



IZA DP No. 3767

The Relative Effectiveness of Selected Active Labour Market Programmes and the Common Support Problem

Gesine Stephan
André Pahnke

October 2008

The Relative Effectiveness of Selected Active Labour Market Programmes and the Common Support Problem

Gesine Stephan

IAB and IZA

André Pahnke

IAB

Discussion Paper No. 3767
October 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

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

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

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

ABSTRACT

The Relative Effectiveness of Selected Active Labour Market Programmes and the Common Support Problem^{*}

For Germany, we analyse the (relative) effects of participation in several active labour market programmes on the employment prospects of participants. First, our results show that different matching algorithms result in different severe problems of common support. Second, we obtain favourable effects of participation in training programmes, which is not true for job creation schemes. Third, while lock-in effects are smaller for shorter programmes, long retraining shows mainly positive effects compared to shorter training at the end of the observation period. Fourth, participants in job creation schemes are too different from participants in training programmes to conduct a reliable comparison.

JEL Classification: J68, J64, J65

Keywords: evaluation of active labour market programmes, propensity score matching, common support problem

Corresponding author:

Gesine Stephan
Institute for Employment Research (IAB)
Regensburger Str. 104
90478 Nuremberg
Germany
E-mail: gesine.stephan@iab.de

^{*} We are grateful to the TrEffeR-team of the German Public Employment Service – in particular Michael Irskens, Susanne Rässler and Torben Schewe – and to Knut Gerlach, Katrin Hohmeyer and Hans Kiesel for support and helpful hints. Note that a previous discussion paper version (IAB Discussion Paper 29/2008) presented separate results for West and East Germany.

I INTRODUCTION

During the last years, the evaluation of active labour market programmes has become a central research topic in many countries: Policy and administration have increasingly been interested in topics such as programme effectiveness and efficiency. At the same time, the development of comprehensive merged data sets – covering times of unemployment, programme participation and employment – laid the groundwork for further research. Although methodological advancements (Imbens, 1999; Lechner, 2001) have extended the often used framework for the estimation of causal treatment effects (Rubin, 1974; Heckman et al., 1999) to pairwise programme comparisons, most evaluation studies have analysed the effects of being or not being in a particular programme.

For Germany, our study analyses the relative effectiveness of participation in different variants of further vocational training, short-training programmes and job creation schemes. Average programme effects on participants are estimated by comparing the group of those joining a particular programme during March 2003 with groups of similar persons that did enter either no programme or one of the other programmes during this month. Outcome variables are cumulated days spent in regular employment during the 3.5 years after programme start as well as the share in regular employment at the end of the observation period. To choose adequate comparison groups we apply propensity score matching; the data used are provided by the TrEffeR-database of the German Public Employment Service.

Our paper complements the existing literature mainly by two aspects: First, the knowledge on the relative effectiveness of participation in different programmes is still sparse for Germany. The results of the few studies available differ partly, we analyse more recent programme entries, and we distinguish between two very different variants of short training programmes. Second, we pay particular attention to the problem of common support that arises when estimat-

ing relative programme effectiveness (Lechner 2000) and conclude that not all programme variants might be compared.

Section II provides a brief survey on the programmes investigated, while Section III sketches previous research results. Section IV depicts the evaluation approach. Section V discusses the applied method, paying particular attention to the common support problem. Section VI then informs about data and variables used in the empirical analysis. Our empirical findings are described in section VII. We draw some conclusions in Section VIII.

II PROGRAMME FEATURES

Unemployment in Germany had been rising for many years. As a consequence, major labour market reforms were enacted from 2003 to 2005. New instruments of active labour market policy were implemented, existing programmes were modified, the German Public Employment Service was reorganised, and former unemployment assistance and social assistance were consolidated into a new means-tested basic social care for needy unemployed job-seekers. Unemployment reached its maximum with on average 4.9 million unemployed persons in 2005. Since 2005 unemployment has declined substantially, to (seasonally adjusted) 3.2 million registered unemployed persons in August 2008. Active labour market programmes have the main objective to improve the employment prospects of participants and to avoid or shorten periods of unemployment. Table 1 shows entries into the most important German labour market programmes from 2000 to 2006 (see Bernhard et al., 2008, for details).

Our analysis is restricted to three of the largest programmes: We analyse variants of further vocational training, short training programmes and job creation schemes. Programme features as well as evaluation results will be described in the following. Further important programmes are in particular wage subsidies, start-up subsidies and contracting-out to private agencies.

Table 1 **Entries in important German labour market programmes during 2000-2006 (in 1000)**

Programme	2000	2001	2002	2003	2004	2005	2006
Further vocational training (<i>Förderung beruflicher Weiterbildung</i>)	523	442	455	255	185	132	247
Short training (<i>Trainingsmaßnahmen</i>)	485	551	865	1064	1188	894	978
Job creation scheme I (<i>Arbeitsbeschaffungsmaßnahmen, SAM</i>)	318	246	215	179	161	80	80
Job creation scheme II (<i>Arbeitsgelegenheiten</i>)	-	-	-	-	-	630	742
Wage subsidy (<i>Eingliederungszuschuss</i>)	152	127	188	183	157	134	217
Start-up subsidy I (<i>Überbrückungsgeld</i>)	93	96	125	159	183	157	108
Start-up subsidy II (<i>Existenzgründungszuschuss</i>)	-	-	-	95	168	91	43
Start-up subsidy III (<i>Gründungszuschuss</i>)	-	-	-	-	-	-	34
Contracting-out to private agencies (<i>Beauftragung privater Dritter</i>)*	-	-	-	-	635	426	301

*) Numbers are available since 2004, while different variants started already in 1998 (contracting-out subtasks of placement) and 2002 (contracting-out all placement services).

Source: Statistics Department of the German Public Employment Service (Data-Warehouse).

For a long time further vocational training (*Förderung beruflicher Weiterbildung, FbW*) has been one of the most important German labour market programmes. During the first half of this decade, the number of entries decreased (as well as the average programme duration); but the number of entries increased again in 2006. Further vocational training updates and extends professional qualifications. It encompasses a range of different treatments, which can be broadly classified in qualification programmes, training within “practice firms” (which offer practical occupational training, but are no “real” companies) and long retraining programmes. The latter might be granted to persons without completed vocational training or to unemployed who did not practice a corresponding job for the last four years. Direct training costs for further vocational training programmes are paid by the Public Employment Service, and participants receive a subsistence allowance that usually equals the unemployment compensation. Since 2003 access to further training programmes is granted through training vouchers that specify training target, programme duration, regional scope and period of validity. A caseworker issues a training voucher to a potential participant, who has then to find a training provider offering an adequate course, who is also willing to sign in the holder of the voucher. Kruppe (2008) showed that hard-to-place unemployed a) receive a voucher less often and b) make use of the voucher less often.

Since 2001, short training programmes (*Trainingsmaßnahmen, TM*) are the programme with the highest number of programme entries. Programme duration is, however, short and varies from two to eight weeks. Short training programmes have a number of different objectives: They could improve knowledge and skills, test the occupational aptitude of the employee, check whether unemployed are suited for further longer-term measures, support job-search by job application training, or verify an employee's availability and willingness to work. Training measures are conducted by providers (classroom) or placement in a firm (firm-internal). Similar to further vocational training programmes, direct programme costs are paid by the Public Employment Service, while participants continue to receive unemployment compensation. Participation in a short firm-external training programme will be often suggested by the caseworker and then conducted by providers (Kurtz 2003). Short firm-internal training requires additionally that a firm is willing to offer the training opportunity to a potential participant. Employers might use the training to test the productivity of the unemployed person without incurring any wage costs; this raises also the danger of deadweight losses if an employer would have hired the unemployed in question anyway.

Job creation programmes had been an important instrument at the beginning of this decade, but then lost importance until 2004. Our analysis will focus on the most important traditional variant of job creation schemes (*Arbeitsbeschaffungsmaßnahmen, ABM*) that are mostly conducted in the public and non-commercial sector. They should provide the last chance to stabilise and qualify unemployed persons for later re-integration into regular employment. The tasks carried out during participation have to be of “additional” nature (they would not be executed without the subsidy) and of public interest. Until 2004, grants paid for job creation schemes were based upon an “allowable” remuneration and covered part of the costs. Currently, a lump sum payment is granted, the amount of which varies with the qualification required. The regular duration of participation in a job creation schemes is limited to 12 months, while exemptions are possible. A further variant (*Strukturanpassungsmaßnahmen, SAM*) had been in place from 2003

to 2004, with the purpose to maintain or improve regional infrastructure and environment whereas providers received a monthly lump sum payment for participants. The regular programme duration was 36 months, but could even be prolonged. In 2005, a new variant of job creation schemes was introduced for the new group of needy job seekers receiving basic social care (*Arbeitsgelegenheiten* or *Ein-Euro-Jobs*), which now plays a major role. This programme provides only a modest additional reimbursement for work.

III LITERATURE REVIEW

A considerable number of papers investigated the effectiveness of further vocational training programmes in Germany, comparing participants with comparable non-participants. Lechner et al. (2005; 2007), Fitzenberger et al. (2006) and Fitzenberger and Völter (2007) analysed the long-run effects up to seven years after programme entry, focusing mainly on unemployment entries during 1993/94. In the long run, they generally found positive effects of further vocational training. However, because programme effects are rather weak, it may take some time until the estimated effect turns positive. For more recent programme entries during the years 2000 to 2002 the evidence is mixed: Wunsch and Lechner (2008) restricted their sample to West Germany and individuals aged 25 and 49. Their results showed that further vocational training – and other programmes – had mostly negative or insignificant effects on employment rates of participants 30 months after programme start. Biewen et al. (2007) found positive effects on employment rates of participants of age 25 to 53 for programmes of short and medium duration in West Germany (but not in East Germany) and particular groups of unemployed. Rinne et al. (2007) obtained – two years after programme entry – positive effects of participation in medium length programmes on the employment probabilities of participants in all subgroups investigated, covering participants of age 17 to 65. Hujer et al. (2006a) applied duration analysis to East German data from the years 1999 to 2002. Their main result is that participa-

tion in further vocational training prolonged unemployment duration during the period investigated.

For short training programmes, Biewen et al. (2007) found mostly positive effects of short training-programs, while Wunsch and Lechner (2008) – who separately analysed short combined measures, jobseeker assessment and short training for minor adjustment of skills – did not. The duration analysis of Hujer et al. (2006b) showed that the risk of entering employment is significantly higher for individuals participating in a short-training programme. Wolff and Jozwiak (2007) distinguished between short classroom training and short training within firms for individuals and investigated the effect on the employment prospects of needy job-seekers, who receive the new basic social care. They showed that both variants had positive effects, which are much larger for short training within firms. Büttner (2008) used data from a social experiment on short-training programmes to test the availability of the unemployed. His main result was that it is the notification of treatment rather than participation that has an effect on leaving unemployment.

Entries into job creation schemes – with an average duration of 9 to 11 months – have been investigated by Caliendo et al. (2006; 2008a; 2008b) and Hujer and Thomsen (2006). The authors applied statistical matching methods; they imposed no age restrictions, but estimated heterogeneous effects for groups of participants. Three years after programme entry (in February 2000) these effects turned out to be mostly negative or insignificant. Exceptions were long-term unemployed, highly qualified men and older women in West Germany. Wunsch and Lechner (2008) also obtained negative effects of participation in this programme on the employment prospects of participants. Hujer and Zeiss (2006) showed for East Germany that participation in job creation schemes increases individual unemployment duration of participants, using the timing-of-events method. Recent results on the new “One-Euro-Jobs” (Hohmeyer and Wolff 2007), introduced in 2005, highlighted the effect heterogeneity of this programme. The authors

found slightly positive effects in particular for participants from West Germany and individuals out of regular employment for a longer time period.

What are the results of cross-programme comparisons for Germany? First evidence was presented by Lechner et al. (2005), who compared participants in practice firms, short and long programmes providing professional skills as well as retraining, for programmes starting during 1994/95. In the long run, seven years after programme start, they obtained few significant differences between programme effects (Table 6.2, 46). Using more recent data, Biewen et al. (2007) conducted pairwise evaluations of several further vocational training programmes and short training programmes. They observed labour market outcomes of participants for a period of 2 to 2.5 years after treatment start. A main result is that participants in short training programmes would not have improved their latter employment rates by attending longer classroom or practical training. Longer classroom training showed no advantage for the treated, compared to the other programme variants. However, practical further training was often more effective for participants than another training variant would have been. Wunsch and Lechner (2008) compared an even wider range of programmes, distinguishing three kinds of short training programmes, four variants of further vocational training, and public employment schemes. Their pairwise comparisons (Table 7, 169) showed at $\alpha = 0.05$ nearly no significant effect of participation in one programme – compared to participation in another programme – on the employment rates 2.5 years after treatment start. Exceptions were long retraining programmes that had a significant negative effect on employment rates of participants compared to most other programmes. Kluge et al. (2007) used a different technique to compare the effects of training of different lengths, restricted to programmes that were not leading to the acquisition of a degree: Estimating a dose-response function and adjusting for covariate-imbalance, they obtained an increasing dose-response for treatments up to 100 days and concluded that longer programs do not add an additional treatment effect. Their outcome variables were the employment probability two years after programme entry and one year after programme exit.

Finally, we sketch the results of several comparative programme studies for other countries: Gerfin and Lechner (2002) evaluated nine different Swiss active labour market programmes, focussing on the first programme participation of an unemployed. Their study showed that temporary wage subsidies – paid in order to temporarily compensate income-losses in comparison to former times of employment (*Zwischenverdienste*) – has been most efficient in integrating participants into regular employment. Gerfin et al. (2004) contrasted employment programmes in non-profit organizations and temporary wage subsidies; again the latter were the more “successful” programme. Sianesi (2008) compared six major Swedish active labour market programmes. Employment subsidies performed best by far; they were followed by trainee replacement and labour market training. For Great Britain, Dorsett (2001) contrasted entries into subsidised employment, full-time education and training, an environmental task force or a voluntary sector, which were different options within the New Deal Programme for Young People. Again, wage subsidies dominated all other options.

Sianesi (2008) summarises as a main result of many micro studies that the more a programme resembles regular employment in the competitive sector, the higher the programme’s benefits to its participants will be. However, the underlying selection process for participation in wage and start-up subsidy programmes – which are most similar to regular work in the private sector – differs to a larger extent from the programmes analysed in our paper: In the case of wage subsidies an employer must be willing to offer at least a subsidised job to an unemployed person, while founding a subsidised new enterprise requires that the founder has the initiative to do so.

IV EVALUATION APPROACH

As most micro studies of active labour market programmes, our evaluation approach is based on the model of potential outcomes. We compare labour market outcomes of – comparable – participants receiving different “treatments”. In the basic version of the potential outcomes model (Rubin, 1974) an individual can potentially be in two states, while the outcome variable

of interest may differ between these states. Based on Rubin’s work, Imbens (1999) and Lechner (2001) provided an extension to the case of multiple states. Denote participation in one of N treatments starting at time t by $S^t \in \{0,1,\dots,N\}$, and let the corresponding potential outcomes at time $t+h$ be given by $\{Y_0^{t+h}, Y_1^{t+h}, \dots, Y_N^{t+h}\}$. Typically also non-participation at time t is interpreted as a particular kind of treatment. Because an individual may enter only one programme at time t , only one element of the latter set is observable, all other outcomes are unobserved “counterfactuals”. In our case the outcome variables under consideration will be the employment rate and cumulated days spent in regular employment after programme entry. Furthermore, t will be March 2003 and h will be 3.5 years.

We assume that the value of the outcome variables for each person is not influenced by the actual participation of other persons (“Stable Unit Treatment Value Assumption” SUTVA). Then the average effect of treatment J on participants in this programme, compared to receiving treatment K instead, is given by

$$(1) \quad \theta_{JK}^{t+h} = E[Y_J^{t+h} - Y_K^{t+h} | S^t = J] = E[Y_J^{t+h} | S^t = J] - E[Y_K^{t+h} | S^t = J].$$

In the following, we will denote participants in programme J as the “treatment group” and participants in programme K as the “comparison group”. It is not possible to observe the average counterfactual outcome that members of the treatment group J would have had, if they had participated in programme K instead $[Y_K^{t+h} | S^t = J]$. Thus one has to find an adequate comparison group to impute the counterfactual outcome (Rubin 1974).

With non-experimental data, statistical matching techniques might be applied to find such a comparison group – but only for those individuals in the treatment group J who have a positive probability to be in programme K instead (“Common Support Condition”). Statistical matching relies on “matching on observables”: Assume that all variables X , determining the participation decision as well as the expected success of a programme, are known and available. Then a comparison group of individuals receiving treatment K , with similar observable characteristics

X to the treatment group J , may be chosen to obtain an estimate for the counterfactual outcome $[Y_K^{t+h} | S^t = J]$. Similar to the binary case (Rubin, 1974), Imbens (1999) and Lechner (1999; 2001) showed that the “Conditional Independence Assumption” (CIA) – formally given by $Y_0^{t+h}, Y_1^{t+h}, \dots, Y_N^{t+h} \perp\!\!\!\perp S^t | X$ – identifies the parameters of interest in the case of multiple treatments.

Thus the identifying assumption of statistical matching techniques is that no unobserved heterogeneity correlated with the selection into programmes and with outcome variables remains after accounting for observable variables. In contrast, duration analysis (in particular the timing-of-events approach by Abbring and van den Berg, 2003 and 2004) allows for selection on unobserved characteristics. But these models impose the identifying assumption that transition processes into labour market programmes as well as across labour market states can be modelled as a multivariate mixed proportional hazard model, while statistical matching is a non-parametric approach.

Sianesi (2004, 2008) and Fredriksson and Johansson (2004) have pointed out that labour market programmes in Europe are ongoing and individuals may take up a programme sooner or later provided they are still eligible. But the unemployed themselves or the caseworker may decide against participation, because they expect or are expected to find regular employment soon. Thus selecting a comparison group of individuals who never participated in any programme would lead to base selection on expected (successful) future outcomes. Steiger (2004) and Stephan (2008) demonstrate empirically, how evaluation results vary with the choice of the classification window. We do not put any restrictions on the future of persons and define non-participation in a particular programme as not taking up this programme during March 2003, but maybe at a later date. Following Sianesi (2004; 2008), we will denote this group as “waiting”. Similarly, also participants in the programmes investigated might take part in another programme later.

V APPLIED METHOD AND COMMON SUPPORT PROBLEM

Rosenbaum and Rubin (1983) have shown that instead of matching on a high-dimensional vector of X -variables it is sufficient to match on the propensity score – the probability to join a programme – to obtain the same probability distribution for treated and non-treated individuals. Similar properties hold in a multiple treatment framework as well (Lechner, 2001). In consequence, the same methods as in a binary treatment framework can be applied. First, propensity scores could be estimated separately for each combination of programmes J and K , using a binary probit or logit model. Second, the complete choice problem can be formulated in one model and estimated with a multinomial probit model, for instance. Lechner (2002) obtained basically the same estimation results, irrespective whether conditional probabilities were derived from a multinomial model or estimated directly. Our analysis is based on the first approach, estimating 81 binary probit models, since we compare participation in nine programme variants with participation in the other eight programme variants as well as with non-participation. Note that estimated effects θ_{JK}^{t+h} and θ_{KJ}^{t+h} are not necessarily symmetric; estimated effect refers to individuals within common support in the treatment group.

While propensity score matching is a useful simplification and asymptotically all estimators should yield the same results, in small samples the choice of the matching algorithm and the determination of the region of common support can be important (Heckman et al., 1997; Smith, 2000; Caliendo and Kopeinig, 2008). This problem is less severe when estimating the effects of program participation compared to non-participation, since the number of potential comparison persons is usually large. However, the number of potential comparison persons can be relatively small in cross-programme comparisons. In consequence, only a low share of the treatment group might be in common support. Furthermore, particular comparison observations might have a strong impact on results if a non-participant may be used as a comparison person more than once in the matching procedure (matching with replacement).

In the following we compare results for two standard propensity score matching algorithms. First, we conduct a nearest neighbour matching without replacement that chooses for each participant the non-participant with the most similar propensity score as a comparison person. This is rather for demonstration purposes, because we will partly end up with very few observations. Second, we perform radius matching (Dehejia and Wahba, 2002) that matches participants with “synthetic comparison persons”, composed of a weighted equivalent of all persons falling within the radius of their propensity score. As more data points are used, radius matching will result in lower variances compared to nearest neighbour matching. However, the bias of the estimates will be higher, because also more different comparison observations are used.

We restrict our analysis on the region of common support and furthermore set a calliper – a maximum distance of propensity scores between treated and comparison persons – of 0.005 for cross-programme comparisons and of 0.0005 for comparisons with non-participants. The programme impact is then estimated as the mean difference in the weighted outcomes of both groups. All estimates are performed using the STATA-module *psmatch2* (Leuven and Sianesi, 2003). Note that variance estimates for estimated treatment effects neglect that the propensity score itself has been estimated (Abadie and Imbens, 2006).

To analyse common support, one possibility would be to present distributions of propensity scores or of relative weights from matching (Black and Smith, 2004); but each would require 162 graphs. Instead, we compute Lechner bounds (Lechner 2000; Caliendo and Kopeinig, 2008) to test for the sensitivity of estimated treatment effects with respect to the common support problem. The lower (upper) Lechner bound is given by the weighted average of a) the estimated average treatment effect and b) the average distance of observations for treated persons throughout common support from the upper (lower) bounded potential outcome. Weights are given by probabilities a) to be or b) not to be within common support. We test if estimated effects will be still significant after computing confidence intervals around these bounds. In the Appendix we will present the underlying shares of individuals in the treatment group that are

within common support. Additionally, we display the maximum weight across comparison persons as a share of treated persons to see how heavily observations are used in constructing the counterfactuals. Note that the weight of a comparison person is always one for nearest neighbour matching without replacement, thus the share of treated persons is $1/N$ (with N as the number of treated persons), but the share maybe smaller or larger using radius matching.

Finally, to test the quality of matching, the mean standardised bias (MSB) (Rosenbaum and Rubin, 1983) between each treated group and its matched comparison group is computed across all variables of X . The standardised bias of a covariate is defined as the difference of means in the treated and matched comparison sample, divided by the square root of the average sample variance. Thus, a lower value of the MSB indicates more similarity between the two groups. While no clear theoretical indication exists as to which remaining bias might be acceptable, Caliendo and Hujer (2006) summarise as a rule of thumb that most studies assess a reduction of the MSB after matching to 3 or 5 percent as sufficient. We follow this suggestion and interpret pairwise comparisons where the remaining MSB after matching exceeded the value of 5 percent as programme types that are in fact not comparable in terms of their participants.

VI DATA AND VARIABLES

The empirical analysis is based on the TrEffeR-data set (Stephan et al., 2006). This administrative data set has been constructed for monitoring purposes of the German Public Employment Service. The current version merges data flows from computer based operative systems of the Public Employment Service on periods of registered job search, registered unemployment, participation in labour market programmes and employment for the period from 2000 to 2007.

The sample analysed here covers individuals of age 25 to 59, who were unemployed for no longer than one year in March 2003. Individuals entering one of these programme variants during March 2003 are included in the cross-programme comparison, while those not entering a

programme are interpreted to be in a state of “waiting”. As is done most often in the literature, only the first programme entry during an unemployment spell is analysed; the distance to the previous unemployment spell has to amount to at least one month.

We restrict our analysis to the programme variants described in Table 2. We also distinguish between different completed programme durations (up to three months, 4 to 6 months, 7 to 12 months and more than 12 months). Our data do not include information on planned programme duration. Presumably, participants exit a programme prior to its planned completion if they find a job during participation or if they do not expect to find a job even with the help of the programme. Hence, completed duration might be correlated with the outcome of treatment as well as with individual characteristics. However, Kluve et al. (2007) found that instrumental variable estimates using planned duration as instruments were not significantly different from estimates utilizing information on completed duration. They conclude that estimates relying on actual training duration do not suffer strongly from endogeneity.

Table 2 **Analysed programme variants**

1. Further vocational training (*Förderung beruflicher Weiterbildung*)

- a) Provision of specific professional skills, which might contain occupation-related training and general training (*berufsbezogene übergreifende Weiterbildung, berufspraktische Weiterbildung oder berufliche Aufstiegsweiterbildung*).
 - b) Practical training in a practice firm, without trainees working in a “real” company (*berufliche Weiterbildung in einer Übungsfirma, Übungswerkstatt oder sonstigen Übungseinrichtung*)
 - c) Long retraining programmes, conducted firm-external within a group (*Gruppenmaßnahme mit Abschluss in anerkanntem Ausbildungsberuf*)
-

2. Short training programme (*Trainingsmaßnahme*)

- a) Short classroom training, aimed at the improvement of knowledge and skills (*Nicht-betriebliche Trainingsmaßnahme zur Vermittlung von Kenntnissen*)
 - b) Short firm-internal training, aimed at the improvement of knowledge and skills (*Betriebliche Trainingsmaßnahme zur Vermittlung von Kenntnissen*)
-

3. Job creation scheme of the traditional type I from Table 1 (*Arbeitsbeschaffungsmaßnahme*)

We estimate programme effects on participants using two outcome variables: First, we compute cumulated days spent in regular, unsubsidised employment during the 3.5 years after programme entry in March 2003. Second, we compute shares in regular, unsubsidised employment 3.5 years after programme entry. Times of subsidised employment (for instance by a wage subsidy) and of “marginal” employment are not interpreted as a “success” in this sense. Outcome variables for “waiting” non-participants are measured since March 15, 2003, and all individuals who had already left unemployment at this date were excluded from the sample.

Cumulated effects display the evolvement of estimated programme effects over the entire observation period of time; they can be computed as the integral over employment shares during each day of the observation period: Thus they account for locking-in effects – times of reduced search – over the time period of programme participation. The share in regular employment at the end of the observation period refers only to one particular reporting day. However, it may be interpreted as an indicator that shows how cumulated days in employment will develop further after the end of the observation period: If the average effect on the share of participants in employment is positive, the average effect on cumulated days will turn more positive over time.

The choice of comparison groups is based on a wide range of individual socio-demographic characteristics (measured at the start of an unemployment spell), unemployment duration in the current spell, (un-)employment history in the two years preceding the analysed unemployment spell, and the regional labour market situation (performance cluster according to Blien et al. 2004). Since the data include information on previous unemployment histories, these should capture most of the effects of unobserved individual factors (Heckman et al., 1999). Mean values of the explaining variables can be found in Table A.1 in the Appendix. It shows that participants in a specific programme differ in fact from the average non-participant as well as from participants in other programmes. In particular, participants in all training programmes seem to be a “positive selection” of unemployed persons, whereas those joining a job creation scheme can be considered as a “negative selection”.

VII EMPIRICAL RESULTS

Common support and matching quality

Table 3 summarizes information on the significance and quality of the results: For each combination of programmes, column I contains information on the sign and significance of the estimated effect (++/- indicates that results that are significant at $\alpha = 0.01$, while +/- shows that they are significant at $\alpha = 0.05$). This information is complemented in column II by information on the sign and significance taking into account Lechner bounds. If results in column I and II differ, estimated treatment effects are sensitive with respect to the common support problem. Furthermore, column III of Table 3 indicates if the mean standardised bias after matching is below “acceptable” values of 3 (##) or 5 (#), respectively. For ease of interpretation, comparisons where we find a significant effect using Lechner bounds as well as satisfying matching quality are shaded in light grey. Background information can be found in Table A2 in the Appendix that shows in Panel 1, which share of the treatment group J is within common support. Panel 2 displays the maximum weight given to a comparison person. Finally, Table A3 in the Appendix displays the mean standardised bias (MSB) before and after matching.

Table 3 indicates that – independently from the matching algorithm chosen – we do not face a common support problem if programme entry during March 2003 is compared with “waiting” during this month. However, the picture is different for pairwise programme comparisons: For nearest neighbour matching without replacement, significant point estimates go hand in hand with insignificant effects when Lechner bounds are taken into account. This problem is less severe for radius matching. Table A.2 shows the underlying reason: At least 70 percent of the treatment group are in common support using radius matching, while the minimum share in common support amounts to only 12 percent for nearest neighbour matching without replacement (when comparing participants in short classroom training with those in longer job creation scheme).

Table 3 Sign and significance of point estimates (I), sign and significance taking into account Lechner bounds (II) and matching quality (III)

Treatment group J (type and duration in months)	Comparison group K (type and duration in months)																			
	Waiting	Provision of skills									Practice firm		Re-training		Short training in firm		Job creation Scheme			
		<4			4-6			7-12			4-6		>12		in firm		4-6		7-12	
		I	II	III	I	II	III	I	II	III	I	II	I	II	I	II	I	II	I	II
1 Cumulated days in regular employment																				
<i>1.1 Nearest neighbour matching without replacement</i>																				
Provision of skills, <4	++ ++ ##			##	++ + ##		##	++ ##	-- -- ##	++ ++ ##	++ #	++ ##								
Provision of skills, 4-6	++ ++ ##		##		++ ##		#	++ ##	-- -- ##	++ ++ ##	++ ##	++ #								
Provision of skills, 7-12		##	--	##	--	##		#	++ ##	-- -- ##		##	++ ##	++ ##						
Practice firm, 4-6		#	-	##		##		#	++ #	-- -- ##		#	++	++ #						
Retraining, >12	-- -- ##	-- -- ##	-- -- ##	-- -- ##	-- -- ##	-- -- ##	--	#		-- -- ##	-- -- ##	--	#	#						
Short training in firm	++ ++ ##	++	##	++	##	++	##	++	##	++	##	++	##	++ #						
Short training in class	++ ++ ##	--	##	--	##		##	#	++ ##	--	##		##	++ ##						
Job creation scheme, 4-6	- - ##	--	#	--	#	--	##	--	++ #	-- -- ##	-- -- ##		##	++ ##						
Job creation scheme, 7-12	-- -- ##	--	#	--	##	--	##	--	#	+	#	-- -- #	-- -- ##	##						
<i>1.2 Radius matching</i>																				
Provision of skills, <4	++ ++ ##			##	++ ++ ##	++	##	++ ++ #	-- -- ##	++ ++ ##	++ ++	++ ++								
Provision of skills, 4-6	++ ++ ##		##		++ ++ ##	+	##	++ ++ ##	-- -- ##	++ ++ ##	++ +	++ ++								
Provision of skills, 7-12	+ + ##	-- -- ##	-- -- ##				#	++ ++ ##	-- -- ##		##									
Practice firm, 4-6	+ + ##	-	#	- - ##		##		++ ++ #	-- -- ##		##		++							
Retraining, >12	-- -- ##	-- -- ##	-- -- ##	-- -- ##	-- -- ##	-- -- #			-- -- ##	-- -- ##	-- --	--	--							
Short training in firm	++ ++ ##	++ ++ ##	++ ++ ##	++ ++ ##	++ ++ ##	++ ++ #	++ ++ #		++ ++ #	++ ++ #	++ ++	++ ++	++ ++							
Short training in class	++ ++ ##	-- -- ##	-- -- ##		##	#	++ ++ ##		-- -- ##				++ +							
Job creation scheme, 4-6	-- -- ##	-- --	-- --	-- -- #			+		-- -- #	-- -- ##		##	++ ##							
Job creation scheme, 7-12	-- -- ##	-- --	-- --	-- -- #	--				-- -- #	-- -- ##	-- -- ##		##							
2. Share in regular employment																				
<i>2.1 Nearest neighbour matching without replacement</i>																				
Provision of skills, <4	++ ++ ##			##	##	##	##	--	##	-- -- ##	+ + ##	++ #	+ ##							
Provision of skills, 4-6	++ ++ ##	++	##		##	#	##		##	##	++ ++ ##	++ ##	++ #							
Provision of skills, 7-12	++ ++ ##		##	##		#	-	##	--	##	++ ++ ##	++ ##	++ ##							
Practice firm, 4-6	++ ++ #		##	##	#		#		##		#	++	+ #							
Retraining, >12	++ ++ ##	++	##	##	++	##	#		##	##	++ ++ ##	++ #	++ #							
Short training in firm	++ ++ ##	++	##	##	++	##	##	##	##		++ ++ ##	++ #	++ #							
Short training in class	++ ++ ##	--	##	--	##	-	#	--	##	--	##	+	##	+ ##						
Job creation scheme, 4-6		##	-	#	--	#	--	##	--	##	--	##		##						
Job creation scheme, 7-12		##	-	#	--	##	--	##	#	--	##		##							
<i>2.2 Radius matching</i>																				
Provision of skills, <4	++ ++ ##			-- - ##	##	##	#		-- -- ##	++ ++ ##	++									
Provision of skills, 4-6	++ ++ ##	++ + ##			##	##	##		-	##	++ ++ ##	++ ++	++ +							
Provision of skills, 7-12	++ ++ ##		##	##		#	##		-- -- ##	++ ++ ##										
Practice firm, 4-6	++ ++ ##		#	##	##		#		- - ##	++ ++ ##			+							
Retraining, >12	++ ++ ##	++ ++ ##	+	##	++ ++ ##	#			##	++ ++ ##	++ ++	++	++							
Short training in firm	++ ++ ##	++ ++ ##		##	++ ++ ##	#	#		##	++ ++ ##	++ ++	++ ++	++ ++							
Short training in class	++ ++ ##		##	-- -- ##	-- -- ##	- - #	-- -- ##		-- -- ##											
Job creation scheme, 4-6		##		-	-- -- #		-- --		-- -- #	-- -- ##				##						
Job creation scheme, 7-12	+	##	-	-	-- -- #	-	-- --		-- -- #		##		##							

++ (+) indicates $\theta_{JK}^{3.5} > 0$ and $\alpha = 0.01$ (0.05).

-- (-) indicates $\theta_{JK}^{3.5} < 0$ and $\alpha = 0.01$ (0.05).

(#) indicates a mean standardised bias < 3 (5).

Shaded in light grey: Significant taking into account Lechner-bounds and MSB < 5.

Source: Own calculations, based on the TrEffeR data.

These results clearly underline that matching without replacement is not advisable if the number of potential comparison persons is small. The other side of the coin is that the same observation may be used very often as a comparison person if matching takes place with replacement. Panel 2 of Table A2 show that the maximum weight as share of the treatment group is partly smaller, partly larger in the “waiting” group compared to nearest neighbour matching without replacement. It is, however, usually larger for cross-programme comparisons. In particular, maximum weights are large if comparison persons have participated in a job creation schemes. They increase up to 10 percent of the treatment group (provision of skills, 4 to 6 months); thus one particular person participating in a job creation scheme is used as a comparison person for one tenth of treated persons. This shades a first doubt on the reliability of the estimates where individuals taking up training are compared to similar persons joining a job creation scheme.

Matching quality is better for radius matching than for nearest neighbour matching without replacement, if we compare programme participants with “waiting” non-participants (Table A3). For cross-programme comparisons, results differ not strongly between both matching algorithms regarding training programmes. Matching quality – measured by the mean standardised bias after matching – is usually good ($MSB < 5$) or very good ($MSB < 3$). This is differently for comparisons including participation in a job creation scheme, if radius matching is applied. While for nearest neighbour matching without replacement matching quality is mostly acceptable for job creations schemes, this is paid for by dropping the majority of cases from common support (Table A2).

As robustness checks, we applied several further matching algorithms. Results of nearest neighbour matching without replacement were rather invariant to another sorting of observations. Smaller variations of the callipers chosen did not have a qualitative impact on results. However, results for comparisons including job creation schemes differ strongly from those

presented above if we match nearest neighbours without any calliper; this implies that treated persons and comparison persons can be rather different in terms of their propensity scores.

Summing up, for our analysis radius matching should be preferred against nearest neighbour matching without replacement; the common support problem is too severe for the latter one. Thus we will restrict the following discussion on the results from radius matching. Furthermore, participants in job creation schemes should in most cases not be compared with participants in training schemes – we cannot achieve a good matching quality and at the same time a satisfying solution of the common support problem.

Joining a programme compared to “waiting”

We begin with a discussion of the mean effects of a programme entry compared to no or a later participation (“waiting”), restricting the following discussion to the findings from radius matching. Results at the end of the observation period are summarised in Table 4. Figure 1 and 2 show the development of outcome variables for the treated group and the comparison group as well as estimated treatment effects over the entire observation period. As can be seen from Figure 2, employment evolves cyclically during the calendar year.

Table 4: Estimated effect of participation in programme J compared to waiting, 3.5 years after programme entry

Regular employment	Treatment group J (type and duration in months)								
	Provision of skills			Practice firm	Re-training	Short training		Job creation scheme	
	<4	4-6	7-12	4-6	>12	in firm	in class	4-6	7-12
Cumulated days	112**	92**	20*	38*	-146**	206**	27**	-40**	-52**
Share	0.09**	0.14**	0.12**	0.12**	0.20**	0.15**	0.05**	0.01	0.03*

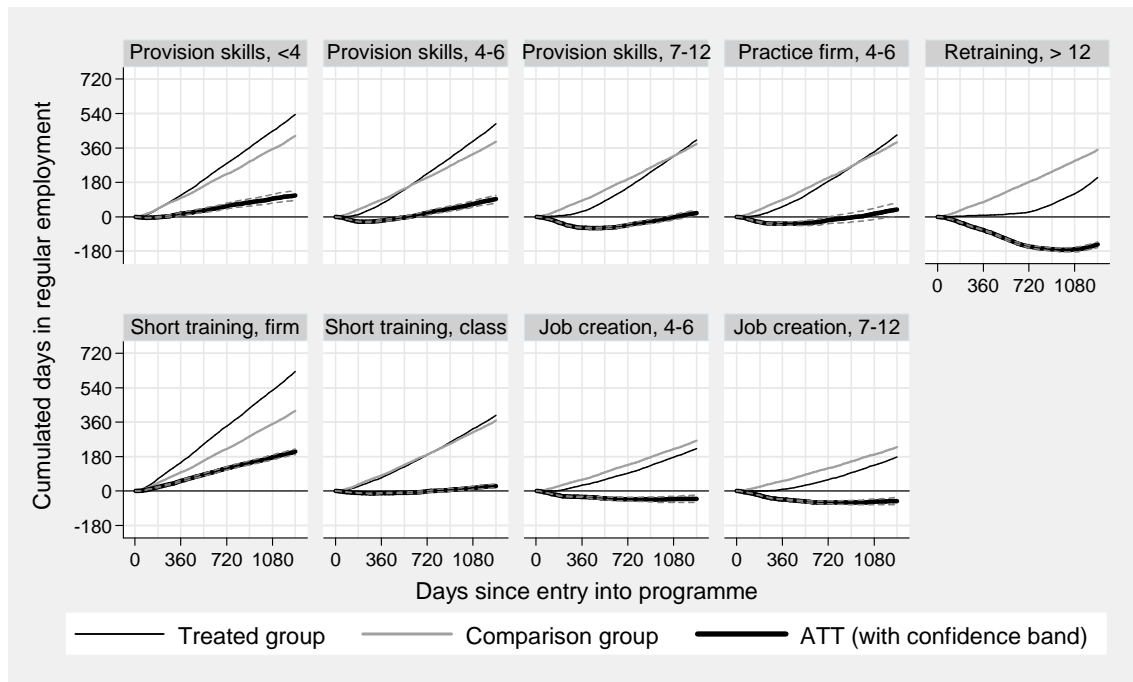
*) $\alpha = 0.05$, **) $\alpha = 0.01$.

Source: Own calculations, based on the TrEffeR data.

Shaded in light grey: Significant taking into account Lechner-bounds and $MSB < 5$.

Matching algorithm: Radius matching with calliper of 0.0005.

Figure 1 Cumulated days in regular employment during the 3.5 years after programme entry: Averages for treatment and comparison group as well as average treatment effects on participants compared to “waiting”

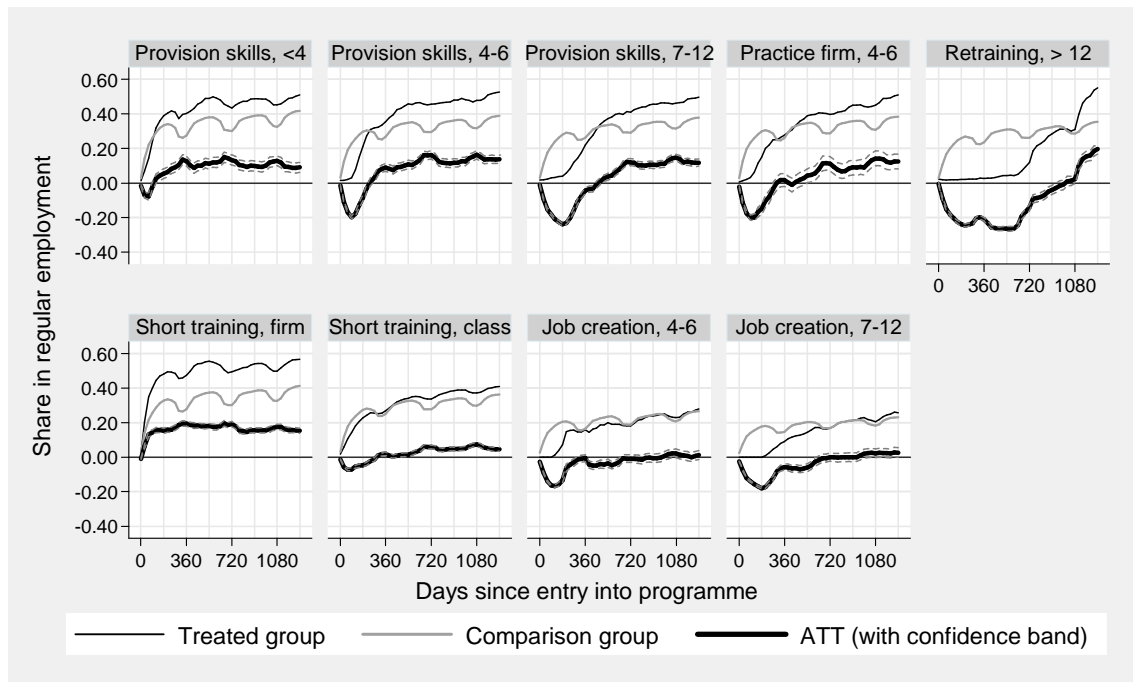


Source: Own calculations, based on the TrEffeR data.
 Matching algorithm: Radius matching with calliper of 0.0005.

Overall, Table 4 and Figure 1 show rather mixed results of programme participation on cumulated days the participants have spent in regular employment during the 3.5 years after treatment start. We find highly significant positive effects of further vocational training providing professional skills with programme durations of up to six months. Longer training providing skills for 7 to 12 months as well as training in practice firms for 4 to 6 months have significant, but small effects on the number of days in regular employment. The cumulated effect of long retraining programmes is significantly negative. However, this is not surprising, because substantial lock-in effects are a necessary side-effect of this kind of programmes. Regarding short training programmes, the effect on days spent in employment amounts to about 200 days for firm-internal programmes, but is small for classroom training. This underlines the fact that it is important to distinguish between different variants of short training programmes. Finally, job

creation schemes with duration between 7 and 12 months have significant negative effects on cumulated days spent in employment.

Figure 2 Share in regular employment during the 3.5 years after programme entry: Averages for treatment and comparison group as well as average treatment effects on participants compared to “waiting”



Source: Own calculations, based on the TrEffeR data.
 Matching algorithm: Radius matching with calliper of 0.0005.

A surprising fact is, however, that we obtain mostly significantly positive effects on shares of participants in regular employment, 3.5 years after programme entry (Table 4 and Figure 2). The only exceptions are shorter job creation schemes that have insignificant effects at the end of the observation period. Furthermore, effects are largest for long retraining schemes and arise – as can be seen in Figure 2 – only at the very end of the observation period. Thus, one can expect that cumulated effects also could have turned positive for all further training programmes if the observation period would have been longer.

These results are in line with those obtained by Biewen et al. (2007) and Rinne et al. (2007), who found also (at least partly) positive employment effects of further vocational training and

short training programme, but different from the findings of Wunsch and Lechner (2008). One reason might be that the latter analysed a shorter time period after programme entry (30 months compared to 42 months in this study), while programme effects might take some time to break even. Second, Wunsch and Lechner (2008) restricted their sample to the age group 25 to 49. However, our results are even more positive, if we introduce a similar age restriction. Third, we analyse only (short-term unemployed) persons who entered a programme during the first year of their unemployment spell. Fourth, Wunsch and Lechner (2008) defined non-participants as persons who did not enter a programme during the 18 months following the inflow date into their sample. In contrast, Biewen et al. (2007) as well as Rinne et al. (2007) performed separate estimates by duration of unemployment at the beginning of a treatment and required only that non-participants did not enter a measure during an accordingly chosen classification window. We defined all those individuals, who did not enter a programme during one single month, as non-participants. For job creation schemes our results are similar to the findings of Caliendo et al. (2006; 2008a; 2008b), who observed a negative impact of participation in job creation schemes on the employment prospects of participants. While we obtain non-negative, but insignificant effects of participation on shares in employment at the end of the observation period, we observe outcome variables for a slightly longer period of time (42 instead of 36 months) and analyse programme entries starting three years later than these authors did (2003 instead of 2000).

Pairwise comparison of programme participations

We turn now to a pairwise comparison of participation in particular programmes, which is presented in Table 5. We usually find a sufficiently “good” comparison group when comparing participation in variants of further vocational training schemes and short training schemes. Regarding further vocational training programmes, it seems that participants in shorter programmes have spent more days in regular employment during the observation period. However, the mostly insignificant differences in shares in employment indicate that this “advantage” may

not continue to increase over time. Thus the more positive effects of shorter programmes on cumulated days spent in employment are clearly an effect of the lock-in effects of participation increasing with programme duration. Furthermore, when Lechner bounds are taken into account, differences in both outcome variables are mostly insignificant comparing training, aimed at the provision of skills, with training in a practice firm.

Table 5 Estimated effect of participation in programme J compared to participation in programme K, 3.5 years after programme entry

Treatment group J (type and duration in months)	Comparison group K (type and duration in months)								
	Provision of skills			Practice firm	Re-training	Short training		Job creation scheme	
	<4	4-6	7-12	4-6	>12	in firm	in class	4-6	7-12
1. Cumulated days in regular employment									
Provision of skills, <4		18	99**	87**	323**	-93**	92**	176**	199**
Provision of skills, 4-6	-14		70**	60*	281**	-128**	68**	141**	130**
Provision of skills, 7-12	-91**	-80**		-37	213**	-197**	-8	8	35
Practice firm, 4-6	-65*	-51*	22		241**	-191**	19	11	105**
Retraining, >12	-237**	-280**	-186**	-206**		-379**	-174**	-138**	-94**
Short training in firm	113**	117**	194**	194**	430**		188**	264**	285**
Short training in class	-71**	-65**	2	6	209**	-201**		57	97**
Job creation scheme, 4-6	-120**	-130**	-86**	-94	59*	-255**	-87**		48**
Job creation scheme, 7-12	-201**	-124**	-121**	-158**	-15	-292**	-93**	-45**	
2. Share in regular employment									
Provision of skills, <4		-0.05**	-0.02	-0.04	-0.05	-0.07**	0.05**	0.13**	0.07
Provision of skills, 4-6	0.06**		0.02	0.01	-0.02	-0.03*	0.09**	0.20**	0.12**
Provision of skills, 7-12	0.02	-0.03		-0.06	-0.05	-0.05**	0.07**	0.09	0.01
Practice firm, 4-6	0.04	-0.01	0.01		-0.06	-0.06*	0.08**	0.07	0.11*
Retraining, >12	0.14**	0.05*	0.07**	0.06		0.03	0.15**	0.21**	0.15**
Short training in firm	0.08**	0.02	0.05**	0.06	0.03		0.11**	0.21**	0.15**
Short training in class	-0.01	-0.09**	-0.07**	-0.07*	-0.12**	-0.13**		0.06	0.01
Job creation scheme, 4-6	-0.04	-0.10*	-0.13**	-0.13	-0.25**	-0.14**	-0.06**		0.01
Job creation scheme, 7-12	-0.10*	-0.09*	-0.13**	-0.17*	-0.30**	-0.15**	-0.03	0.00	

*) $\alpha = 0.05$, **) $\alpha = 0.01$.

Source: Own calculations, based on the TrEffeR data.

Shaded in light grey: Significant taking into account Lechner-bounds and $MSB < 5$.

Matching algorithm: Radius matching with calliper of 0.005.

Lock-in effects are naturally largest for long retraining programmes, and thus retraining performs worst among all programmes when comparing cumulated days in regular employment during the 3.5 years after programme start. However, the share in employment at the end of the observation period is marginally significantly larger than is the case for several shorter further vocational training programmes. This implies that the effectiveness of retraining programmes

measured by cumulated days in employment could be supposed to increase further over time. For this programme, Lechner et al. (2005, Table 6.1, 38) found rather positive – but only partly significant results – for an observation period of seven years, compared to shorter programmes.

The “winners” in the pairwise comparison are obviously short training programmes conducted within firms: Participants spend more days in regular employment than those in every other programme, while they would have fared far worse in most other programmes. The explanation at hand is that short firm-internal training programmes are – similarly to wage subsidies – not only a training programme, but rather a training programme in combination with access to a firm. They offer participants the possibility not only to increase their productivity, but also to convey their productivity to a potential employer. They may thus be used as a kind of cheap probation period by employers, where the “wage” is paid by the Public Employment Service.

In contrast, participants in short classroom training have spent less days in regular employment during the observation period than those individuals participating in further vocational training aimed at the provision of skills with an duration of up to six months; and also participants in the latter fared better with their training compared to a short classroom training. Thus one might conclude that previous results of Biewen et al. (2007) and Wunsch and Lechner (2008), who obtained no advantages of participation in further vocational training compared to short programmes, are partly a result of the fact that these studies did not distinguish between variants of short training programmes.

Finally, as has already been discussed, it is mostly not possible to find a satisfactory comparable group of individuals participating in a training programme for participants in job creation schemes, vice versa. Differences in characteristics of participants are due to the age structure of participants (for instance, retraining is typically granted for younger workers, while participants in job creation schemes are in average older) and to time spent in unemployment during the last years, but also due to the regional distribution of programme assignment (the overwhelming

majority of entries in job creation schemes in March 2003 took place in Berlin and within other East German urban areas with high unemployment).

VIII CONCLUSIONS

For Germany, our paper estimated the effects of participation in further training programmes, short training programmes and job creation schemes, starting in March 2003, as well as the relative effectiveness of these programmes. The analysis was restricted to persons of age 25 to 59 and on the first programme during the first year of an unemployment spell.

First, our analysis demonstrated that estimates of relative programme effectiveness face more severe problems of common support than the more often conducted estimates of programme effects compared to a state of “waiting” – simply because the number of potential comparison persons participating in other programmes is smaller. Thus it is important to match with replacement. Furthermore, for participants in job creation schemes it is mostly difficult to find an adequate comparison group from participants in training programmes, vice versa. That might be taken as a hint that participants in this programme – whose objective is explicitly not to achieve employment but to increase employability – are in fact a strongly selected group of hard-to-place individuals. Other programmes available in the time period investigated might not have been suitable for this group. Thus, it remains at least questionable whether participation in another programme could have improved the labour market prospects of participants in job creation schemes.

Second, compared to non-participation in the sense of “waiting”, we find that participation in further vocational training aimed at the provision of skills as well as short training programmes increase the number of days that participants have spent in regular employment during the 3.5 years after programme start. In contrast, participation in retraining and job creation schemes

decreases the number of days. However, all programmes had a positive impact on the share of participants in regular employment at the end of the observation period.

Third, the pairwise comparison of programmes conveys the impression that across further vocational training programmes, shorter programmes perform overall better; this is mostly the result of shorter lock-in effects. Also participation in further training aimed at the provision of skills with duration of up to six months has been more advantageous for treated persons than short classroom training would have been. In contrast, short firm-internal training is by the far the most “successful” programme in the portfolio of programmes included in this investigation. This is, however, probably related to the fundamentally different design of this programme, which requires an employer willing to offer a training opportunity.

Overall, our study indicates that in the longer run participation in training programmes – but not in job creation schemes – supports the reintegration of previously unemployed workers into the labour market. Comparing programmes, some variants of further vocational training programmes (aimed at the provision of skills and with duration of up to 6 months) have been relatively more effective – but at the same time also more costly – than short classroom training. Participants in job creation schemes differ too strongly from participants in training schemes to conduct a reliable comparison. Still, our results refer to the impact of programmes on individual employment prospects of participants; Layard et al. (1991, 481) emphasise that for a judgment of the welfare costs of a programme much more has to be taken into account: Beneath the programme costs and the employment effects at the macro level, output benefits, psychic benefits and social costs, as well as the distributional incidence of active labour market programmes have to be considered.

REFERENCES

- ABADIE, A. and IMBENS, W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74, 235-267.
- ABBRING, J.H. and VAN DEN BERG, G. (2003). The non-parametric identification of treatment effects in duration models. *Econometrica*, 71, 1491-1518.
- ABBRING, J.H. and VAN DEN BERG, G. (2004). Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data model. *Empirical Economics*, 29, 5-20.
- BERNHARD, S., HOHMEYER, K., JOZWIAK, E., KOCH, S., KRUPPE, T., STEPHAN, G. and WOLFF, J. (2008). Aktive Arbeitsmarktpolitik in Deutschland und ihre Wirkungen. IAB-Forschungsbericht 2/2008.
- BIEWEN, M., FITZENBERGER, B., OSIKOMINU, A. and WALLER, M. (2007). Which program for whom? Evidence on the comparative effectiveness of public sponsored training programs in Germany. IZA Discussion Paper 2885.
- BLACK, D.A. and SMITH, J.A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, 21, 99-124
- BLIEN, U., HIRSCHENAUER, F., ARENDT, M., BRAUN, H.-J., GUNST, D.-M., S., KLEINSCHMIDT, H., MUSATI, M., ROß, H., VOLKKOMMER, D. and WEIN, J. (2004). Typisierung von Bezirken der Agenturen für Arbeit. *Zeitschrift für Arbeitsmarktforschung*, 37, 146-175.
- BÜTTNER, T. (2008). Ankündigungseffekt oder Maßnahmewirkung? Eine Evaluation von Trainingsmaßnahmen zur Überprüfung der Verfügbarkeit. *Zeitschrift für ArbeitsmarktFor-schung*, 41, 25-40.
- CALIENDO, M. and HUJER, R. (2006). The microeconomic estimation of treatment effects – An overview. *Allgemeines Statistisches Archiv*, 90, 197-212.
- CALIENDO, M., HUJER, R. and THOMSEN, S. (2006). Sectoral heterogeneity in the employment effects of job creation schemes in Germany. *Jahrbücher für Nationalökonomie und Statistik*, 226, 139-179.
- CALIENDO, M., HUJER, R. and THOMSEN, S. (2008a). Identifying effect heterogeneity to improve the efficiency of job creation schemes in Germany. *Applied Economics*, 40, 1101-1122.
- CALIENDO, M., HUJER, R. and THOMSEN, S. (2008b). The employment effect of job creation schemes in Germany, A microeconomic evaluation. In D. Millimet, J. Smith and E. Vytlacil (eds.), *Advances in Econometrics*, Vol. 21, Estimating and Evaluating Treatment Effects in Econometrics, New York: Elsevier-Science, 383-430.
- CALIENDO, M. and KOPEINIG, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22, 31-72.
- DEHEJIA, R. H. and WAHBA, S. (2002). Propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 84, 151-161.
- DORSETT, R. (2001). The new deal for young people: Relative effectiveness of the options in reducing male unemployment. PSI Discussion Paper 7.

- FITZENBERGER, B., OSIKOMINU, A. and VÖLTER, R. (2006). Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany. IAB-Discussion Paper 17/2006.
- FITZENBERGER, B. and VÖLTER, R. (2007). Long-run effects of training programs for the unemployed in East Germany. *Labour Economics*, 14, 730-755.
- FREDRIKSSON, P. and JOHANSSON, P. (2004). Dynamic treatment assignment – The consequences for evaluations using observational data. IZA Discussion Paper 1062.
- GERFIN, M. and LECHNER, M. (2002). Microeconomic evaluation of the active labour market policy in Switzerland. *The Economic Journal*, 112, 854-893.
- GERFIN, M., LECHNER, M. and STEIGER, H. (2004). Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes. *Labour Economics*, 12, 807-835.
- HECKMAN, J.J., ICHIMURA, H. and TODD, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies*, 64, 605-654.
- HECKMAN, J.J., LALONDE, R.L. and SMITH, J.A. (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter, D. Card (eds.), *Handbook of Labor Economics* 3, Amsterdam: Elsevier., 1865-2097.
- HOHMEYER, K. and WOLFF, J. (2007). A fistful of Euros: Does One-Euro-Job participation lead means-tested benefit recipients into regular jobs and out of unemployment benefit II receipt? IAB-Discussion Paper 32/2007.
- HUJER, R. and THOMSEN, S. (2006). How do employment effects of job creation schemes differ with respect to the foregoing unemployment duration. ZEW-Discussion Paper 06-047.
- HUJER, R. and ZEISS, C. (2006). The effects of job creation schemes on the unemployment duration in East Germany. IAB-Discussion Paper 16/2006.
- HUJER, R., THOMSEN, S. and ZEISS, C. (2006a). The effects of vocational training programmes on the duration of unemployment in Eastern Germany. *Allgemeines Statistisches Archiv*, 90, 299-322.
- HUJER, R., THOMSEN, S. and ZEISS, C. (2006b). The effects of short-term training measures on the individual unemployment duration in West Germany. ZEW-Discussion Paper 06-065.
- IMBENS, G.W. (1999). The role of propensity score in estimating dose response functions. NBER Technical Working Paper 237.
- KLUGE, J., SCHNEIDER, H., UHLENDORFF, A. and ZHAO, Z. (2007). Evaluating continuous training programs using the generalized propensity score, IZA Discussion Paper 3255.
- KRUPPE, T. (2008). Selektivität bei der Einlösung von Bildungsgutscheinen, IAB-Discussion Paper 17/2008.
- KURTZ, B. (2003). Trainingsmaßnahmen – was verbirgt sich dahinter? IAB-Werkstattbericht 8/2003.
- LAYARD, R., NICKELL, S. and JACKMAN, R. (2001). *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford: Oxford University Press.

- LECHNER, M. (2000). A note on the common support problem in applied evaluation studies. SIAW-Discussion Paper.
- 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 Policies in Europe*, Heidelberg: Physica/Springer, 43-58.
- LECHNER, M. (2002): Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *The Review of Economics and Statistics*, 84, 205-220.
- LECHNER, M., MIQUEL, R. and WUNSCH, C. (2005). Long-run effects of public sector sponsored training. IAB-Discussion Paper 3/2005.
- LECHNER, M., MIQUEL, R. and WUNSCH, C. (2007). The curse and blessing of training the unemployed in a changing economy: The case of East Germany after unification. *German Economic Review*, 8, 468-509.
- LEUVEN, E. and SIANESI, B. (2003). psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing and covariate imbalance testing. <http://ideas.repec.org/c/boc/bo-code/s432001.htm>.
- RINNE, U., SCHNEIDER, M. and UHLENDORFF, A. (2007). Too bad to benefit? Effect heterogeneity of public training programs. IZA Discussion Paper 3240.
- ROSENBAUM, P.R. and RUBIN, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-50.
- RUBIN, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- SIANESI, B. (2004). An evaluation of the swedish system of active labor market programs in the 1990s. *The Review of Economics and Statistics*, 86, 133-155.
- SIANESI, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics*, 15, 392-421.
- SMITH, J. (2000), A critical survey of empirical methods for evaluating active labor market policies. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*. 136, 1-22.
- STEIGER, H. (2004). Is less more? A look at nonparticipation in Swiss active labour market programmes, mimeo.
- STEPHAN, G. (2008). The effects of active labor market programs in Germany: An investigation using different definitions of non-treatment. IAB-Discussion Paper 12/2008.
- STEPHAN, G., RÄSSLER, S. and SCHEWE, T. (2006). Das TrEffeR-Projekt der Bundesagentur für Arbeit. *Zeitschrift für ArbeitsmarktForschung*, 39, 447-465.
- WOLFF, J. and JOZWIAK, E. (2007). Does short-term training activate means-tested unemployment benefit recipients in Germany? IAB-Discussion Paper 29/2007.
- WUNSCH, C. and LECHNER, M. (2008). What did all the money do? On the general ineffectiveness of recent West German labour market programmes. *Kyklos*, 61, 134-174.

APPENDIX

Table A.1 Variable means (0 = no, 1 = yes) before matching

Variables	Waiting	Provision of skills			Pract. firm 4-6	Retraining >12	Short training		Job creation scheme	
		<4	4-6	6-12			in firm	in class	4-6	6-12
a) Socio-demographic characteristics										
Female	0.42	0.32	0.44	0.43	0.53	0.56	0.33	0.52	0.44	0.47
Age 25-29	0.14	0.18	0.16	0.12	0.14	0.24	0.22	0.16	0.08	0.06
Age 30-34	0.16	0.18	0.20	0.21	0.18	0.25	0.20	0.18	0.09	0.08
Age 35-39	0.18	0.20	0.24	0.22	0.16	0.26	0.20	0.21	0.12	0.12
Age 40-44	0.16	0.19	0.19	0.22	0.22	0.16	0.17	0.19	0.14	0.14
Age 45-49	0.14	0.15	0.13	0.14	0.19	0.07	0.12	0.14	0.18	0.19
Age 50-54	0.13	0.08	0.07	0.07	0.09	0.02	0.07	0.10	0.24	0.24
Age 54-59	0.09	0.01	0.01	0.01	0.02	0.00	0.02	0.03	0.16	0.17
Health problems	0.12	0.04	0.05	0.05	0.05	0.05	0.07	0.06	0.11	0.20
Slightly disabled	0.03	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.06
Severely disabled	0.04	0.01	0.02	0.03	0.02	0.00	0.03	0.02	0.08	0.12
Married	0.54	0.53	0.56	0.54	0.58	0.52	0.48	0.52	0.62	0.57
Married and female	0.25	0.16	0.25	0.25	0.31	0.32	0.15	0.29	0.29	0.29
Foreigner	0.13	0.08	0.11	0.08	0.09	0.08	0.08	0.11	0.04	0.05
Without secondary degree	0.12	0.06	0.07	0.03	0.05	0.05	0.06	0.07	0.07	0.08
Secondary degree (Hauptschule)	0.46	0.34	0.35	0.24	0.36	0.31	0.39	0.34	0.32	0.36
Secondary degree (Realschule)	0.30	0.44	0.38	0.40	0.45	0.49	0.38	0.38	0.48	0.42
Secondary degree (Gymnasium)	0.12	0.17	0.21	0.33	0.14	0.15	0.17	0.21	0.13	0.15
Without vocational training	0.32	0.18	0.22	0.14	0.19	0.31	0.19	0.21	0.14	0.21
Vocational training	0.62	0.75	0.67	0.69	0.75	0.65	0.74	0.68	0.78	0.70
University degree	0.06	0.07	0.10	0.17	0.06	0.04	0.08	0.11	0.07	0.09
Unemployment benefit receipt	0.50	0.68	0.63	0.61	0.66	0.54	0.66	0.57	0.69	0.66
Unemployment assistance receipt	0.15	0.10	0.10	0.11	0.11	0.19	0.11	0.12	0.19	0.22
No benefit receipt	0.35	0.22	0.27	0.28	0.23	0.27	0.23	0.31	0.12	0.13
b) Month of program entry										
1st month of unemployment	0.12	0.11	0.10	0.11	0.09	0.18	0.16	0.10	0.06	0.07
2nd month of unemployment	0.12	0.13	0.11	0.09	0.09	0.10	0.15	0.13	0.07	0.09
3rd month of unemployment	0.17	0.15	0.13	0.13	0.16	0.11	0.16	0.18	0.08	0.08
4th month of unemployment	0.12	0.11	0.11	0.10	0.11	0.09	0.11	0.11	0.10	0.11
5th month of unemployment	0.09	0.11	0.10	0.09	0.09	0.09	0.09	0.10	0.08	0.07
6th month of unemployment	0.08	0.09	0.10	0.10	0.10	0.08	0.07	0.09	0.10	0.08
7th month of unemployment	0.06	0.06	0.08	0.08	0.06	0.07	0.06	0.07	0.09	0.10
8th month of unemployment	0.05	0.06	0.07	0.06	0.07	0.06	0.04	0.05	0.09	0.09
9th month of unemployment	0.06	0.07	0.06	0.08	0.08	0.07	0.06	0.06	0.10	0.07
10th month of unemployment	0.04	0.04	0.05	0.06	0.05	0.05	0.04	0.04	0.08	0.08
11th month of unemployment	0.04	0.04	0.05	0.04	0.03	0.04	0.04	0.04	0.07	0.08
12th month of unemployment	0.04	0.04	0.05	0.06	0.06	0.05	0.03	0.04	0.09	0.08
c) 2-year-history preceding current spell										
Unemployed up to 1 month	0.42	0.54	0.57	0.55	0.55	0.47	0.48	0.55	0.37	0.31
Unemployed 1-6 months	0.18	0.21	0.18	0.18	0.20	0.16	0.22	0.18	0.17	0.17
Unemployed 7-12 months	0.18	0.15	0.14	0.14	0.13	0.14	0.17	0.13	0.20	0.23
Unemployed 13-18 months	0.12	0.06	0.06	0.08	0.07	0.11	0.08	0.08	0.15	0.16
Unemployed 19-24 months	0.10	0.04	0.05	0.06	0.05	0.11	0.04	0.06	0.11	0.14
Participation in active labour market programmes	0.27	0.27	0.23	0.28	0.24	0.35	0.34	0.26	0.37	0.48
Period of sickness	0.14	0.08	0.09	0.08	0.11	0.10	0.09	0.09	0.17	0.20
Sanction imposed	0.03	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.01	0.01
d) Regional labour market situation										
East Germany, worst situation	0.04	0.04	0.03	0.05	0.04	0.06	0.05	0.05	0.19	0.06
East Germany, bad situation	0.19	0.24	0.18	0.21	0.20	0.31	0.16	0.19	0.32	0.44
East Germany, high unemployment	0.04	0.04	0.02	0.05	0.05	0.03	0.03	0.04	0.06	0.04
Metropolitan area, high unemployment	0.10	0.11	0.08	0.13	0.01	0.10	0.07	0.08	0.34	0.27
Metropolitan area, medium unemployment	0.11	0.15	0.14	0.16	0.04	0.10	0.11	0.13	0.03	0.06
Above average unemp., moderate dynamics	0.06	0.06	0.06	0.05	0.09	0.05	0.07	0.06	0.02	0.02
Rural area, average unemployment	0.06	0.09	0.07	0.05	0.09	0.02	0.08	0.04	0.00	0.02
Rural area, below average unemployment	0.13	0.09	0.14	0.10	0.14	0.18	0.18	0.11	0.01	0.03
Metropolitan area, good situation, high dynamics	0.09	0.03	0.08	0.08	0.07	0.05	0.08	0.16	0.01	0.02
Rural area, good situation and high dynamics	0.03	0.03	0.03	0.02	0.01	0.04	0.02	0.01	0.01	0.02
Small-business dominated, good situation	0.09	0.06	0.12	0.08	0.17	0.05	0.09	0.09	0.02	0.02
Region with very good situation	0.05	0.05	0.04	0.02	0.07	0.03	0.05	0.04	0.00	0.00
Mean programme duration		57	149	278	154	820	30	44	174	318
Cumulated days in employment after 3.5 years	324	535	486	402	428	205	625	395	223	178
Share in regular employment after 3.5 years	0.32	0.51	0.53	0.50	0.51	0.55	0.57	0.41	0.28	0.26
Observations	192460	1146	1744	2332	532	1121	3967	7792	1199	1039

Source: Own calculations, based on the TrEffeR data.

APPENDIX

Table A.2 Share of treatment group J within common support and maximum weight of a comparison person as share of treated persons

Treatment group J (type and duration in months)	Comparison group K (type and duration in months)									
	Waiting	Provision of skills			Practice firm	Re-training	Short training		Job creation scheme	
		<4	4-6	7-12	4-6	>12	in firm	in class	4-6	7-12
1 Share of treatment group J within common support										
<i>1.1 Nearest neighbour matching without replacement</i>										
Provision of skills, <4	1.00		0.93	0.92	0.38	0.55	0.99	0.99	0.37	0.36
Provision of skills, 4-6	1.00	0.61		0.85	0.30	0.47	0.96	0.99	0.28	0.28
Provision of skills, 7-12	1.00	0.45	0.64		0.21	0.39	0.85	0.97	0.30	0.27
Practice firm, 4-6	1.00	0.82	0.98	0.92		0.63	0.98	1.00	0.40	0.42
Retraining, >12	1.00	0.56	0.74	0.80	0.30		0.91	0.98	0.33	0.31
Short training in firm	1.00	0.29	0.42	0.50	0.13	0.26		0.92	0.17	0.17
Short training in class	1.00	0.15	0.22	0.29	0.07	0.14	0.47		0.13	0.12
Job creation scheme, 4-6	1.00	0.35	0.40	0.58	0.18	0.31	0.57	0.82		0.67
Job creation scheme, 7-12	0.99	0.40	0.47	0.61	0.22	0.33	0.63	0.87	0.78	
<i>1.2 Radius matching</i>										
Provision of skills, <4	1.00		0.99	0.98	0.93	0.93	1.00	1.00	0.90	0.88
Provision of skills, 4-6	1.00	0.99		0.99	0.96	0.96	0.99	1.00	0.99	0.97
Provision of skills, 7-12	1.00	0.99	1.00		0.89	0.98	0.99	1.00	0.97	0.99
Practice firm, 4-6	1.00	0.97	0.99	0.97		0.89	1.00	1.00	0.75	0.78
Retraining, >12	1.00	0.97	0.98	0.99	0.90		0.99	1.00	0.91	0.85
Short training in firm	1.00	1.00	1.00	1.00	0.98	0.95		1.00	1.00	0.98
Short training in class	1.00	1.00	1.00	0.99	0.99	0.94	0.99		0.93	0.96
Job creation scheme, 4-6	1.00	0.93	0.87	0.95	0.84	0.75	0.92	0.97		0.97
Job creation scheme, 7-12	0.99	0.94	0.94	0.93	0.79	0.70	0.89	0.98	0.98	
2 Maximum weight of a comparison person as share of treated persons										
<i>2.1 Nearest neighbour matching with replacement (1/N)</i>										
Provision of skills, <4	0.001		0.001	0.001	0.002	0.002	0.001	0.001	0.002	0.002
Provision of skills, 4-6	0.001	0.001		0.001	0.002	0.001	0.001	0.001	0.002	0.002
Provision of skills, 7-12	0.0004	0.001	0.001		0.002	0.001	0.001	0.000	0.001	0.002
Practice firm, 4-6	0.002	0.002	0.002	0.002		0.003	0.002	0.002	0.005	0.004
Retraining, >12	0.001	0.002	0.001	0.001	0.003		0.001	0.001	0.003	0.003
Short training in firm	0.0003	0.001	0.001	0.001	0.002	0.001		0.0003	0.001	0.002
Short training in class	0.0001	0.001	0.001	0.000	0.002	0.001	0.0003		0.001	0.001
Job creation scheme, 4-6	0.001	0.002	0.002	0.001	0.005	0.003	0.001	0.001		0.001
Job creation scheme, 7-12	0.001	0.002	0.002	0.002	0.004	0.003	0.002	0.001	0.001	
<i>2.2 Radius matching</i>										
Provision of skills, <4	0.0001		0.003	0.005	0.015	0.018	0.002	0.002	0.043	0.033
Provision of skills, 4-6	0.001	0.008		0.004	0.024	0.016	0.002	0.001	0.106	0.048
Provision of skills, 7-12	0.001	0.009	0.008		0.013	0.016	0.003	0.003	0.058	0.043
Practice firm, 4-6	0.0001	0.016	0.004	0.008		0.027	0.005	0.002	0.065	0.055
Retraining, >12	0.001	0.016	0.009	0.006	0.025		0.008	0.003	0.054	0.029
Short training in firm	0.0003	0.004	0.003	0.003	0.017	0.013		0.002	0.040	0.033
Short training in class	0.0003	0.012	0.004	0.004	0.018	0.012	0.003		0.027	0.021
Job creation scheme, 4-6	0.001	0.036	0.026	0.020	0.097	0.051	0.019	0.008		0.017
Job creation scheme, 7-12	0.001	0.028	0.037	0.014	0.068	0.060	0.014	0.007	0.006	

Source: Own calculations, based on the TrEffeR data.

APPENDIX

Table A.3 Mean standardized bias before and after matching

Treatment group J (type and duration in months)	Comparison group K (type and duration in months)									
	Waiting	Provision of skills			Practice firm	Re-training	Short training		Job creation scheme	
		<4	4-6	7-12	4-6	>12	in firm	in class	4-6	7-12
1 Mean standardised bias before matching										
Provision of skills, <4	11.1		6.8	8.6	9.0	11.9	6.1	8.1	20.4	21.4
Provision of skills, 4-6	9.5	6.8		6.1	8.3	11.0	7.3	5.4	21.8	22.0
Provision of skills, 7-12	11.6	8.6	6.1		10.6	11.7	10.3	7.5	20.2	20.9
Practice firm, 4-6	12.3	9.0	8.3	10.6		13.4	9.9	9.1	19.6	21.4
Retraining, >12	12.5	11.9	11.0	11.7	13.4		12.7	11.5	17.6	17.8
Short training in firm	9.7	6.1	7.3	10.3	9.9	12.7		7.9	22.2	22.6
Short training in class	8.3	8.1	5.4	7.5	9.1	11.5	7.9		21.2	21.3
Job creation scheme, 4-6	19.0	20.4	21.8	20.2	19.6	17.6	22.2	21.2		7.6
Job creation scheme, 7-12	19.1	21.4	22.0	20.9	21.4	17.8	22.6	21.3	7.6	
2 Mean standardised bias after matching										
<i>2.1 Neighest neighbour matching without replacement</i>										
Provision of skills, <4	3.0		1.4	1.9	2.5	2.5	2.3	2.4	3.2	2.6
Provision of skills, 4-6	1.9	1.7		1.4	3.3	2.2	1.3	1.7	2.7	3.4
Provision of skills, 7-12	2.1	1.3	1.4		3.3	2.0	1.1	1.5	2.5	2.2
Practice firm, 4-6	3.4	2.7	2.2	3.4		3.9	2.8	3.3	5.1	4.9
Retraining, >12	2.2	2.5	2.1	1.7	4.3		2.3	2.0	3.9	3.1
Short training in firm	1.8	2.3	1.5	1.3	3.0	2.5		1.0	3.4	3.6
Short training in class	1.5	3.0	2.2	1.4	4.0	2.5	1.2		2.7	2.7
Job creation scheme, 4-6	2.2	3.1	3.2	2.7	5.3	4.2	2.7	2.2		1.6
Job creation scheme, 7-12	1.9	3.4	2.9	2.4	4.0	3.7	3.1	2.9	1.6	
<i>2.2 Radius matching</i>										
Provision of skills, <4	0.4		1.0	1.2	2.8	3.7	0.8	0.8	6.6	5.5
Provision of skills, 4-6	0.3	1.3		0.9	2.2	2.3	0.9	0.4	9.8	8.9
Provision of skills, 7-12	0.5	1.5	1.2		3.6	2.7	1.4	0.7	8.5	6.8
Practice firm, 4-6	0.6	3.1	1.3	2.2		4.6	1.3	0.7	8.0	5.4
Retraining, >12	0.4	2.3	1.5	1.7	3.6		1.5	0.9	6.7	5.4
Short training in firm	0.2	1.3	1.2	1.4	3.4	3.4		0.6	9.1	7.0
Short training in class	0.2	2.5	1.1	1.4	3.5	2.5	0.9		10.1	6.6
Job creation scheme, 4-6	1.7	7.8	6.1	4.0	7.6	7.9	3.8	2.9		2.4
Job creation scheme, 7-12	1.0	6.4	6.6	3.4	8.3	6.6	3.8	2.4	1.6	

Source: Own calculations, based on the TrEffeR data.