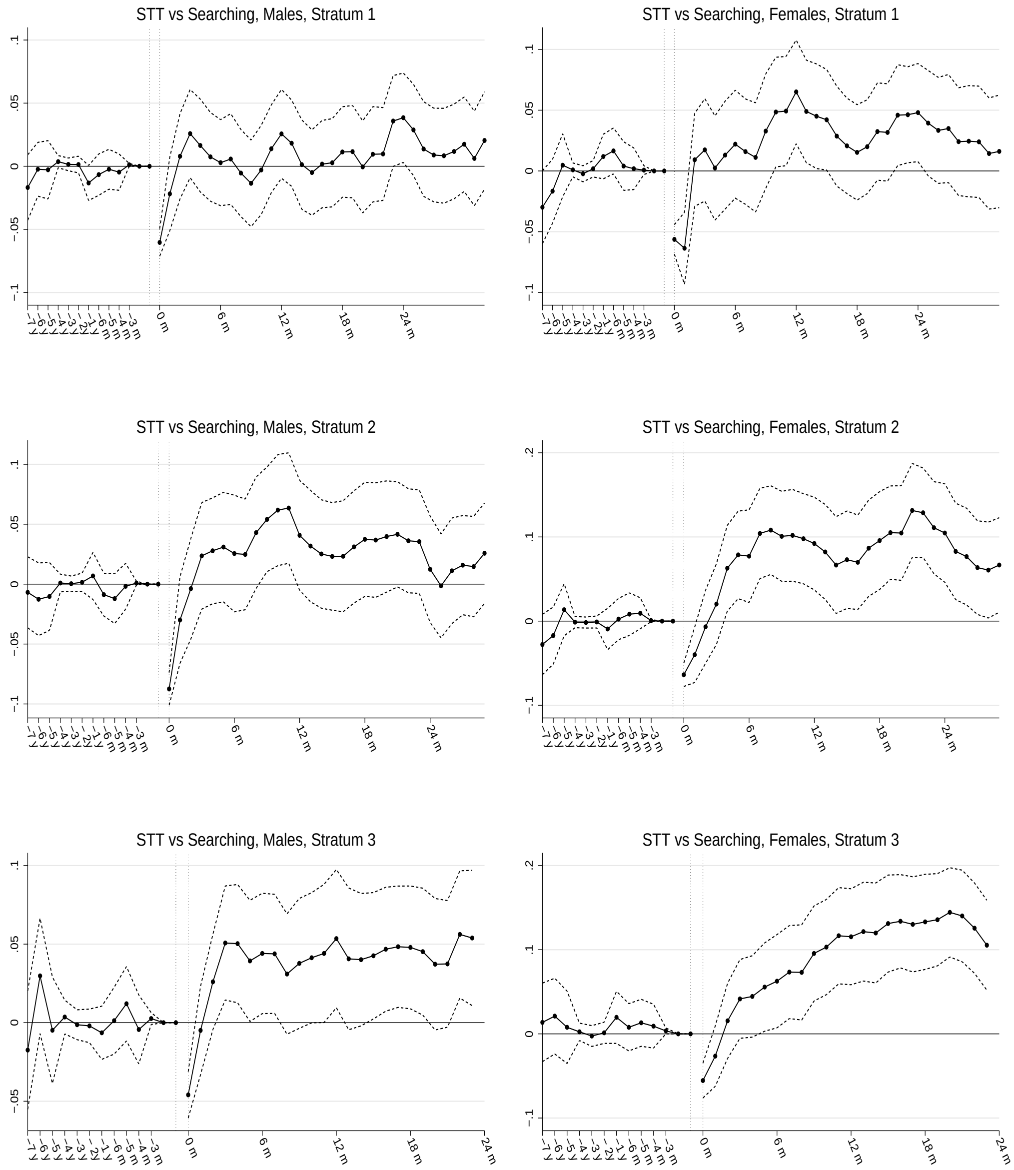
Figures

Figure 1: Active Labor Market Policies in Germany



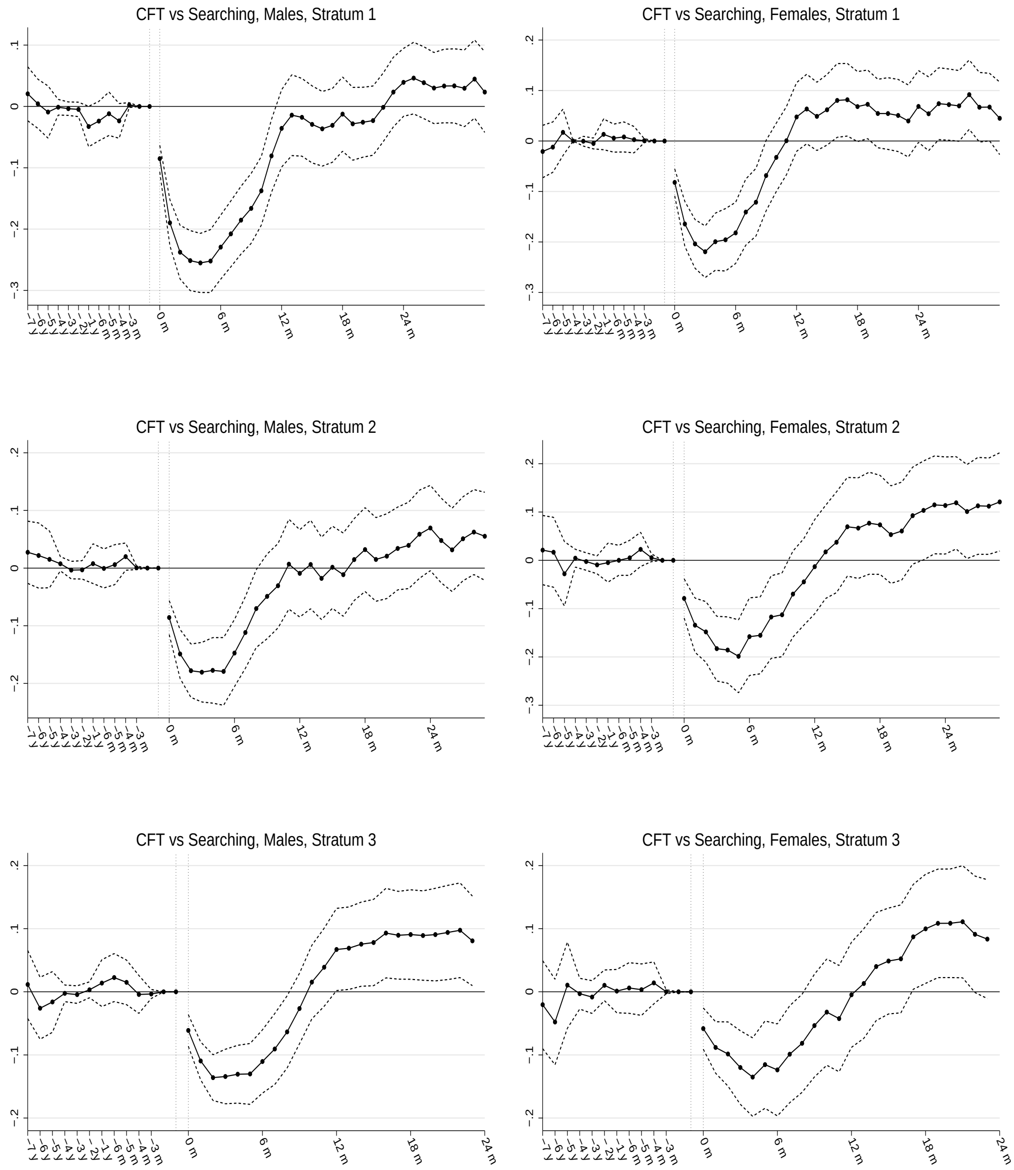
Source: Bundesagentur für Arbeit (2002b-2005a).

Figure 2: Employment Effects for STT vs. Searching



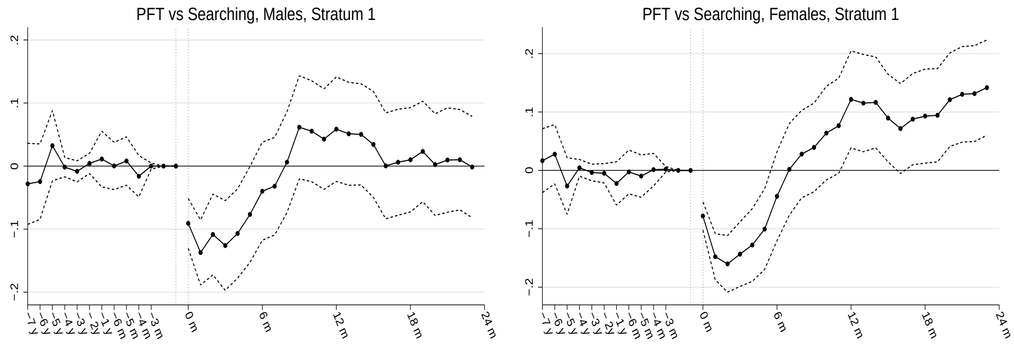
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figure 3: Employment Effects for CFT vs. Searching



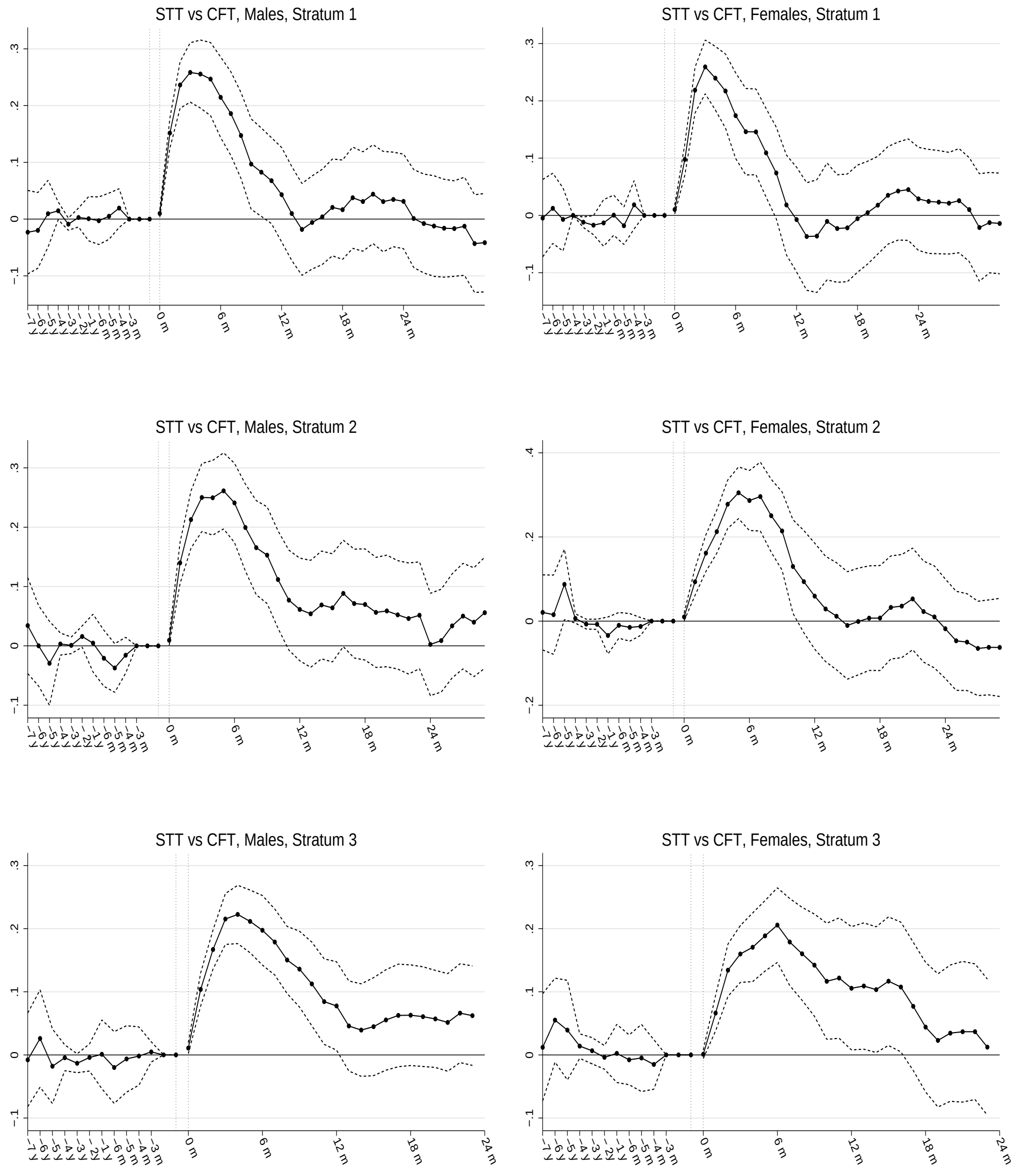
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figure 4: Employment Effects for PFT vs. Searching



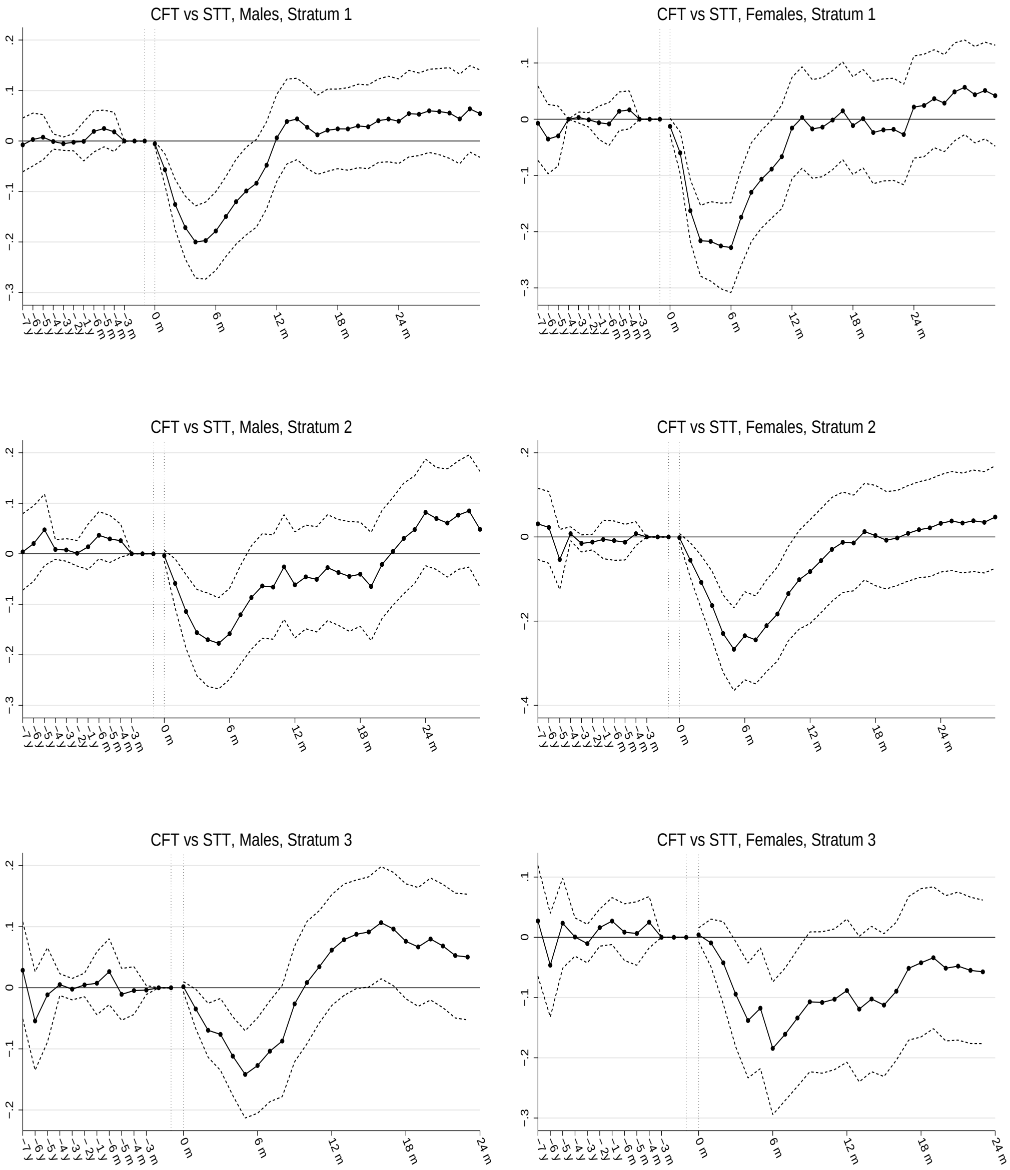
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

Figure 5: Employment Effects for STT vs. CFT



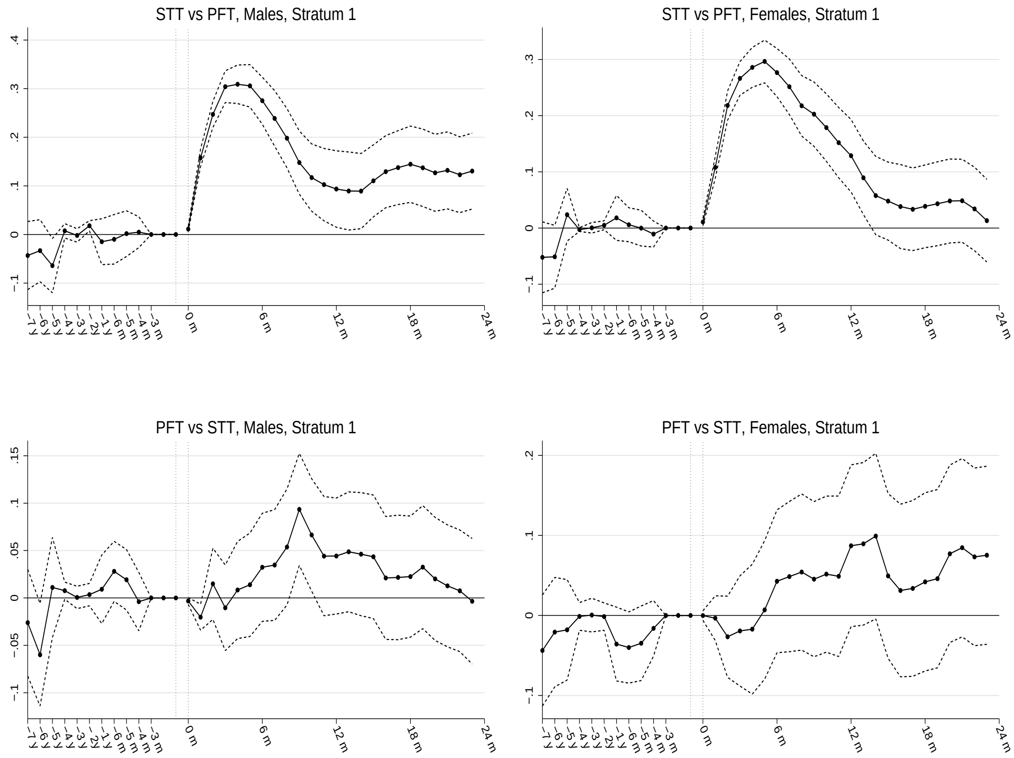
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment ($\geq$ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figure 6: Employment Effects for CFT vs. STT



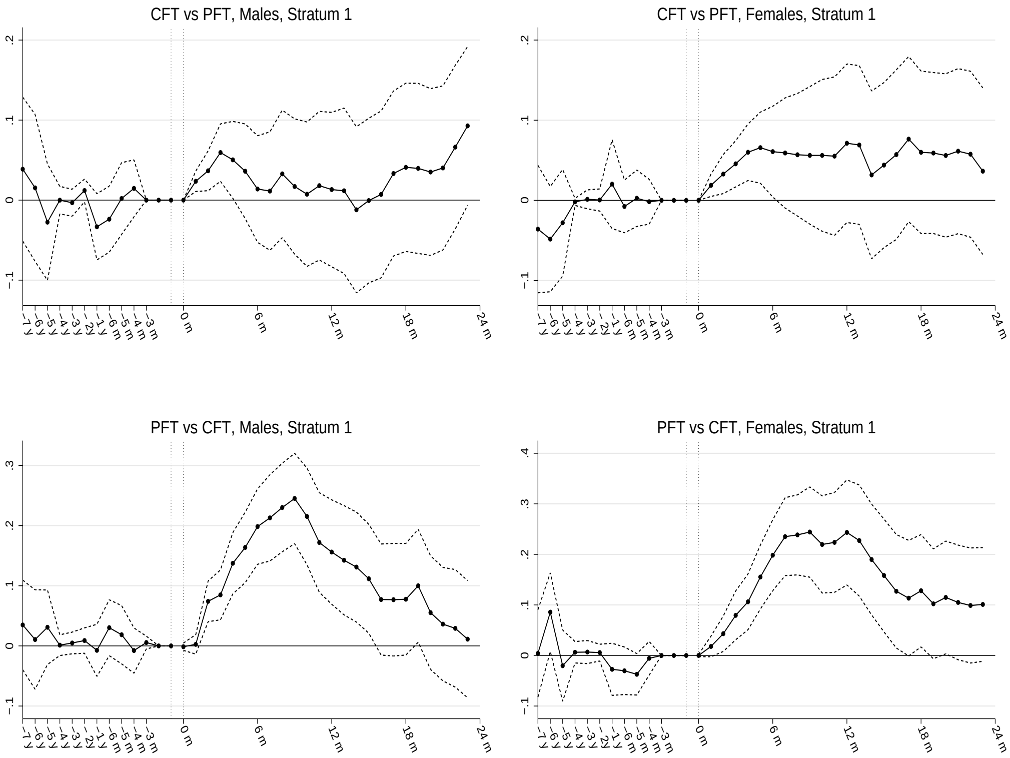
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment ($\geq$ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figure 7: Employment Effects for STT vs. PFT and PFT vs. STT



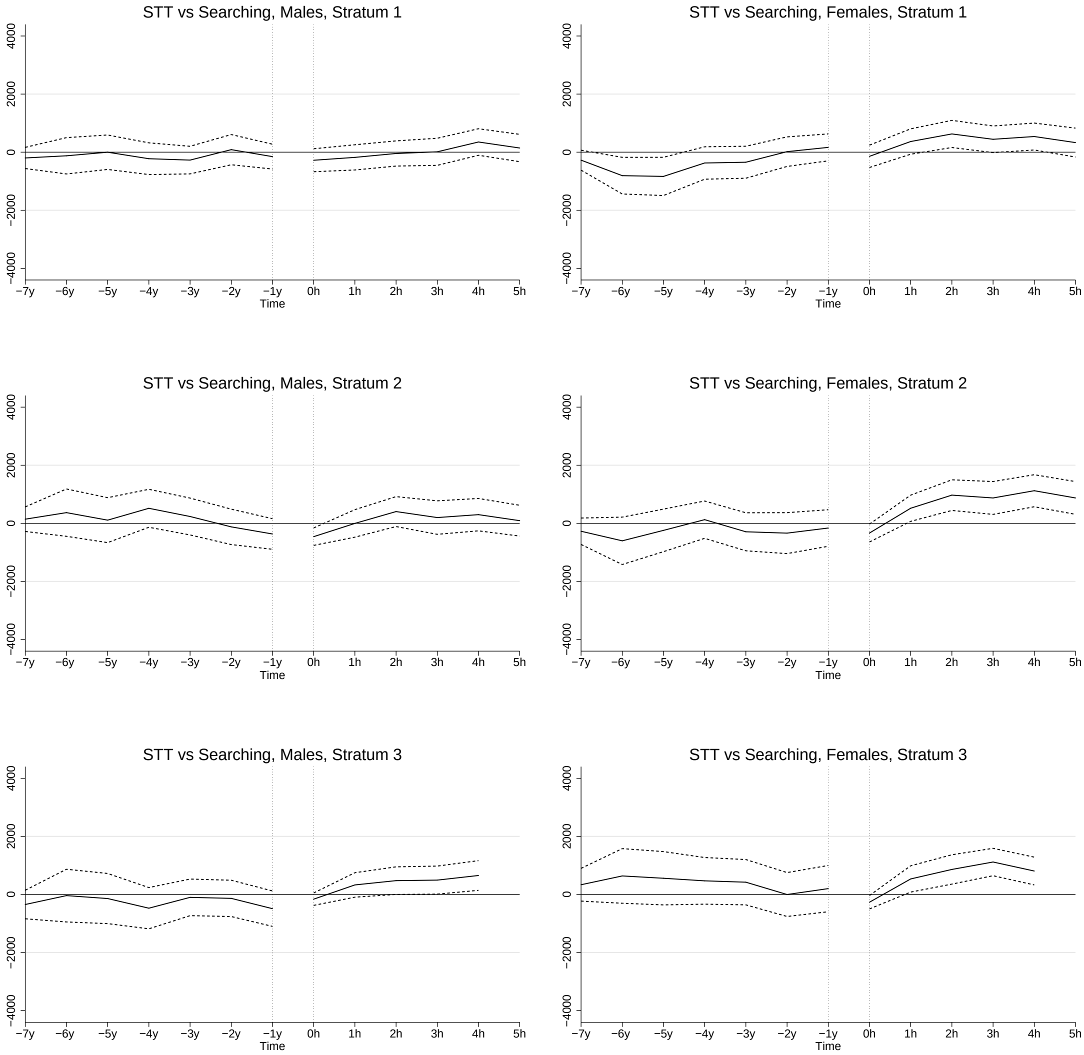
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

Figure 8: Employment Effects for CFT vs. PFT and PFT vs. CFT



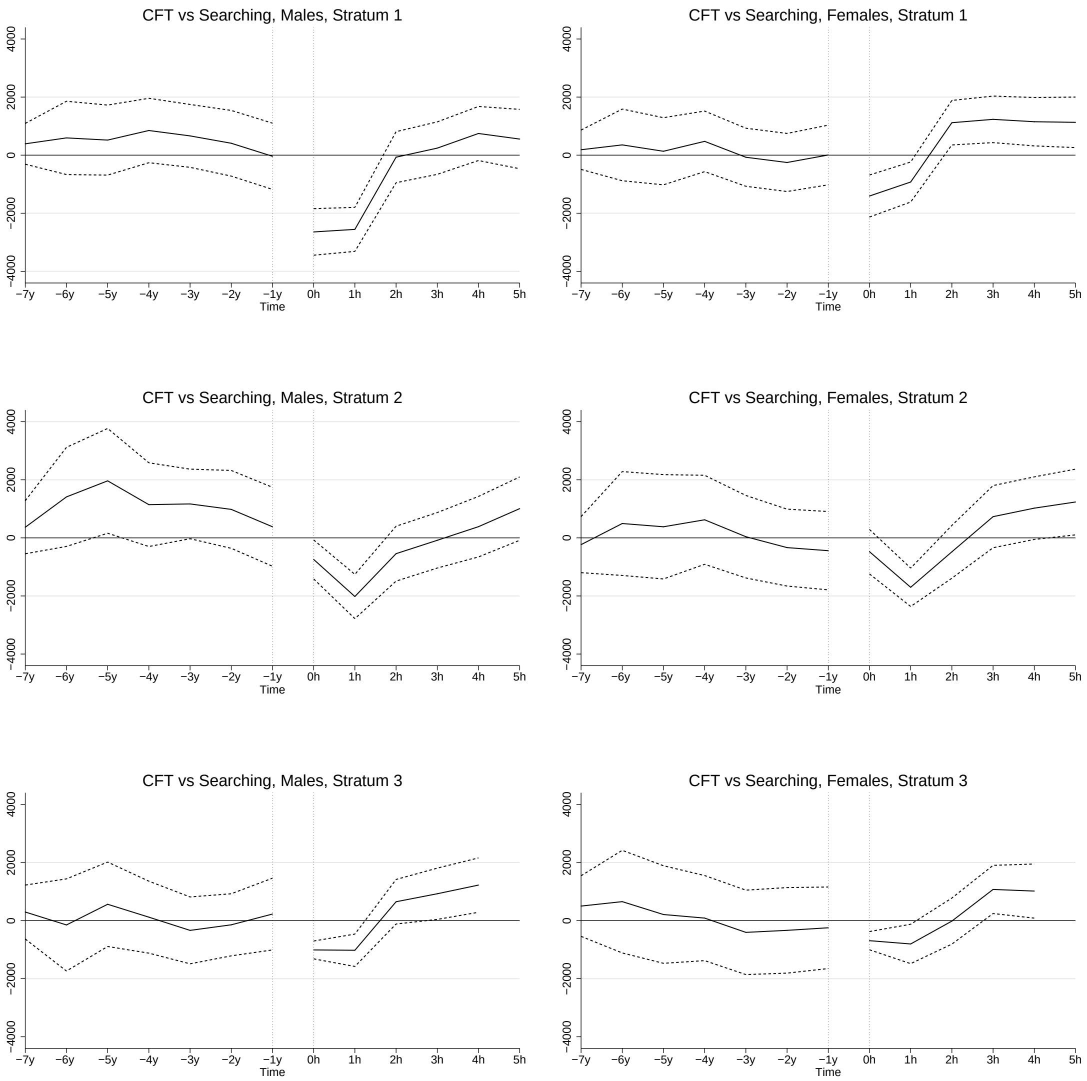
Note: Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

Figure 9: Earnings Effects for STT vs. Searching



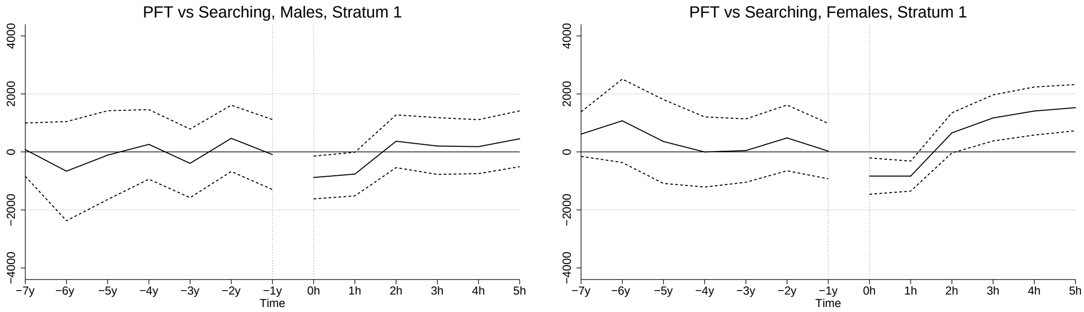
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 10: Earnings Effects for CFT vs. Searching



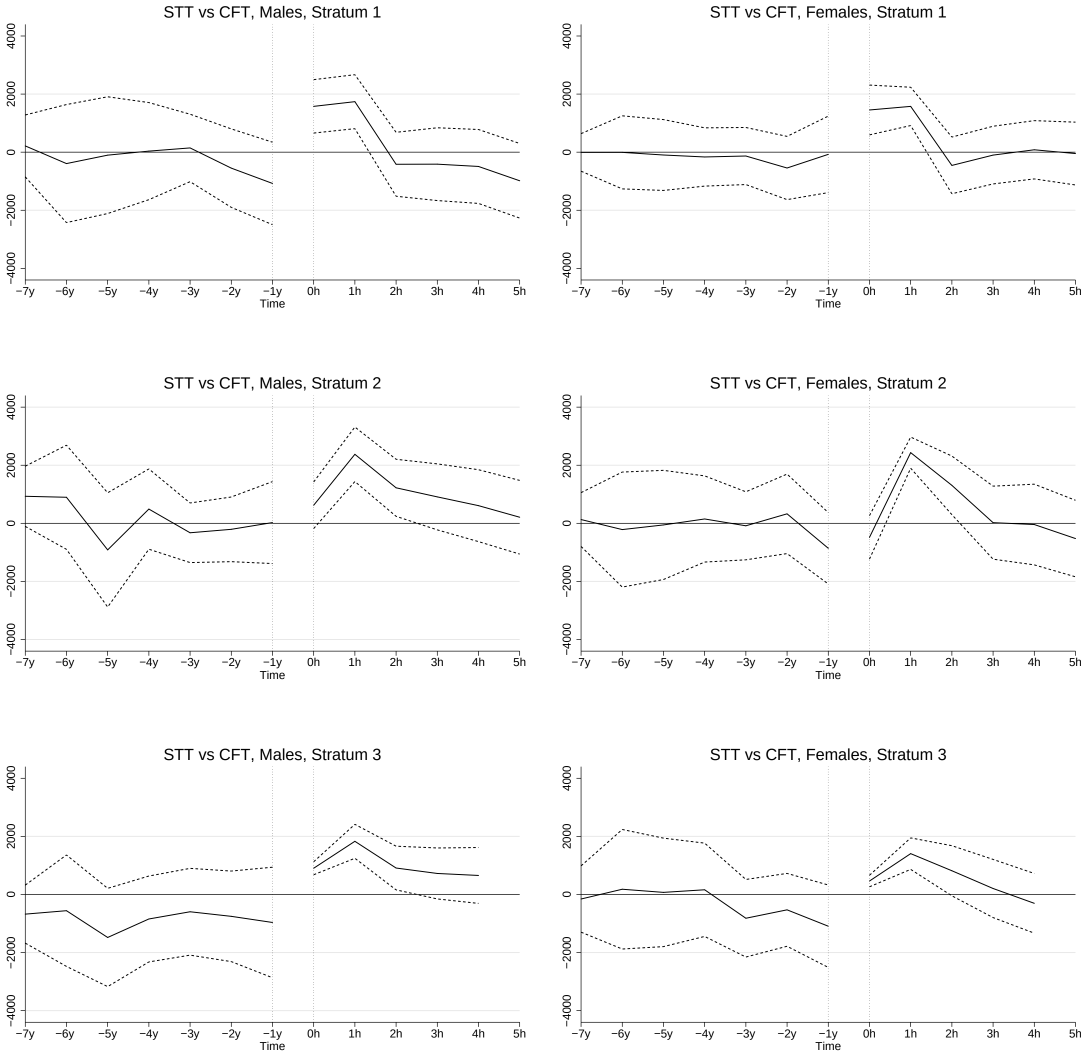
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 11: Earnings Effects for PFT vs. Searching



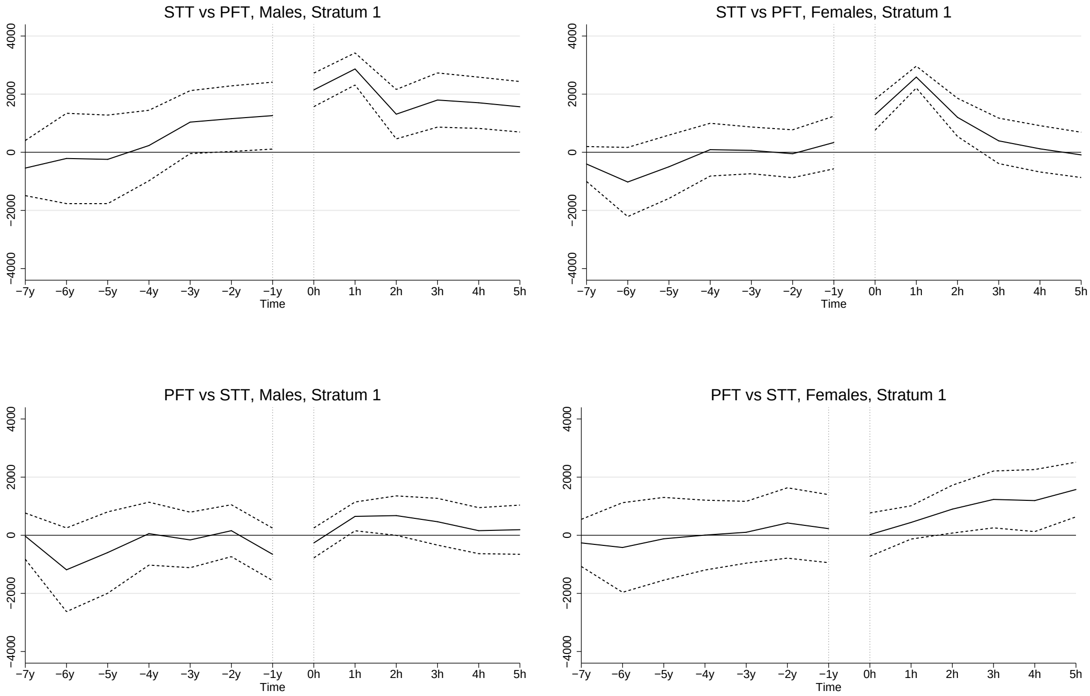
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 12: Earnings Effects for STT vs. CFT



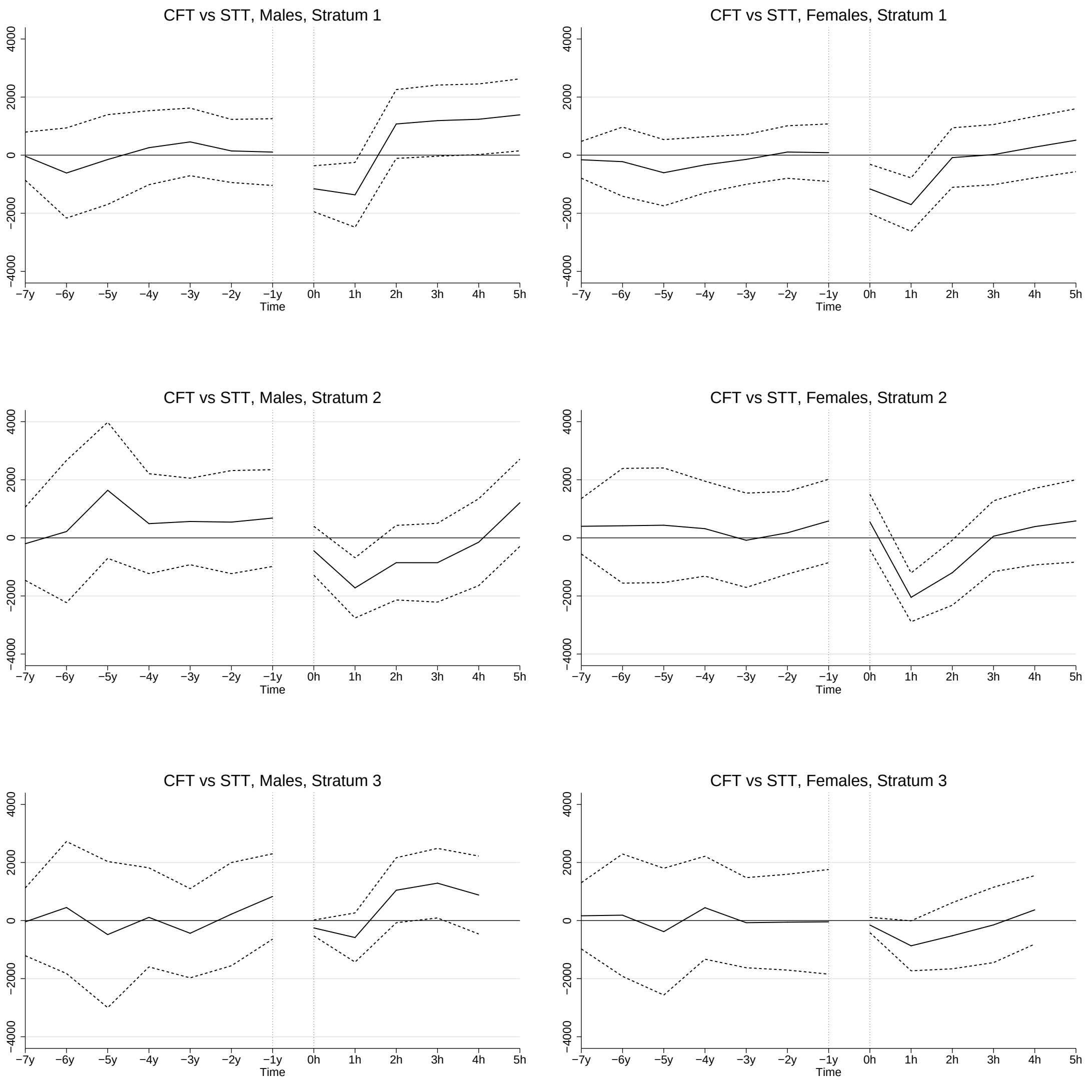
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 13: Earnings Effects for STT vs. PFT and PFT vs. STT



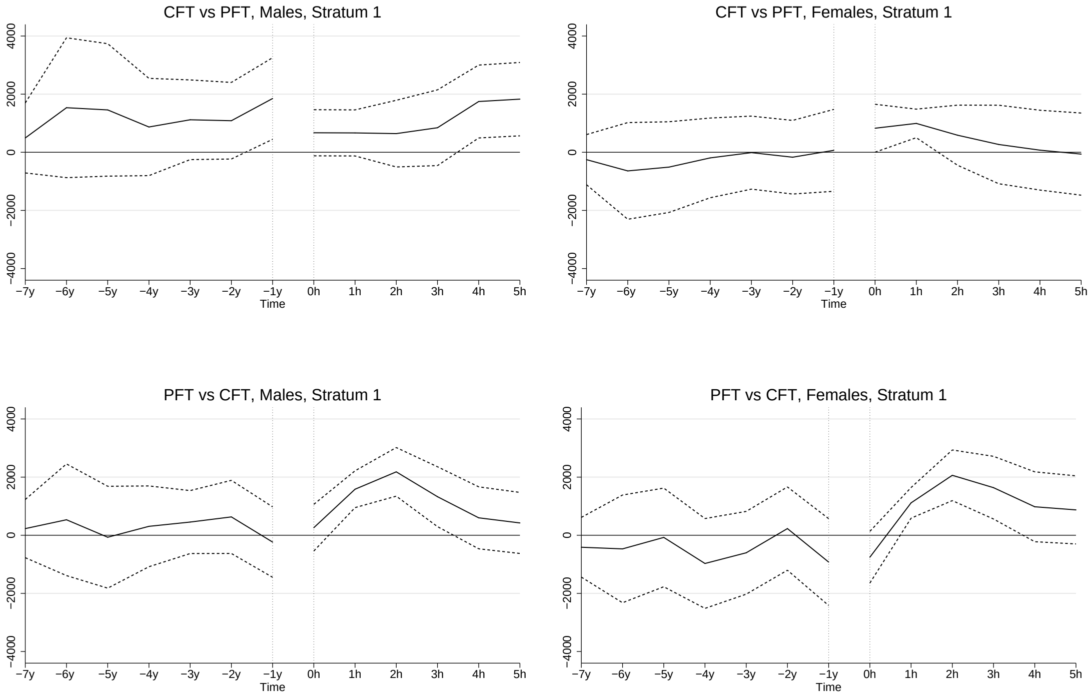
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 14: Earnings Effects for CFT vs. STT



Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

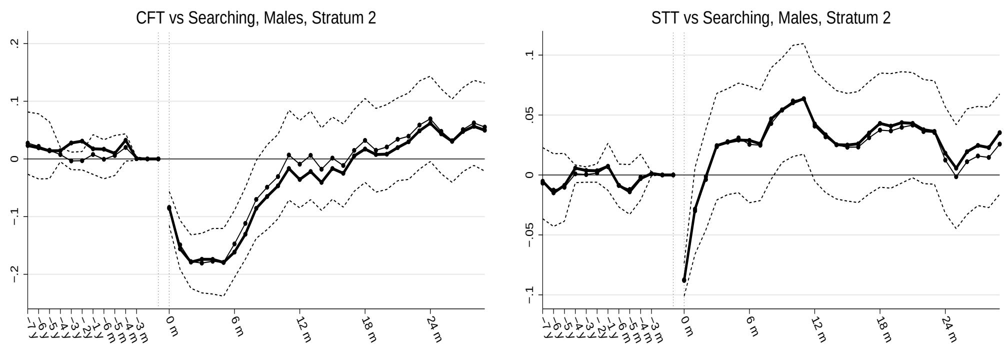
Figure 15: Earnings Effects for CFT vs. PFT and PFT vs. CFT



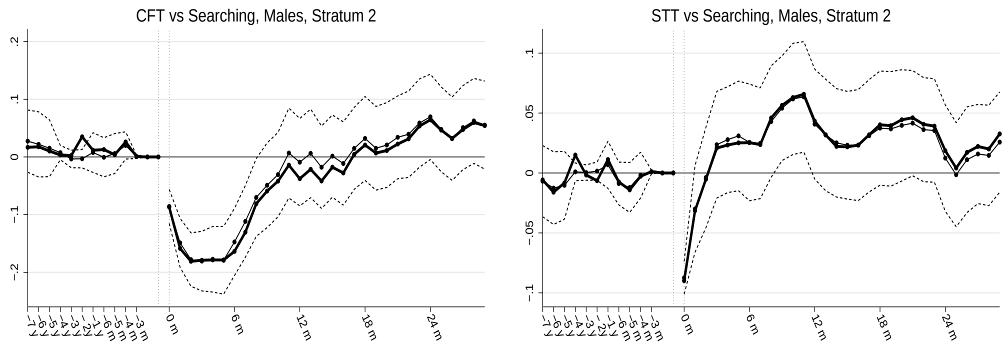
Note: The abscissae measures pre-unemployment (< 0) and post-treatment (≥ 0) time intervals. Pre-unemployment intervals cover a period of one year (=364 days), post-treatment intervals half a year (=182 days). The difference in average real earnings (in Euro, reference year 2000) during the corresponding time interval is given on the ordinate.

Figure 16: Sensitivity Analysis 1: Employment History

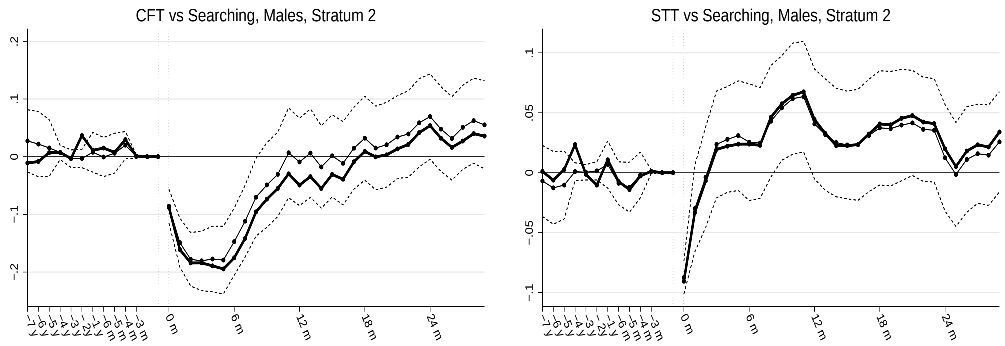
Step 1: No exact matching on employment histories



Step 2: In addition, no history sequence dummies in propensity score



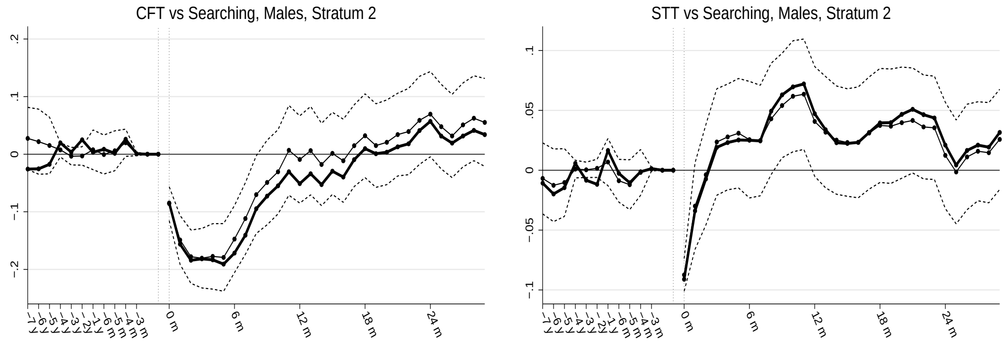
Step 3: In addition, the seven-year history information is dropped from in propensity score



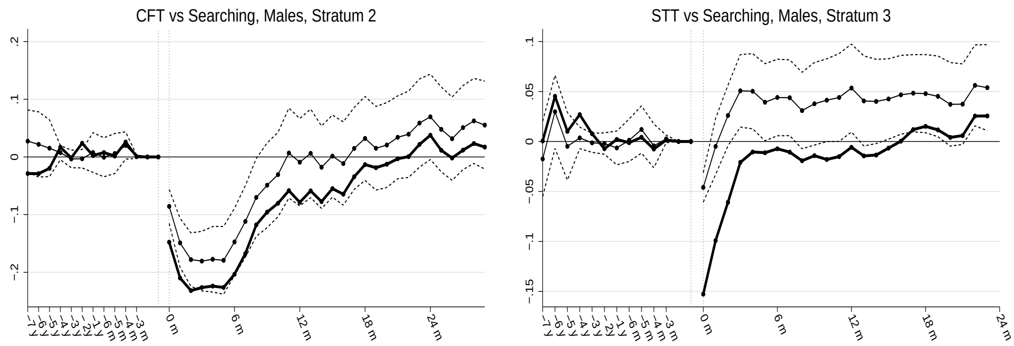
Note: The bold line refers to the alternative specification. The thin line and the confidence interval refers to the benchmark specification.

Figure 17: Sensitivity Analysis 1: Employment History

Step 4: In addition, everything related to benefit information dropped



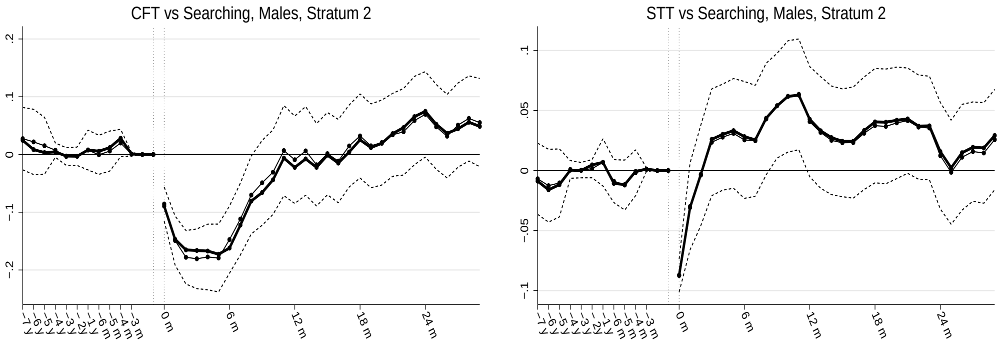
Step 5: In addition, no exact matching on unemployment duration



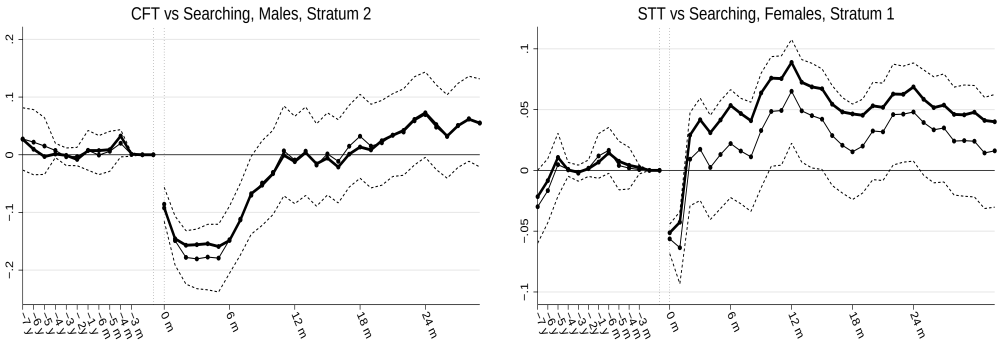
Note: The bold line refers to the alternative specification. The thin line and the confidence interval refers to the benchmark specification.

Figure 18: Sensitivity Analysis 2: Rich Personal Information and Specification Search

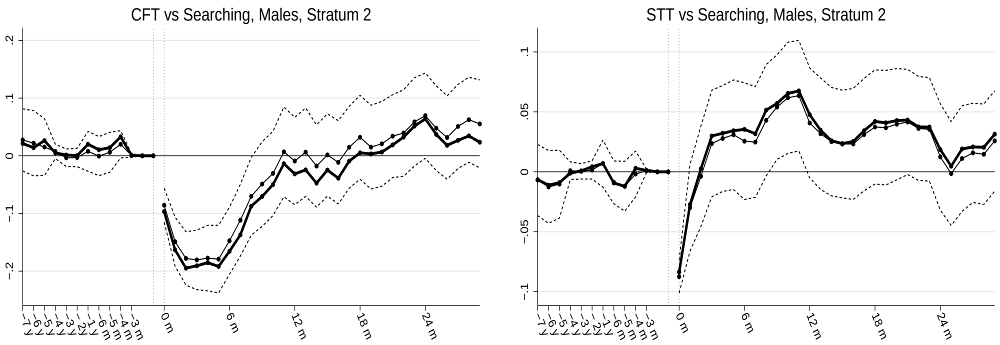
Step 1: No personality variables



Step 2: In addition, no rich personal information



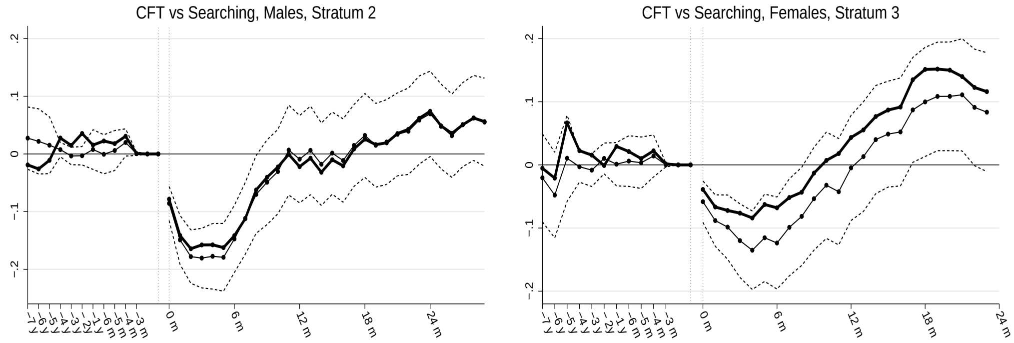
No detailed specification search



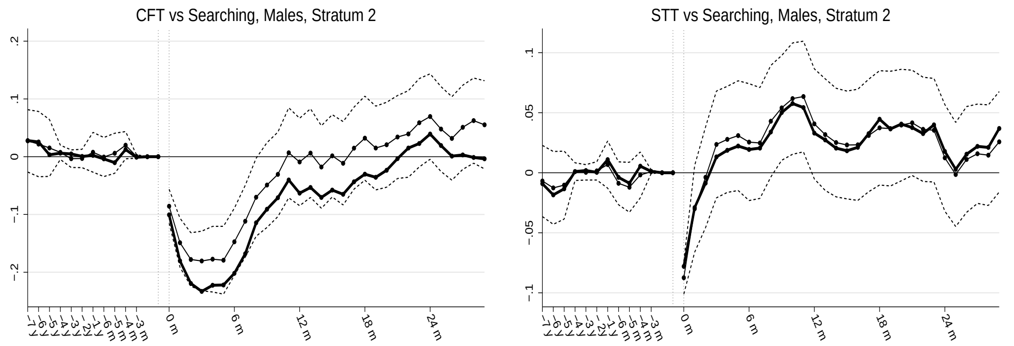
Note: The bold line refers to the alternative specification. The thin line and the confidence interval refers to the benchmark specification.

Figure 19: Sensitivity Analysis 3: Comparison with MTG and Information on Other Programs

Comparison with specification similar to MTG



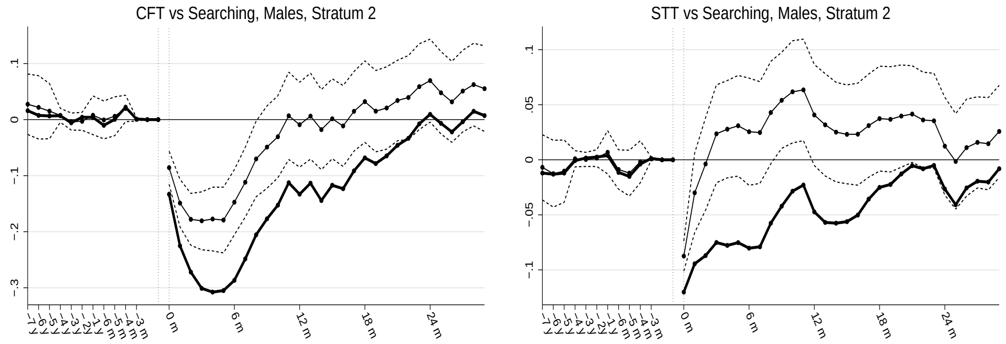
Ignoring information on other programs



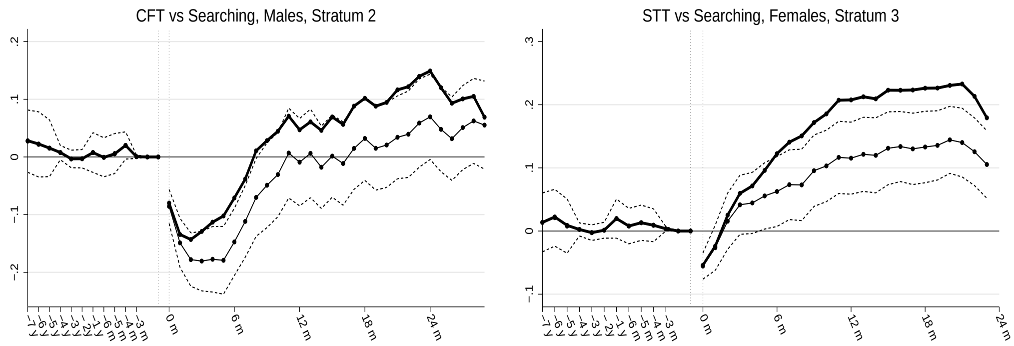
Note: The bold line refers to the alternative specification. The thin line and the confidence interval refers to the benchmark specification.

Figure 20: Sensitivity Analysis 4: Future Participation

Excluding future participants from the beginning of stratum



Excluding future participants from the month they enter a program



Note: The bold line refers to the alternative specification. The thin line and the confidence interval refers to the benchmark specification.

Tables

Table 1: Average Expenditures per Participant in Short-Term Training, Further Training and Retraining in Germany in 2000-2003

	2000		2001		2002		2003	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Short-term training	580	1,2	570	1,1	658	0,9	538	1
Further/retraining	1627	8,2	1668	9,3	1686	9,1	1555	10,5
– subsistence allowance	1152		1178		1188		1156	
– training costs	640		664		681		631	

Note: Columns labeled with a (1) contain the average monthly expenditures (in Euro) per participant, columns labeled with a (2) display the average duration of the program in months. Source: Bundesagentur für Arbeit (2001,2002a-2004a).

Table 2: Cumulated Treatment Effects After 24 Months

	Males			Females		
	Stratum 1	Stratum 2	Stratum 3	Stratum 1	Stratum 2	Stratum 3
STT vs Searching	0.098 (0.311)	0.636 (0.416)	0.907 (0.366)**	0.544 (0.382)	1.788 (0.488)***	2.136 (0.507)***
STT vs CFT	2.202 (0.656)***	2.814 (0.710)***	2.477 (0.543)***	1.715 (0.689)**	2.588 (0.879)***	2.453 (0.750)***
STT vs PFT	3.859 (0.576)***	.	.	3.086 (0.524)***	.	.
CFT vs Searching	-2.510 (0.463)***	-1.168 (0.591)**	0.077 (0.526)	-0.885 (0.576)	-0.833 (0.799)	-0.209 (0.699)
CFT vs STT	-1.096 (0.699)	-1.513 (0.899)*	0.181 (0.801)	-1.817 (0.754)**	-2.076 (0.961)**	-2.046 (1.007)**
CFT vs PFT	0.674 (0.730)	.	.	1.247 (0.730)*	.	.
PFT vs Searching	-0.297 (0.703)	.	.	0.722 (0.688)	.	.
PFT vs STT	0.645 (0.506)	.	.	1.020 (0.898)	.	.
PFT vs CFT	2.741 (0.673)***	.	.	3.471 (0.817)***	.	.

Note: Monthly ATTs (i.e. the difference in employment probabilities) are summed over the first 24 months from the program start onwards. For each comparison pair the treatment named first corresponds to the active treatment and the second one to the control treatment.

Appendix

7.1 Descriptive Statistics and Variable Description

Table 3: Entries into Active Labor Market Programs in Germany and West Germany from 2000–2004 (in Thousand)

	2000	2001	2002	2003	2004
			Germany		
Public sponsored training programs*	1,154	1,069	1,457	1,502	1,435
– Further/retraining	552	450	456	255	185
– Short-term training	477	565	877	1,064	1,188
Employment subsidies	459	465	538	808	909
Public Job Creation	314	246	217	189	166
Placement and advisory services	601	742	934	1,460	2,795
Specific measures for young adults	446	496	447	389	408
Other	391	478	469	217	105
Total	3,365	3,496	4,062	4,565	5,818
			West Germany		
Public sponsored training programs*	770	643	972	985	958
– Further/retraining	338	261	273	161	124
– Short-term training	286	339	545	690	789
Employment subsidies	225	206	245	365	451
Public Job Creation	89	73	63	39	42
Placement and advisory services	279	296	375	640	1,447
Specific measures for young adults	364	191	210	262	270
Other	125	370	289	17	85
Total	1,852	1,778	2,154	2,308	3,253

Source: Bundesagentur für Arbeit, Arbeitsmarkt 2002b-2005b, own calculations.

Note: Programs marked with a * are the ones that are analyzed in the paper. All other programs are not explicitly analyzed, but we take into account whether a person participated in such a program when evaluating the programs we are interested in.

Table 4: Sample Sizes

	Males	Females
	Stratum 1 (0-3 Months)	
Searching	29351	18409
STT	912	693
CFT	389	344
RT	263	262
	Stratum 2 (4-6 Months)	
Searching	18529	12572
STT	547	409
CFT	251	194
	Stratum 3 (7-12 Months)	
Searching	10996	8421
STT	662	497
CFT	270	201
	Aggregated Stratum 1 for PFT (0-12 Months)	
Searching	25854	16060
STT	2120	1593
CFT	915	741
PFT	263	234

Searching	:	Being unemployed at the beginning of stratum and not having started a program during stratum
STT	:	Short–term training
CFT	:	Classroom further training
PFT	:	Practical further training
RT	:	Retraining

Table 5: Variable Definitions

Name	Definition
female	1 if female, 0 otherwise
agegroup	Dummies for the following agegroups: 1) 25 - 29 years, 2) 30 - 34 years, 3) 35 - 39 years, 4) 40 - 44 years, 5) 45 - 49 years, 6) 50 - 54
foreigner	1 if citizenship is not German, 0 otherwise
ethnicgerman	1 if ethnic German, i.e. returned settler from former German settlements, 0 otherwise
qualification	Dummies for the following categories: 1) no degree, 2) vocational training degree, 3) university or technical college degree
schooling	Dummies for the following categories: 1) no schooling degree, 2) Hauptschulabschluss or Mittlere Reife /Fachoberschule (degrees reached after completion of the 9th or 10th grade), 3) Fachhochschulreife or Abitur/Hochschulreife (degrees reached after completion of the 12th or 13th grade)
health	Dummies for the following categories: 1) no health problems mentioned, 2) health problems but considered without impact on placement, 3) health problems considered to have an impact on placement
pasthealth	Same categories as health, but referring to the past two years before the beginning of the unemployment spell
disabled	1 if disabled, 0 otherwise
land	Ten categories for the West-German Bundesländer: 1) Schleswig-Holstein, 2) Hamburg, 3) Niedersachsen, 4) Bremen, 5) Nordrhein-Westfalen, 6) Hessen, 7) Rheinland-Pfalz, 8) Baden-Württemberg, 9) Bayern, 10) Saarland

<continued on next page>

Table 5: Variable Definitions <continued>

Name	Definition
area	Alternative regional classification aggregating the German Bundesländer into 4 categories: 1) Schleswig Holstein, Niedersachsen, Bremen, Hamburg; 2) Nordrhein-Westfalen, 3) Hessen, Rheinland-Pfalz, Saarland; 4) Baden-Württemberg, Bayern
region	Classification provided by the Federal Employment Office of the district of residence into five groups according to the local labor market conditions: 1) regions with very bad labor market conditions, 2) metropolitan areas with high unemployment, 3) regions with medium unemployment, 4) metropolitan areas with very good labor market conditions, 5) non-metropolitan areas with very good labor market conditions
family	Dummies for the following categories: 1) missing, 2) living alone, 3) not married but living together with at least one person, 4) single parent, 5) married
married	Dummies for the following categories: 1) missing, 2) married, 3) not married
child	1 if at least one child, 0 otherwise
youngchild	1 if at least one child younger than 10 years, 0 otherwise
occupation	Dummies for the following categories: 1) missing, 2) elementary occupations, 3) skilled agriculture and fishery workers, 4) craft, machine operators, and related, 5) service workers, 6) clerks, 7) technicians and associate professionals, 8) professionals and managers
industry	Dummies for the following categories: 1) agriculture, forestry, fishing, 2) manufacturing, 3) construction, 4) trade, transportation, restaurants, 5) finance, insurance, accomodation, professional services, 6) public and private Services

<continued on next page>

Table 5: Variable Definitions <continued>

Name	Definition
occhange	Dummies for the following categories: 1) missing, 2) if the person wishes to work in the same occupation as in the last employment, 3) otherwise
parttime	1 if last employment was part-time, 0 otherwise
whitecollar	Dummies for the following categories: 1) missing, 2) if the previous employment was a white-collar job, 3) if it was a blue-collar job
problemgroup	1 if participation in a program providing psychosocial support within the last three years, 0 otherwise
onlyparttime	1 if information available that only part-time job is desired, 0 otherwise
endlastjob	Dummies for the following categories: 1) missing and other reasons, 2) layoff, 3) voluntary quit, 4) fixed term contract
quarter	Quarter of the end of the last employment: 1) 1st quarter 2000, 2) 2nd quarter 2000, 3) 3rd quarter 2000, 4) 4th quarter 2000, 5) 1st quarter 2001, 6) 2nd quarter 2001, 7) 3rd quarter 2001, 8) 4th quarter 2001, 9) 1st quarter 2001
penalty	1 if the unemployed was suspended from benefits within the last three years, 0 otherwise
motivationlack	1 if within the last three years there is information, that the person did not appear regularly at the labor office, on lack of cooperation, availability or similar, 0 otherwise
pasttreatcancel	1 if in the past the person dropped out of a program, 0 otherwise
pastincapacity	1 if incapacity of work due to illness, parental leave, cure or therapy within the last three years, 0 otherwise

<continued on next page>

Table 5: Variable Definitions <continued>

Name	Definition
proposals	Number of placement proposals divided by the days since the beginning of the unemployment spell and the start date of the spell from which the information is taken
dapp	1 if employed as apprentice within the last three years before the beginning of the unemployment spell, 0 otherwise
countemp, countub, countua, countsp, countoos, countcontact	number of days within the last three years before the beginning of unemployment spent in regular employment, receiving unemployment benefits, unemployment assistance, subsistance payments, out of sample, in contact with the labor office, respectively
demp6, demp12, demp24, demp6_12, demp12_24	1 if in regular employment 6, 12, 24, 6 and 12, and 12 and 24 months, respectively, before the beginning of the unemployment spell
waged	daily wage in the last job(s) before the beginning of the unemployment spell
ddssec, ddcens, ddmarg	Dummies if waged is censored: ddsec is 1 if earnings are within the social security thresholds, ddcens is 1 if earnings are above the social security threshold, ddmarg is 1 if earnings are below the social security threshold (the social security threshold is the threshold below/above which earnings are not subject to social security contributions anymore)
lnwage, lnwagedsq	log(waged) and log(waged) squared interacted with ddssec
wage	Total earnings in the last year before the beginning of the unemployment spell
dssec, dcens, dmarg	Censoring dummies referring to wage (see above)
lnwage, lnwagesq	log(wage) and log(wage) squared interacted with dssec

<continued on next page>

Table 5: Variable Definitions <continued>

Name	Definition
ur_yb, ur_qb, ur_qb3, ur_qb6, ur_qb12, ur_qb24	Unemployment rate in the individual's home district in the calendar year before the beginning of unemployment, in the last month of the quarter before the beginning of unemployment, and in the last month of the quarter before the beginning of the stratum, respectively

Note: If not mentioned otherwise, variables are defined relative to the beginning of the time window of elapsed unemployment duration.

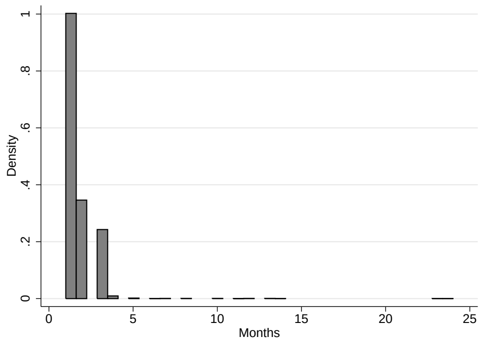
Table 6: Descriptive Statistics for Selected Variables

Variable	Mean	SD	Min	Median	Max
			Males		
age	36.90	7.72	25	36	53
foreigner	.173	.379	0	0	1
schooling1	.124	.329	0	0	1
schooling2	.741	.438	0	1	1
schooling3	.135	.342	0	0	1
qualification1	.357	.479	0	0	1
qualification2	.591	.492	0	1	1
qualification3	.052	.222	0	0	1
countemp	801.05	287.08	86	873	1096
			Females		
age	37.94	7.89	25	37	53
foreigner	.108	.310	0	0	1
schooling1	.072	.259	0	0	1
schooling2	.732	.443	0	1	1
schooling3	.195	.396	0	0	1
qualification1	.321	.467	0	0	1
qualification2	.606	.489	0	1	1
qualification3	.073	.259	0	0	1
countemp	776.92	312.86	86	845	1096

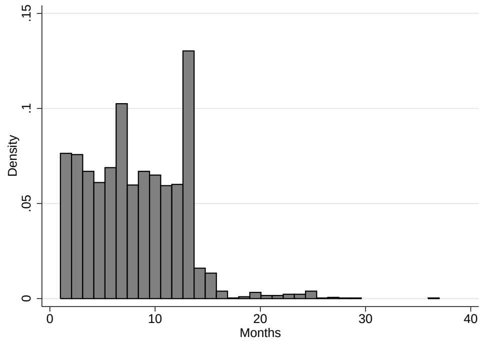
Table 7: Descriptive Statistics on Program Durations in the Analysis Sample

	N	Min	Mean	Median	75th Perc.	95th Perc.	Max
STT	6158	1	1.6	1	2	3	24
CFT	2893	1	8.2	8	12	14	37
PFT	740	1	6.5	7	8	12	26
Total	10857	1	5.8	2	8	24	53

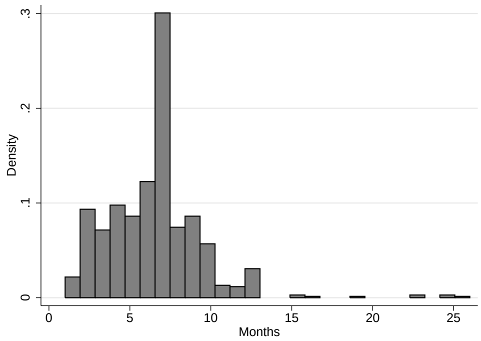
Figure 21: Densities of Program Durations in the Analysis Sample



Short-Term Training



Classroom Further Training



Practical Further Training

7.2 Regression for Effect Heterogeneity

7.2.1 Test for Effect Heterogeneity

The following table reports the Wald-tests for effect heterogeneity. Concretely, we test the joint significance of the coefficient estimates of the ex post outcome regression in equation (4) for the interaction effects involving the treatment dummy variable D_i, i.e. we test the hypothesis $H_0 : \boldsymbol{\delta} = 0$ based on the bootstrap estimates for the variance–covariance matrix.

Table 8: Wald-Tests for Effect Heterogeneity

	χ^2-Test-Statistic (p-Value)		
	Stratum 1	Stratum 2	Stratum 3
STT, Males	14.95 (0.780)	22.02 (0.339)	19.75 (0.473)
STT, Females	23.81 (0.251)	16.02 (0.715)	20.38 (0.434)
CFT, Males	16.22 (0.703)	26.43 (0.152)	15.51 (0.747)
CFT, Females	8.10 (0.991)	21.96 (0.343)	15.64 (0.739)
PFT, Males	18.28 (0.559)		
PFT, Females	35.72 (0.017)**		

Note: *** = statistically significant at 1 %, ** = at 5 %, * = at 10 %, bootstrap standard errors.

7.2.2 Mismatch Corrected ATT and Uncorrected ATT

The following table reports the estimated mismatch corrected ATT (γ, see equation (4)) and the corresponding uncorrected ATT after matching for comparison. This allows us to assess the quality of alignment of treated and control individuals with respect to the covariates included in the ex post outcome regression.

Table 9: Mismatch Corrected ATT and Uncorrected ATT

	STT, Males		
	Stratum 1	Stratum 2	Stratum 3
Corr. ATT with CI	0.005 [-.021, .030]	0.025 [-.009, .060]	0.037 [.007, .067]
Uncorr. ATT	0.004 (0.013)	0.026 (0.017)	0.037 (0.015)
	STT, Females		
	Stratum 1	Stratum 2	Stratum 3
Corr. ATT with CI	0.021 [-.010, .052]	0.073 [.033, .112]	0.089 [.047, .131]
Uncorr. ATT	0.023 (0.016)	0.074 (0.020)***	0.088 (0.021)***
	CFT, Males		
	Stratum 1	Stratum 2	Stratum 3
Corr. ATT with CI	-0.109 [-.148, -.070]	-0.054 [-.100, -.007]	-0.006 [-.050, .038]
Uncorr. ATT	-0.104 (0.019)***	-0.050 (0.025)**	0.002 (0.022)
	CFT, Females		
	Stratum 1	Stratum 2	Stratum 3
Corr. ATT with CI	-0.038 [-.086, .010]	-0.036 [-.102, .031]	-0.016 [-.075, .042]
Uncorr. ATT	-0.036 (0.024)	-0.035 (0.033)	-0.010 (0.029)
	PFT, Males		
			Stratum 1
Corr. ATT with CI			-0.024 [-.083, .034]
Uncorr. ATT			-0.013 (0.029))
	PFT, Females		
			Stratum 1
Corr. ATT with CI			0.021 [-.037, .079]
Uncorr. ATT			0.030 (0.029)

Note: Outcome variable is the ATT averaged over the first 24 month since the beginning of treatment. For the corrected ATT estimates we report as CI the 95% confidence interval.

7.3 Future Participation

The following table shows the share of those in the searching group who will be treated at some future point in time (i.e. within the next 35 months and in the same unemployment spell).

Table 10: Future Participation

	Males	Females
	Stratum 1 (0-3 Months)	
No future treatment	80.19	80.48
Future treatment	19.81	19.52
	Stratum 2 (4-6 Months)	
No future treatment	77.34	79.73
Future treatment	22.66	20.27
	Stratum 3 (7-12 Months)	
No future treatment	78.99	85.20
Future treatment	21.01	14.80