Identifying Effect Heterogeneity to Improve the Efficiency of Job Creation Schemes in Germany

Marco Caliendo,[*] Reinhard Hujer[†] and Stephan L. Thomsen[‡]

[†]*DIW, Berlin and IZA, Bonn*

[‡]*J.W.Goethe-University, Frankfurt/Main, IZA, Bonn, ZEW, Mannheim*

[§]*J.W.Goethe-University, Frankfurt/Main*

Abstract

Previous empirical studies of job creation schemes in Germany have shown that the average effects for the participating individuals are negative. However, we find that this is not true for all strata of the population. Identifying individual characteristics that are responsible for the effect heterogeneity and using this information for a better allocation of individuals therefore bears some scope for improving programme efficiency. We present several stratification strategies and discuss the occurring effect heterogeneity. Our findings show that job creation schemes do neither harm nor improve the labour market chances for most of the groups. Exceptions are long-term unemployed men in West and long-term unemployed women in East and West Germany who benefit from participation in terms of higher employment rates.

Keywords: Evaluation – Active Labour Market Policy – Targeting – Efficient Allocation – Effect Heterogeneity
JEL Classification: C13, H43, J68

[*]Marco Caliendo is Senior Research Associate at the German Institute for Economic Research (DIW Berlin) and Research Fellow of the IZA, Bonn, e-mail: mcaliendo@diw.de.

[†]Reinhard Hujer is Professor of Statistics and Econometrics at the J.W.Goethe-University of Frankfurt/Main, and Research Fellow of the IZA, Bonn and the ZEW, Mannheim, e-mail: hujer@wiwi.uni-frankfurt.de.

[‡]Stephan L. Thomsen is Research Assistant at the Institute of Statistics and Econometrics, J.W.Goethe-University of Frankfurt/Main, e-mail: sthomsen@wiwi.uni-frankfurt.de.

1 Introduction

The permanent integration into regular employment is the primary purpose of active labour market policy (ALMP) in Germany. To achieve this goal, the Federal Employment Agency (FEA) spends substantial amounts on measures like vocational training programmes (VT), job creation schemes (JCS) and special promotion for disabled people and aspirants for vocational rehabilitation. ALMP was first introduced in Germany in the late 1960s. Since then, the labour market experienced several important changes, caused by the oil price shocks during the 1970s and the growth of the labour market after the German Re-Unification in 1990. The set of programmes was gradually adjusted to these changes. Despite these reforms and large spending on ALMP, the German labour market is still plagued by high and persistent unemployment. Therefore, evaluating ALMP has become a major topic and was also legally anchored in the reformed legal basis for ALMP in 1998 (Social Code III). The main question to be answered is, if programmes improve the employment chances of participants.

In this paper we evaluate the effects of JCS for the participating individuals. JCS, which have been one major element of ALMP in Germany over the last years, are a form of subsidised employment and aim at the stabilisation and qualification of unemployed persons with disadvantages on the labour market. The main purpose of these programmes is the (re-)integration of unemployed persons into the first labour market.[1] Recent empirical studies of JCS for Germany have shown that the average effects for the participating individuals are negative (see e.g. Caliendo, Hujer, and and Thomsen, 2005a and 2005b).[2] The reasons for these findings have to be analysed. One possible explanation may be the poor quality of programmes in conjunction with often cited stigma- and 'locking-in'-effects. But leaving this argument aside for a moment, the results may also come from inefficient allocation mechanisms. The central motivation in this context is that programme impacts are heterogeneous (Manski, 1997 and 2000) and therefore negative average effects may not apply for all strata of the population. As Heckman, Lalonde, and Smith (1999) point out, negative mean impacts may be acceptable if most participants benefit from participation. Abandoning the 'common effect' assumption of treatment effects and identifying the individuals who gain from the programmes is an obvious opportunity to improve their future efficiency. If we are able to identify the personal characteristics which are responsible for the effect heterogeneity in individual impacts, we can use this knowledge for a better future allocation of individuals to programmes. A

[1] Other purposes of JCS, for example the relief of the stock of unemployed in regions with great imbalances of the labour market, are secondary only and will not be evaluated here.

[2] This is also a common finding in the recent evaluation literature of ALMP programmes in Europe. Whereas ALMP were seen as a reasonable opportunity to reduce and avoid unemployment for a long time, the international experiences with the implemented programmes show a mixed picture. The majority of programmes seem to be ineffective in terms of their goals. As the overviews by Martin and Grubb (2001) for OECD countries and Calmfors, Forslund, and Hemström (2002) for Sweden clarify, ALMP are in their present design and implementation not able to achieve a lasting reduction of unemployment.

good example is a situation where we find e.g. that a certain programme works for older participants but does not work for younger participants at all. If in the past more younger individuals have been allocated to the programme, the average effect of the programme may have been negative. Knowing the sources of effect heterogeneity would have helped to achieve a better allocation.

Our evaluation focuses on two main issues: First, we analyse if individuals gain on average from participation. To do so, we use matching methods to estimate the average treatment effect on the treated. Thereby we take gender-specific and regional differences into account. Since the average effects may not apply to all strata of the population, we examine different sources of effect heterogeneity in a second step. We start with a selection of special problem-groups of the labour market like long-term unemployed or individuals without professional training and estimate their treatment effects separately. After that, we construct a simple indicator, which we call target score, based on the individual's number of disadvantages on the labour market, to analyse whether programme effects differ corresponding to the individual labour market obstacles. If programmes are tailored to the needs of the most disadvantaged, one would expect stronger effects for persons with a higher target score. Finally, we use the estimated participation probability to answer the question, whether a higher participation probability correlates with a higher programme effect.

We use data on participants who started their JCS programme in February 2000, and on a comparison group of nonparticipants who were eligible for participation at the end of January 2000, but did not participate in February. We observe the employment status of our sample until December 2002, i.e. almost three years after programmes have started.

The paper is organised as follows: In the following section we briefly review some stylised facts of ALMP and JCS in Germany. We present the data used in section three and introduce the econometric methodology in section four. In section five we discuss the results for the main population. After that, we present the results of the target approaches. Finally, section seven concludes.

2 Some Stylised Facts of Active Labour Market Policy and Job Creation Schemes in Germany

The legal basis for ALMP in Germany is the Social Code III. ALMP are part of the employment promotion and primarily aim at the permanent (re-)integration of unemployed persons into regular employment. According to Social Code III, employment promotion should help to achieve the balancing of labour demand and supply. Therefore, unemployment should be circumvented by an efficient filling of vacancies and the increase of the individual employment chances due to an upgrade of the worker's human capital.

Although ALMP have a long tradition in Germany, their importance increased after the German Re-Unification in 1990. Especially in the eastern part, ALMP were implemented on a large scale to cushion the strong employment reduction in the first years of the transition process. During the last decade two major instruments characterised German ALMP: First, VT programmes that aim at a qualification transfer to circumvent and solve structural problems on the labour market. Second, JCS whose main purpose is to stabilise and qualify unemployed workers for later re-integration into regular employment, but which are also used to relieve tense labour market situations in regions with high unemployment rates.

Promotion of JCS[3] can be authorised if they support activities which are of value for society and additional in nature. Furthermore, individuals have to be placed, whose last chance to stabilise and qualify for later re-integration into regular employment is participation in these schemes. Additional in nature means that the activities could not be executed without the subsidy. Measures with a predominantly commercial purpose have been excluded explicitly up to January 2002. The majority of activities is conducted in the public and non-commercial sector. Financial support for JCS is obtained as a wage subsidy to the employer. Even though JCS should be co-financed measures where between 30 and 75 percent of the costs are subsidies by the FEA and the rest is paid by the supporting institution, exceptions can be made in the direction of a higher subsidy-quota (up to 100 percent). The legal requirements for individuals to enter JCS have been relaxed by the Social Code III amendment (Job-AQTIV-Gesetz) in January 2002. Before that time, potential participants had to be long-term unemployed (more than one year) or unemployed for at least six months within the last twelve months. Furthermore, they had to fulfil the conditions for the entitlement of unemployment compensation. In addition, the local placement officers were allowed to place up to five percent of the allocated individuals who did not meet these conditions (Five-Percent-Quota). Further exceptions are made for young unemployed (under 25 years) without professional training, short-term unemployed (with at least three months of unemployment) placed as tutors, and disabled who could be stabilised or qualified.[4] The subsidy is in general paid for 12 months, but may be extended up to 24 months or even 36 months under special circumstances. Participants are allowed to do a practical training for up to 40 percent of the time and a VT for up to 20 percent. Together, both types of training should not exceed 50 percent of the programme duration. Priority should be given to projects which enhance the chances for permanent jobs, support structural improvement in social or environmen-

[3] The legal basis for JCS is §§ 260–271, 416 Social Code III. They have been the second most important instrument of ALMP in Germany regarding the fiscal volume and the number of promoted individuals. For 2002 the number of promoted individuals in JCS amounts to 112,462 in East and 52,229 in West Germany. These figures correspond to spendings from 1,639.5 million euro in East and 693.5 million euro in West Germany.

[4] With the 2002 amendment, unemployed individuals whose only occupation opportunity is participation in JCS can be placed in programmes independently of the preceding unemployment duration. In addition, the Five-Percent-Quota was augmented up to ten percent.

tal services or aim at the integration of extremely hard-to-place individuals. Participation in JCS results from placement by the local labour office. Unemployed individuals who cannot be integrated into regular employment or do not fit the conditions for another instrument of ALMP may be offered a place in a JCS. The responsible caseworker may cancel a running programme at any time if the participant can be placed into regular employment. If an unemployed person rejects the JCS offer or if a participant denies a career counselling by the placement officer, the labour office can stop the payment of unemployment benefits for up to twelve weeks.

3 Data Set

The data used for the empirical analysis contain information on all participants who were placed in a JCS in February 2000, and on a comparison group of unemployed persons who were eligible for participation in January 2000, but did not enter those schemes in February. Information on nonparticipants and participants were merged from several sources of the FEA. Central source for the information derived on participants is a prototype version of the programme participants master data set ('Maßnahme-Teilnehmer-Gesamtdatei', MTG). This data set includes information from the job-seekers data base ('Bewerberangebotsdatei', BewA), an adjusted version of this data set for statistical purposes (ST4) and the particular information of subsidised employment programmes (ST11TN). For this reason, the MTG contains a large number of attributes to describe individual aspects on the one hand and on the other hand provides a reasonable basis for the construction of the comparison group. The included attributes can be split into four classes: socio-demographic and qualification information, labour market history and particular programme information.[5] The information on the comparison group is derived from the BewA with the additional attributes of the ST4. Therefore, almost all characteristics in the analysis for the comparison as well as for the treatment group originate from the same data sources (see Appendix B for more details). The information is completed by a characterisation of the regional labour market situation by a classification of similar and comparable labour office districts (see Blien *et al.* (2004) and appendix C).[6]

For the outcome variable we use information from the Employment Statistics Register ('Beschäftigtenstatistik', BSt), which includes information on the total population of persons who are registered in the social security system. These are employees and participants of several ALMP programmes, but

[5] The final version of the MTG includes information on all ALMP programmes of the FEA.

[6] The value of good data is an essential building block for a valid evaluation. As for example Heckman, Ichimura, Smith, and Todd (1998) point out, having access to a geographically-matched comparison group administered the same questionnaire as programme participants matters in deriving effective non-experimental estimators of programme impacts.

no self-employed or pensioners. We define only regular employment as a success, whereas all kinds of subsidised employment or participations in ALMP programmes are defined as a failure. While this definition might conflict with the institutional setting, it reflects the economic point of view to measure the integration ability of JCS into non-subsidised employment.[7] To identify spells of regular employment without further promotion, we use the excerpted information of the final version of the MTG on the individual's time spent in ALMP programmes. We observe the labour market outcome for the participating and nonparticipating group until December 2002. Our analysis in the following parts refer to this last month of the observation period, i.e. all employment effects of JCS are estimated for December 2002, that is 35 months after programmes have started. We exclude information on participants in Berlin.[8] Our final sample consists of 11,151 participants and 219,622 nonparticipants. Previous empirical findings have shown that the effects of JCS differ with respect to region and gender (see Hujer, Caliendo, and Thomsen, 2004). Therefore, we separate our analysis by these characteristics, i.e. we separately estimate the effects for men and women in East and West Germany.

4 Econometric Methodology

Estimation of treatment effects based on non-experimental data requires consideration of some identifying issues. As we want to compare participation in one specific programme with nonparticipation, we can use the potential outcome framework with two potential outcomes Y^1 (individual receives treatment) and Y^0 (individual does not receive treatment). The actually observed outcome for any individual i can be written as: $Y_i = Y_i^1 \cdot D_i + (1 - D_i) \cdot Y_i^0$, where $D \in \{0, 1\}$ is a binary treatment indicator. The treatment effect for each individual i is the difference between her potential outcomes $\Delta_i = Y_i^1 - Y_i^0$. Since one of the outcomes is unobservable for each individual, there is no opportunity to calculate individual effects directly. Thus, we have to concentrate on population averages of gains from treatment. A common evaluation parameter is the average treatment effect on the treated (ATT), which focusses explicitly on the effects of those for whom the programme is actually intended. It is given by:

$$\Delta_{ATT} = E(\Delta \mid D = 1) = E(Y^1 \mid D = 1) - E(Y^0 \mid D = 1). \tag{1}$$

Given equation (1), the problem of selection bias is straightforward to see, since the second term on the right hand side of equation (1) is unobservable.

[7] Only the first programme participation is evaluated, any participation in later programmes is viewed as an outcome of the first treatment and is defined as a failure.

[8] The special situation of the labour market in the capital city requires a separate evaluation of the integration effects of JCS into regular employment. The small number of participants aggravates the interpretation of the results.

If the condition $E(Y^0 \mid D = 1) = E(Y^0 \mid D = 0)$ holds, we can use the nonparticipants as an adequate control group. However, this identifying assumption is likely to hold only in randomised experiments. Consequently, estimating the ATT by the difference between the subpopulation means of participants $E(Y^1 \mid D = 1)$ and nonparticipants $E(Y^0 \mid D = 0)$ will lead to a selection bias, which may be caused by observable (e.g. age, skill differences) or unobservable factors (e.g. motivation). For both cases different estimation strategies are available.[9] If we are willing to assume that all relevant attributes for selection are observable, the matching estimator is an appealing choice. It is based on the idea that if individuals are similar conditional on all relevant variables, further differences in the labour market outcome between participants and nonparticipants result from the programme only.[10] It is well known that matching on X can become hazardous when X is of high dimension ('curse of dimensionality'). To deal with this dimensionality problem, Rosenbaum and Rubin (1983) suggest the use of balancing scores $b(X)$, i.e. functions of the relevant observed covariates X such that the conditional distribution of X given $b(X)$ is independent of the assignment to treatment. For participants and nonparticipants with the same balancing score, the distributions of the covariates X are the same, i.e. they are balanced across the groups. The propensity score $P(X)$, i.e. the probability of participating in a programme, is one possible balancing score. It summarises the information of the observed covariates X into a single index function. Rosenbaum and Rubin (1983) show that if treatment assignment is strongly ignorable given X, it is also strongly ignorable given any balancing score. Since we focus on ATT, it is sufficient to assume that (in the notation of Dawid (1979)):

$$Y^0 \text{ II } D|P(X), \tag{2}$$

where II denotes independence. Similar to randomisation in a classical experiment, matching balances the distributions of all relevant[11] pre-treatment characteristics X in the treatment and comparison group, and thus achieves independence between the potential outcomes and the assignment to treatment. Hence, if the mean exists, $E(Y^0 \mid P(X), D = 1) = E(Y^0 \mid P(X), D = 0) = E(Y^0 \mid P(X))$ and the missing counterfactual mean can be constructed from the outcomes of nonparticipants. In order for both sides of the equations to be well defined simultaneously for all $P(X)$, it is usually additionally assumed that

$$Pr(D = 1 \mid X) < 1 \tag{3}$$

for all X. This implies that the support of X is equal in both groups, i.e. $S = \text{Support}(X|D = 1) = \text{Support}(X|D = 0)$. These assumptions are sufficient for identification of (1), because the moments of the distribution of Y^1 for the treated are directly estimable.

[9] See for example Heckman *et al.* (1999), Angrist and Krueger (1999) or Blundell and Costa-Dias (2002) .

[10] See Imbens (2004) or Smith and Todd (2005) for a recent review regarding matching methods.

[11] Relevant variables are all those covariates that jointly determine assignment to treatment and the potential outcomes.

Several matching methods have been suggested in the literature. Good overviews can be found in Heckman *et al.* (1998) and Smith and Todd (2005). The choice of the matching method usually involves a trade-off between matching quality and variance. First, one has to decide on how many nonparticipating individuals to match to a single treated individual. Nearest-neighbour (NN) matching only uses the participant and its closest neighbour. Therefore it minimises the bias but might also involve an efficiency loss, since a large number of close neigbours is disregarded. Kernel-based matching on the other hand uses more nonparticipants for each participant thereby reducing the variance but possibly increasing the bias. Finally, using the same nonparticipating individual more than once (NN matching with replacement) may possibly improve the matching quality, but increases the variance. We have also tested the sensitivity of the results with respect to different matching methods. It turns out that the results are not sensitive to the choice of the matching estimator and therefore we will use and present only one matching strategy, namely nearest-neighbour (NN) matching without replacement and a caliper of 0.02 (see appendix A for details).

5 Empirical Impacts of Job Creation Schemes

5.1 Estimating the Propensity Score

We have estimated the propensity scores using binary logit models with participation as dependent variable. To take account for regional heterogeneity and to allow for gender-specific interaction effects, we have estimated separate models for men and women in East and West Germany.[12] Several model specifications have been tested for the selection of variables to be included in the model. Our final specification contains explanatory variables like age, marital status, the number of children, nationality and health restrictions that describe the sociodemographic background of individuals. Furthermore, qualification is included by characteristics like professional training, the occupational group, the professional rank and previous work experience. The influence of the individual labour market history is given by the unemployment duration, the number of (successless) placement propositions, the duration of the last occupation, the last contact to the personal caseworker, whether the person is an aspirant for vocational rehabilitation, existing placement restraints due to health restrictions and information on an ALMP participation in the past. The regional context is considered by using the classification of the FEA for comparable labour office districts (see Appendix C). Table 1 presents the estimation results for the

[12] We have also estimated the propensity scores for the two regions using dummy variables for sex. However, using the results of the two estimations ignores possible gender-specific interaction effects and the fact, that the coefficients in the estimation differ in their significance and magnitude. This leads to a worse matching quality in the sense that the balancing of covariates after matching is reduced, i.e. the standardised bias (see below) is higher.

participation probability in JCS for the four main groups. Additionally, the number of observations in the four participating and nonparticipating groups are included.

It becomes obvious that allocation differs by regions. The coefficients of the sociodemographic variables show that the participation probability of men in West Germany decreases with age, while in East Germany older men and women are more likely to participate. This indicates the slightly different purpose of the programmes in East and West Germany. Especially in East Germany, JCS function as a relief for the labour market and are used as a bridge to retirement. Furthermore, it has to be noted that German nationals are more likely to participate than foreigners. This may be due to the fact that other measures of ALMP (e.g. language courses) are preferred for foreigners. Regardless of region, health restrictions increase the individual participation probability. This finding indicates an allocation according to the legal basis.

The coefficients for the qualification characteristics emphasise gender specific differences in the allocation. A higher qualification increases the participation probability in both regions for women, whereas the coefficients are insignificant for higher qualified men. The positive coefficients may be seen as an indication that for higher qualified women it is even harder to return to regular employment, so they are willing to participate in a JCS to finish unemployment. As expected, work experience reduces the participation probability of all groups. Work experience is in general an important criterion for placement into regular employment. The finding indicates that experienced workers have other opportunities on the labour market. Since unemployment duration is an eligibility criterion for participation, its influence is of major importance. We included unemployment duration in three categories, up to 13 weeks, between 13 weeks and one year, and for more than one year. As expected, participation probability increases with unemployment duration.

< Include table 1 about here. >

The number of (successless) placement propositions is an indicator for bad labour market opportunities and the coefficient affirms allocation according to the law. A last interesting point to note is that placement restrictions annotated by the caseworker harm the participation probability. This is somewhat surprising, because JCS should even be offered to these groups.

The coefficients for the regional context are in reference to the labour office districts with the best labour market environment. More severe labour market conditions correlate with a decrease in the participation probabilities in both parts. For men in East Germany, living in labour office districts with average labour market opportunities bears the clearest reduction of participation probability, while analogously for West German women and men living in labour office districts dominated by large cities with

an above average unemployment shows the strongest decrease. The better the labour market conditions in the respective labour office district, the more likely are the unemployed persons to participate.

5.2 Matching Quality and First Results

Quality of Propensity Score Estimation and Matching Before we present the results, we first have to check the quality of our propensity score estimation and second, the success of the matching procedure in balancing the covariates between treatment and comparison group.

Our model specification for the propensity score estimation was based on specification tests to identify the relevant variables.[13] One simple method to validate the ability of a good prediction is the computation of hit-rates, i.e. the proportion of persons with a correct prediction of their status (participation and nonparticipation). As becomes obvious from table 2, these hit-rates lie between 70.6 percent for men and 75.7 percent for women in West Germany. For East Germany, the hit-rates are 74.2 for men and 72.2 percent for women. This implies a quite accurate underlying model. However, the aim of propensity score matching is not to maximise the hit-rate, but to balance the covariates between treatment and comparison groups. Since we do not condition directly on all covariates but on the propensity score, we have to check the ability of the matching procedure to balance the relevant covariates. We do so by comparing the absolute bias between the respective participating and nonparticipating groups before and after matching took place. One suitable indicator to assess the distance in the marginal distributions of the X-variables is the standardised bias (SB) suggested by Rosenbaum and Rubin (1985). For each covariate X it is defined as a percentage of the quotient between the difference of the sample means in the treated and (matched) comparison subsamples and the square root of the average of the sample variances in both groups. The SB before and after matching are given by

$$SB_{\text{before}} = 100 \cdot \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5 \cdot (V_1(X) + V_0(X))}}, \qquad SB_{\text{after}} = 100 \cdot \frac{(\overline{X}_{1M} - \overline{X}_{0M})}{\sqrt{0.5 \cdot (V_{1M}(X) + V_{0M}(X))}}, \tag{4}$$

where X_1 (V_1) is the mean (variance) in the treated group before matching and X_0 (V_0) the analogue for the comparison group. X_{1M} (V_{1M}) and $X_{0M}(V_{0M})$ are the corresponding values after matching. This is a common approach used in many evaluation studies, e.g. by Sianesi (2004). To abbreviate the documentation, we present only the means of the SB before and after matching for the four main groups (Table 2). While the mean SB lies between 10.83 and 14.62 percent before matching, it reduces to 1.60 to 3.20 percent after matching.

[13] See Caliendo (2005) for an overview regarding such specification tests and other issues concerning the implementation of matching estimators.

< Include table 2 about here. >

Sianesi (2004) additionally suggests to re-estimate the propensity score on the matched sample, that is only on participants and matched nonparticipants and compare the pseudo-R^2's before and after matching. The pseudo-R^2 indicates how well the regressors X explain the participation probability. After matching there should be no systematic differences in the distribution of the covariates between both groups. Therefore, the pseudo-R^2 after matching should be fairly low. As the results from Table 2 show, this is true for our estimation. The results of the F-tests (with degrees of freedom in brackets) point in the same direction indicating a joint influence before, and no joint influence after matching.

First Results All estimated effects in the later sections of this paper correspond to December 2002, the last month of our observation period. We are aware of the fact that consideration of only this month bears some shortcomings for a valuable interpretation of the programme effects. Since December 2002 is almost three years after programmes have started, and with respect to the average duration of programmes of twelve months for the majority of participants almost two years after the programmes have ended, the estimates do not provide evidence for the development of the employment effects. Hence, to give an idea of the time path of the effects, figure 1 presents the estimated effects for the four main groups between February 2000 and December 2002. For the further analysis, our focus will be on the mid-term effects of job creation schemes.

< Include figure 1 about here. >

At the beginning of the observation period, the programme effect is expected to be overlayed by so-called 'locking-in'-effects (van Ours, 2004) due to a reduced search intensity of the participants. This reduced search intensity is plausible for participants, since they are occupied by participation and spend less time on job search. Thus, a valid interpretation of the programme effects on the employment rates should start after the majority of participants has left the programmes, i.e. after twelve months. Since the purpose of JCS is to stabilise and qualify unemployed persons for the re-integration into regular employment, we would expect increasing employment rates after the programmes have ended. We find these 'locking-in'-effects for all groups (see figure 1). After this initial fall there is a clear rising tendency for the groups in West Germany and a moderate rising tendency for the groups in East Germany. For the smallest group, women in West Germany, there is the strongest rise in the employment rates with significant positive effects at the end of the observation period in December 2002. The effects for men in West Germany are also rising, but the effects are insignificant at the end of our observation period, i.e.

an increase in the employment chances by participation cannot be established. While the effects in West Germany are clearly rising, we find a stepwise increase with relatively constant levels over one-year-periods in East Germany. Besides that, the 'locking-in'-effects during the first year after programmes start are not as strong as in the West. This finding can be interpreted as an indication of worse outside options for the nonparticipants.

Although the effects show a rising tendency for all groups, a significant increase of the employment rates due to participation can only be stated for women in West Germany, who have a significant positive effect of 4.6 percentage points in December 2002. For men in West Germany we do not find any significant effects in December 2002, whereas men in East Germany have a significant negative effect of -2.9 percentage points. For women in East Germany the effect is slightly better but still significantly negative at -1.4 percentage points. So it seems that JCS rather decrease than increase the employment prospects of participants. Of course, due to the strong 'locking-in'-effects, the starting position for the participants is on average lower than for nonparticipants. However, since we observe the outcomes until 35 months after start of the programmes and almost two years after the majority of the individuals has left the programmes, a successful programme should overcompensate for this initial fall.

6 Targeting

Clearly, as already mentioned, one possible explanation for the discouraging results in the previous section may be the poor quality of the programmes in conjunction with stigma- and 'locking-in'-effects. Another possible cause might be an inefficient allocation of participants. Since programme effects are heterogeneous (Manski, 1997 and 2000), the average effects depicted in the above section must not apply to all strata of the population. Negative mean impacts may be acceptable if the majority of participants gains from the programme (Heckman *et al.*, 1999). Abandoning the 'common effect' assumption of treatment effects and identifying the individuals that benefit from the programmes is an obvious opportunity to improve the future efficiency of ALMP. If we are able to identify the individual characteristics, which are responsible for the effect heterogeneity in individual impacts, we can use this knowledge to suggest allocation rules for a better future allotment of programme participants.

The potential improvement of allocation mechanisms is a much discussed topic in the recent evaluation literature (see for example Lechner and Smith, 2005, Frölich, Lechner, and Steiger, 2003, and Frölich, 2001). An optimal allocation should guarantee the best results according to the underlying programme goal, where two goals - efficiency and equity - can be distinguished. If the goal is efficiency, programmes target at the maximisation of the impacts of the outcome of interest. If the goal is equity, treatment is

administered to those individuals identified as 'neediest', i.e. for example those individuals with the lowest predicted re-employment probabilities (Plesca and Smith, 2002). Frölich *et al.* (2003) distinguish between statistical and non-statistical allocation mechanisms.

Caseworker discretion is the most common non-statistical allocation mechanism. Potential programme participants are interviewed by their personal caseworker and allocation to programmes depends on the caseworker's evaluation of the unemployed person's capabilities, the individual's interests and the availability of slots in the particular programmes. The crucial feature of the caseworker allocation mechanism for an optimal allocation of unemployed persons to programmes is the knowledge of the characteristics of the unemployed person, the situation on the local labour market and the programme providers as well as the professional expertise of the caseworker (Lechner and Smith, 2005). There are only a few studies that examine the quality of caseworker allocation in Europe. Frölich (2001) analyses the effects of caseworker allocation in Sweden; Lechner and Smith (2005) and Frölich *et al.* (2003) evaluate the effectiveness of Swiss caseworkers in comparison to a simulated targeting system. The results indicate that caseworker allocation lacks the ability to achieve the expected programme goals. Reasons for the ineffectiveness of the caseworker allocation may be lack of knowledge of caseworkers regarding the effectiveness of certain programmes. Caseworkers have to build expectations about impacts of programmes on a very uncertain basis. Additionally, the broad variety of available programmes makes it difficult to select an optimal strategy for a specific person (Frölich *et al.*, 2003). Another issue concerns possible 'cream-skimming'. The experiences from the Job Training Partnership Act (JTPA) showed that tying the funding to the performance of local programmes as measured by job placement rates creates the incentive to serve the most able applicants, without regarding how much different groups might have benefited from programmes (see for example Bell and Orr, 2002).

Statistical allocation mechanisms avoid these possible problems by relying on some model indicating the individual gains of participation in a specific programme. Up to now, there is no consistent classification of statistical treatment rules. OECD (2002) defines 'profiling' as 'a procedure where a numerical score, calculated on the basis of multivariate information, determines the referral of a job-seeker to further employment services'. Based on this definition, we will present three approaches to identify potential sources of effect heterogeneity, which could be used – if successful – for a better targeting in future. At first, we will select target groups with disadvantages on the labour market, e.g. long-term unemployed persons. In a second step, we will use these definitions and build a simple index that we call 'target score'. The target score simply sums up the number of individual disadvantages. If programmes are tailored to the needs of the most disadvantaged on the labour market, we would expect higher impacts for persons with higher target scores. For the evaluation of the effects in the target groups and

for the target scores, we estimate separate propensity scores for each group and category considered.[14] Finally, we test whether the effects differ corresponding to different participation probabilities. To do so, we stratify our sample in 20 sub-samples along the propensity score of the participants and use a stratification matching estimator.

6.1 Effects for Selected Target Groups

Identifying groups of participants who benefit from programmes is a central purpose of programme evaluation. Recent evaluation studies of JCS in Germany (Hujer *et al.*, 2004) and experiences from abroad (Martin and Grubb, 2001) recommend a tighter targeting of programmes to individuals with disadvantages on the labour market. Selecting persons that are supposed to have a below average employability is a sensible first approach to identify possible effect heterogeneity due to personal characteristics. Several groups of individuals who should be promoted predominantly are defined in Social Code III. These are long-term unemployed persons, individuals with health restrictions or persons who aspire for vocational rehabilitation.[15] Further target groups are young ond older unemployed as well as workers without any professional training. In addition, JCS should be particularly applied to individuals with special placement restrictions.

Our selection is oriented on these legal definitions. We estimate the effects for participants younger than 25 years and for participants older than 50 years respectively. Further groups are long-term (more than one year when programmes start) unemployed persons, individuals with special placement restrictions due to health restrictions and aspirants for vocational rehabilitation. Additionally, we select four groups of persons who are hard-to-place. The first group contains individuals with more than five (unsuccessful) placement propositions by the local labour offices, the second group are persons who have already participated in an ALMP programme before unemployment. Group three contains individuals without professional training and the last group are people without any work experience.

< Include table 3 about here. >

Table 3 contains the shares of individuals in each of the selected groups differentiated by treatment status. For most of the groups, the results show significant differences of the shares between treatment and comparison group. Thus, one can assume that these characteristics affect the allocation decision to some extent. Surprisingly, long-term unemployment (more than 52 weeks) which is expected to

[14] The results of these estimations and the standardised biases before and after matching are available on request by the authors.

[15] This are especially persons who are no more able to work in their profession due to health restrictions, and therefore should receive a promotion for vocational rehabilitation.

be an important selection criterion (in accordance to the law), differs only for men in East Germany. Additionally, the share of aspirants for vocational rehabilitation of this group and the proportions of men and women without work experience in the region are approximately equal for participants and nonparticipants. This shows once again the different purpose of JCS in East and West Germany.

Further notable findings are the different proportions of participants between the regions. While the share of younger unemployed (below 25 years) in West Germany is clearly larger in the participants' group, the situation in East Germany is the other way round. Older unemployed are more likely to participate here. These differences have to be interpreted in light of the different labour market situation in East and West Germany and the consequently different purpose of JCS in both regions. Placing a larger share of young unemployed into programmes in West Germany complies to the law that postulates stabilising efforts for later re-integration. In East Germany, JCS are used to relieve the labour market and therefore older unemployed are more likely to participate than younger ones.

Besides the age differences, it has to be mentioned that persons with a larger number of placement propositions or who have participated in an ALMP programme before unemployment are more frequent in the participating group. This confirms our expectations that the number of successless placement propositions directly indicates the placement difficulties. Furthermore, earlier participation may identify so-called 'programme careerists'. These are persons who are assigned to ALMP programmes subsequently with short spells of unemployment between the single measures.

Table 4 presents the employment effects in December 2002 for these nine groups with further distinction for gender and region as above. It becomes obvious that programme effects are heterogeneous across the selected groups. Whereas the results for the four main groups showed insignificant effects for men in West Germany, men and women in East Germany suffered from participation and women in West Germany benefited on average from programmes. Consideration of the effects for the selected groups of male participants in West Germany shows, that the effects are for almost all groups insignificant, too, but with one exception. The group of long-term unemployed men benefits from participation and has an employment rate which is 5.03 percentage points higher compared to the rate of matched nonparticipants in December 2002. The female counterparts in that region are the only group who benefited on average from participation. With regard to the results in table 4, it becomes clear that this finding does not hold for all groups. While three groups clearly gain from participation, i.e. older unemployed (12.67 percentage points), long-term unemployed (11.25 percentage points), and hard-to-place women indicated by the number of placement propositions (7.79 percentage points), the others do not experience any enhancement of the employment chances. Anyhow, the three significant effects are above the effects for the whole sample of females in West Germany.

< Include table 4 about here.>

Turning to the estimates for the East German groups reveals a quite similar picture. Again, most of the estimates are statistically insignificant and participants do neither suffer nor benefit form participation at all in December 2002. Whereas the results for men in this region have been significantly negative on average, this finding is confirmed by the result of one group only, namely participants who have participated in an ALMP programme before (-3.36 percentage points). All other estimates do not show significant differences to the nonparticipants' outcomes. Regarding women, we find long-term unemployed to benefit from participation (2.45 percentage points). No significant differences in the employment rates can be established for the remaining groups.

Together with the results for the West German groups, especially long-term unemployed participants seem to benefit from programmes (except for men in East Germany). This finding is somewhat satisfactory since JCS are especially arranged for this group. Although the effects refer to one single month only, the results are plausible. Since occupations in JCS have to be additional in nature, i.e. they are not allowed to compete with regular jobs to avoid substitution effects, the qualifying elements for market-competitive jobs have to be assumed to be negligible. Thus, the stabilising elements in the design of JCS (to keep in touch with the labour market) may be more important for this group. Furthermore, participation in JCS comes along with a stigmatisation of the participant if potential employers suspect a reduced productivity. However, long-term unemployment is a stigma itself and hence the additional stigma-effect of JCS might be of minor importance. To the contrary, for these groups participation must be seen as an indicator for individual motivation to change the personal situation. Hence, the stigma-effect of JCS may be more important for short-term unemployed and younger persons.

Summarising the findings for the selected target groups leads us to three recommendations. First, due to the unsatisfactory results for most of the groups where no differences in the employment rates between participants and nonparticipants could be established, JCS have to be reviewed critically in terms of their goals. Nevertheless, they are no complete failure for some participants as the results especially for long-term unemployed indicate. Second, a tighter targeting of programmes to persons for whom the possible negative aspects (like stigmatisation, lack of human capital transfer etc.) are only of minor importance for the individual labour market prospects, should help to increase programme efficiency. Third, since long-term unemployed persons are not the majority of unemployed in Germany, the number of promotions should be reduced significantly. JCS are definitely sensible for the most disadvantaged workers, but no means for reducing unemployment permanently for all unemployed persons.

6.2 Effects for Target Groups Using Target Scores

The results in the previous section show that JCS do not work for most of the analysed groups. Nevertheless, as the estimates are significantly positive especially for the most disadvantaged persons, the long-term unemployed, the question arises whether a higher number of explicit labour market disadvantages correlates with gains from participation. To answer this question, we build a simple index which we call 'target score' as the sum of the individual number of disadvantages from section 6.1. Without any particular weighting, each disadvantage adds one point to the target score. Persons who do not belong to any of the categories in section 6.1, have a target score of zero. The maximum level is eight, since the categories for the age groups are mutually exclusive. For example, if an individual is below 25 years old and has no professional training, she is assigned a target score of two. If an individual belongs to three of the target groups, the target score is three, and so on. Due to a small number of individuals with a target score of more than five, we summarise these persons in one group, i.e. target score five (and more); the other categories refer to the actual number of disadvantages. We estimate the programme effect on the employment rates in December 2002 within each category of the target score.

< Include table 5 about here.>

If programmes are tailored to the needs of the most disadvantaged and if a higher target score indicates higher need of assistance than we would expect better outcomes for higher scores. The estimates of the effects in December 2002 are given in table 5. Ignoring the significance of the estimates at first, the results show non-negative effects for all groups in West Germany with a target score greater or equal three. For the lower target score groups, the picture is not that homogeneous. While men in West Germany with a target score of one or two are harmed, women with the same score seem to benefit. In East Germany, groups with a target score of less than three have reduced employment rates in December 2002. For women with more disadvantages there seems to be no effect, while for men the estimates tend to be negative except for a target score of three.

The tendencies in the results for West Germany support the hypothesis that a higher target score coincides with a higher need of assistance and a better fit of programmes for those groups, but a clear statement is hampered due to the insignificant estimates for most groups. It is self-evident that our construction of the target score is very simple and is not guided by some strong theory. First, the different targeting criteria are included with the same weights and clearly may not have the same importance for the individual employability. Second, the selection of groups is incomplete. There are other characteristics that increase or decrease the individual employability. Third, the construction of the target score

leaves room for further effect heterogeneity. The target score just notes the number of single targets, but does not identify clear sets of disadvantages where participation improves the employability.

Unfortunately, considering the significance of the results shows that our assumption cannot be empirically approved. For each of the West German groups only one estimate for the higher target scores is significant. For men with a target score of five, i.e. five or more disadvantage criteria on the labour market, the employment rates increase by 14.49 percentage points after participation, for women with a target score of four by 11.76 percentage points. For the other groups the estimates are insignificant, i.e. no clear increase or decrease in the employment rates by participation can be established. The estimates for East Germany show a slightly different picture. The results illustrate that allocating individuals without any of the selected targeting criteria and therefore a target score of zero to programmes, reduces the employment rates in December 2002 by 10.14 for men and 8.12 percentage points for women. Analogously to the finding for West Germany, there are no further significant results. Since our construction of the target score is very simple, it has to be reviewed, whether incorporation of further selection criteria and/or a different weighting of the single targets may improve the significance of the estimates. Although the estimates are unsatisfying yet, the usage of the target score provides some practical utility to identify possible sources for effect heterogeneity.

6.3 Targeting by Stratification Matching

The estimated propensity score reflects the individual participation probability conditional on the relevant observable characteristics. If allocation to the programme is target-oriented, a higher participation probability should also correlate with a higher impact of treatment. Clearly, this argument only holds, if the programmes are tailored according to the needs of the participants. If this is not the case, i.e. if the programmes have the same effects for all participants, individuals with low participation probabilities may benefit more since a high participation probability can to some extent be interpreted as an indicator for bad labour market prospects. Furthermore, an interesting opportunity arises, if the empirical evidence supports a positive relationship between a higher participation probability and a higher impact of treatment. If this is the case, the estimated participation probability could be used as an allocation instrument, i.e. persons with higher propensity score values should be primarily allocated to programmes.

An intuitively appealing method to check this hypothesis is stratification matching, also known as blocking or subclassification. The idea is to divide the sample of participants and nonparticipants conditional on the propensity score into several strata. Within these strata, participants and nonparticipants should have approximately the same probability of treatment. The average treatment effect is estimated within each stratum as if random assignment holds. Estimation of the treatment effect for the treated is

carried out by weighting the within-strata average treatment effects by the number of treated units. Stratification matching can be interpreted as a crude form of non-parametric regression where the unknown function is approximated by a step function with fixed jump points (Imbens, 2004). An important issue in employing this estimator is to make sure that the covariates are balanced within each stratum. The distribution among the treatment and comparison group should be balanced, if the true propensity score is constant. Comparison of the distribution of covariates of both groups within strata yields a possibility to assess the adequacy of the statistical model.

To check our hypothesis whether a higher participation probability correlates with a higher programme impact, we divide our samples into twenty subclasses each. This division is based on the estimated propensity scores of the participants.[16] Therefore, we have the same number of participants in each stratum, but different numbers of nonparticipants with approximately the same scores as the participants. Individuals with the lowest participation probabilities are placed in stratum 1, persons with the highest participation probabilities are placed in stratum 20. It can be seen that this stratification leaves meaningful numbers of observations in each stratum except for women in West Germany.

< Include tables 6 and 7 about here. >

The estimated treatment effects for each stratum are presented in table 6 for East Germany and in table 7 for West Germany. The effectiveness of the programmes can be estimated by comparing the employment rates of participants and nonparticipants in December 2002 given by $E(Y_1)$ and $E(Y_0)$ in the tables. The average treatment effect within each stratum, i.e. the difference of the mean outcomes of the participants and the nonparticipants is also given (Δ). The last lines of the tables provide the average treatment effect on the treated. Obviously, these effects are similar to those estimated with the NN-matching estimators in section 5. In addition to the mean outcomes and the effects, the tables also present the results of the hypothesis testing of equal propensity scores in the treatment and comparison group. We tested the null hypothesis (H_0) that the difference of the mean propensity scores in both groups is zero. Therefore, the alternative hypothesis (H_A) imposes inequality of the propensity score. The p-values of the H_A are given in the tables; if we reject the hypothesis due to a larger value than 0.05, we assume equality of the propensity scores and therefore balancing of the covariates among both groups.[17]

[16] Due to the large number of observations in our samples, using the whole range of the propensity scores of participants and nonparticipants leads to a skewed stratification. Hence, we refer to the propensity scores of the participants only to reduce this skewness. The choice of twenty strata for each of the four groups emerged from balancing tests of the propensity score among treated and comparison persons using a smaller number of blocks.

[17] We also checked the balancing property of stratification by comparing the means of the incorporated variables in the logit models for participants and nonparticipants within each stratum as suggested by Rosenbaum and Rubin (1983). The results for selected variables are available on request by the authors.

The results of the hypothesis tests show that the division into twenty strata provides approximately equal propensity scores for most groups. The equality is hampered only for the groups at the borders of the propensity score range. For men in West Germany, strata 1, 5, 7 and 20 are imbalanced, for women in the same region so are strata 1, 17 and 19. In East Germany the strata with lower participation probabilities are imbalanced. For women the propensity scores are not balanced in 1 and 2, for men in 1 and 3, but also in stratum 19. Although we find significant treatment effects for several strata, these findings do not assist our hypothesis. Taking a look at the results for East Germany (table 6), we find that for the first four strata (except for women in stratum 1) allocation of persons with a low participation probability has a tendential negative influence on the employment chances in December 2002. For men in this region, this tendency is stable for participants up to stratum 14; from stratum 15 onwards the direction of the effects changes to positive. For women we could not establish a clear distinction, since most of the effects are insignificant. For participants in West Germany (table 7) our hypothesis cannot be empirically approved either. One can somehow see that higher participation probabilities correlate with higher impacts, but these findings may be inconsistent as the balancing tests above show. It seems that the participation probability is no adequate measure for effect heterogeneity here and successful integration into regular employment depends on different compositions of the individual characteristics than selection into programmes.

7 Conclusion

Previous empirical studies of JCS in Germany have shown that the average effects for participating individuals are negative. Whereas this inefficiency may be due to the poor quality of programmes, it may be also driven by an inefficient allocation of potential participants to programmes. Allocation of individuals into programmes in Germany is accomplished by caseworker discretion. On the one hand, a positive aspect of this mechanism is that decisions are based on personal contact. On the other hand, since ALMP consist of very different programmes, caseworkers may lack knowledge regarding programme impacts. Since this problem is not specific to Germany, the topic of a potential improvement of allocation mechanisms has become important in recent literature. Broadly, two categories can be distinguished: Non-statistical allocation mechanisms like caseworker discretion and statistical allocation mechanisms called profiling or targeting. Since statistical allocation systems are not introduced in the German labour market yet, there is no empirical evidence for their effectiveness.

In this paper we estimate the average treatment effects for men and women in East and West Germany participating in JCS. Following that we use three strategies to identify possible effect heterogeneity. We

use data on all participants, who started a JCS in February 2000, and on nonparticipants from January 2000, who were eligible to participate, but did not enter those schemes in February. The employment effects of JCS are evaluated in December 2002. The results show positive effects for women in West Germany and negative effects for men and women in East Germany. Men in West Germany do neither suffer nor benefit from participation.

For the three approaches used to analyse effect heterogeneity, we select target groups with disadvantages on the labour market oriented by the definition of the legal basis in a first step. Our findings show that JCS do neither harm nor improve the labour market chances for most of the groups. Exceptions are long-term unemployed men in West Germany, long-term unemployed women in both regions, older women and women who are hard-to-place in West Germany, who benefit from participation. Given these results and remembering that (re-)integration into regular employment is the main purpose, it has to be recommended that JCS should be targeted to those benefiting groups and should not be used on large scale. In a second step, we use these definitions to build up a simple indicator (target score) as the sum of the individual number of disadvantages. If programmes are tailored to the needs of the more disadvantaged persons on the labour market, we expect positive impacts for groups with a higher score. Unfortunately, most of the estimates are insignificant and although the expected tendency is observable, one has to be cautious with interpretation. Finally, we implement stratification matching to analyse if a higher participation probability also correlates with higher impacts. No clear picture can be revealed. The estimated participation probability is no adequate measure for effect heterogeneity here and successful integration into regular employment is determined by different compositions of the individual attributes than selection into programmes. Even though the results could not confirm some of our hypotheses, they show that heterogeneity in treatment effects is an important topic which has to be considered more accurately in further research. We have also shown that this might be a way to improve efficiency of ALMP and hence to allocate scarce resources more effectively.

Acknowledgements The authors thank Richard Blundell and Jeff Smith for valuable comments. The paper has also benefited from fruitful discussions at the annual meeting of the Italian Association of Labour Economists (AIEL) in 2004. Financial support of the Institute for Employment Research (IAB) within the project 'Effects of Job Creation and Structural Adjustment Schemes' is gratefully acknowledged. A supplementary appendix is available on request by the authors and can be downloaded from http://www.wiwi.uni-frankfurt.de/professoren/hujer/papers/identifying_anc_appendix.pdf. An earlier version of this paper has circulated as 'Effect Heterogeneity, Profiling and Targeting - How Can We Improve the Efficiency of Labour Market Policies?'. All remaining errors are our own.

References

ANGRIST, J. D., AND A. B. KRUEGER (1999): "Empirical Strategies in Labor Economics", in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, pp. 1277–1366. Elsevier Science B.V.

BELL, S., AND L. ORR (2002): "Screening (and Creaming?) Applicants to Job Training Programs: The AFDC Homemaker-Home Health Aide Demonstrations," *Labour Economics*, 9, 279–301.

BLIEN, U., F. HIRSCHENAUER, M. ARENDT, H. J. BRAUN, D.-M. GUNST, S. KILCIOGLU, H. KLEINSCHMIDT, M. MUSATI, H. ROSS, D. VOLLKOMMER, AND J. WEIN (2004): "Typisierung von Bezirken der Agenturen der Arbeit," *Zeitschrift für Arbeitsmarktforschung*, 37(2), 146–175.

BLUNDELL, R., AND M. COSTA-DIAS (2002): "Alternative Approaches to Evaluation in Empirical Microeconomics," *Portugese Economic Journal*, 1, 91–115.

CALIENDO, M. (2005): "Microeconometric Evaluation of Labour Market Policies," *Lecture Notes in Economics and Mathematical Systems*, No. 568, Springer-Verlag, Berlin, forthcoming.

CALIENDO, M., R. HUJER, AND S. L. THOMSEN (2005a): "The Employment Effects of Job Creation Schemes in Germany - A Microeconometric Evaluation," Discussion Paper No. 1512, IZA.

——— (2005b): "Sectoral Heterogeneity in the Employment Effects of Job Creation Schemes in Germany," Working Paper, J.W. Goethe-University.

CALMFORS, L., A. FORSLUND, AND M. HEMSTRÖM (2002): "Does Active Labour Market Policy Work? Lessons From The Swedish Experience," CESifo Working Paper No. 675(4), CESifo.

COCHRAN, W., AND D. RUBIN (1973): "Controlling Bias in Observational Studies," *Sankyha*, 35, 417–446.

DAWID, A. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B*, 41, 1–31.

FRÖLICH, M. (2001): "Treatment Choice Based on Semiparametric Evaluation Methods," Discussion Paper No. 2001-16, Department of Economics, University of St. Gallen.

FRÖLICH, M., M. LECHNER, AND H. STEIGER (2003): "Statistically Assisted Programme Selection - International Experiences and Potential Benefits for Switzerland," *Swiss Journal of Economics and Statistics*, 139, 311–331.

HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.

HECKMAN, J., R. LALONDE, AND J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics Vol.III*, ed. by O. Ashen- felter, and D. Card, pp. 1865–097. Elsevier, Amsterdam.

HUJER, R., M. CALIENDO, AND S. L. THOMSEN (2004): "New Evidence on the Effects of Job Creation Schemes in Germany - A Matching Approach with Threefold Heterogeneity," *Research in Economics*, 58(4), 257–302.

IMBENS, G. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics*, 86(1), 4–29.

LECHNER, M., AND J. SMITH (2005): "What is the Value Added by Caseworkers," *Labour Economics*, forthcoming.

LEUVEN, E., AND B. SIANESI (2003): "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing," Statistical Software Components s432001, Boston College Department of Economics, revised 30.April 2004.

MANSKI, C. (1997): "The Mixing Problem in Programme Evaluation," *The Review of Economic Studies*, 64, 537–553.

——— (2000): "Using Studies of Treatment Response to Inform Treatment Choice in Heterogeneous Populations," Technical Working Paper, No. 263, NBER.

MARTIN, P., AND D. GRUBB (2001): "What works and for whom: A review of OECD countries experiences with active labour market policies," *Swedish Economic Policy Review*, 8, 9–56.

OECD (2002): Employment Outlook. Paris.

PLESCA, M., AND J. SMITH (2002): "Rules Versus Discretion in Social Programs: Empirical Evidence on Profiling in Employment and Training Programs," Working Paper, University of Maryland.

ROSENBAUM, P., AND D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–50.

——— (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistican*, 39, 33–38.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–145.

RUBIN, D. (1974): "Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies," *Journal of Educational Psychology*, 66, 688–701.

SIANESI, B. (2004): "An Evaluation of the Active Labour Market Programmes in Sweden," *The Review of Economics and Statistics*, 86(1), 133–155.

SMITH, J., AND P. TODD (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125(1/2), 305–354.

VAN OURS, J. (2004): "The Locking-In Effect of Subsidized Jobs," *Journal of Comparative Economics*, 32(1), 37–52.

Tables

Tab. 1: ESTIMATION RESULTS OF THE LOGIT-MODELS FOR THE PROPENSITY SCORE

	West Germany				East Germany			
	Men		Women		Men		Women	
Variable	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Constant	**-1.1739**	0.2731	**-3.1254**	0.4533	**-5.7880**	0.3659	**-8.0021**	0.3944
Socio-Demographic Variables								
Age	**-0.0599**	0.0145	-0.0067	0.0235	**0.0901**	0.0141	**0.1702**	0.0136
Age(squared)	*0.0004*	0.0002	-0.0003	0.0003	**-0.0008**	0.0002	**-0.0019**	0.0002
Married	**-0.1676**	0.0612	**-0.4483**	0.0761	**0.2683**	0.0506	**0.1145**	0.0344
Number of children	*0.0653*	0.0281	-0.0183	0.0439	-0.0335	0.0266	-0.0238	0.0184
German	**0.4402**	0.0683	*0.2825*	0.1211	**0.6284**	0.1966	**0.7082**	0.2432
Health restrictions								
No health restrictions	Ref.		Ref.		Ref.		Ref.	
Acc. DoR,[1] 80% and over	**0.9160**	0.1826	**1.3404**	0.2578	*0.5491*	0.2758	**1.1375**	0.2442
Acc. DoR, 50% to under 80%	**0.8052**	0.1267	**0.6433**	0.1978	**0.4991**	0.1270	**0.6032**	0.1242
Acc. DoR, 30% to under 50%	**1.1190**	0.3658	**1.9871**	0.4246	**0.5691**	0.1925	**0.7999**	0.1954
Acc. DoR, 30% to under 50%, no equalis.[2]	0.2757	0.1570	0.0651	0.2685	-0.0708	0.1721	-0.0725	0.1826
Other health restrictions	-0.0472	0.0892	-0.0751	0.1390	**-0.1918**	0.0716	*-0.1422*	0.0608
Qualification Variables								
Professional training								
Without compl. prof. training, no CSE	Ref.		Ref.		Ref.		Ref.	
Without compl. prof. training, with CSE	**-0.3364**	0.0622	0.2294	0.1334	0.1015	0.0823	**0.3428**	0.0865
Industrial training	**-0.6738**	0.0692	-0.0808	0.1399	*-0.1777*	0.0748	**0.3315**	0.0820
Full-time vocational school	**-0.7639**	0.2685	-0.0734	0.2432	-0.3223	0.2594	**0.8588**	0.1384
Technical school	-0.0987	0.1756	**0.7183**	0.1927	0.2227	0.1231	**1.0166**	0.0977
Polytechnic	0.3534	0.2009	**1.4983**	0.2144	-0.0135	0.2058	**1.0388**	0.1794
College, University	0.2399	0.1577	**1.0221**	0.1869	0.0810	0.1354	**0.9004**	0.1272
Occupational group								
Plant cultivation, breeding, fishery	*0.2222*	0.0927	0.2628	0.2501	0.0092	0.0828	**0.2370**	0.0670
Mining, mineral extraction	-0.5605	0.4657	–	–	-0.7494	0.5154	–	–
Manufacturing	Ref.		Ref.		Ref.		Ref.	
Technical professions	**-0.5810**	0.1544	-0.1609	0.2605	-0.1954	0.0999	**0.2149**	0.0819
Service professions	**-0.3077**	0.0544	**0.3167**	0.0995	**-0.1739**	0.0478	0.0127	0.0406
Other professions	0.1023	0.1533	0.3933	0.2628	**-1.1891**	0.2170	**-1.2092**	0.2860
Professional rank								
Worker, not skilled worker	Ref.		Ref.		Ref.		Ref.	
Worker, skilled worker	**-0.5499**	0.0982	-0.1637	0.1944	**-0.1811**	0.0597	0.0657	0.0525
White-collar worker, simple occupations	0.0163	0.1152	0.1490	0.1256	0.1809	0.1067	**0.2197**	0.0605
White-collar worker, advanced occupations	0.0877	0.1536	**0.5131**	0.1624	-0.2838	0.1662	-0.0404	0.1215
Other	-0.0112	0.0563	0.1512	0.1054	0.0345	0.0528	*0.1004*	0.0437
Qualification (with work experience)	**-0.3397**	0.0745	**-0.3139**	0.1017	**-0.2279**	0.0695	*-0.1175*	0.0527
Career Variables								
Duration of last employment (months)	**-0.0046**	0.0005	**-0.0033**	0.0007	**-0.0038**	0.0004	**-0.0028**	0.0003
Duration of unemployment (weeks)								
Up to 13 weeks	Ref.		Ref.		Ref.		Ref.	
Between 13 and 52 weeks	**0.2055**	0.0616	0.0698	0.0889	**0.4673**	0.0561	**0.2509**	0.0511
More than 52 weeks	**0.3087**	0.0678	0.0888	0.0974	**0.4498**	0.0599	**0.1694**	0.0509
Number of placement propositions	**0.0494**	0.0028	**0.0530**	0.0042	**0.0610**	0.0030	**0.0919**	0.0031
Last contact to job center (weeks)	-0.0013	0.0125	**0.0520**	0.0177	**-0.1204**	0.0114	**-0.0644**	0.0085
Rehabilitation attendant	-0.1533	0.1185	0.0696	0.2039	**0.2958**	0.0939	0.1535	0.1024
Placement restrictions	**-0.3396**	0.0989	-0.2654	0.1546	**-0.3164**	0.0870	**-0.3000**	0.0825
Programme before unemployment								
No further education or programme	Ref.		Ref.		Ref.		Ref.	
Further education compl., cont. education	**0.2292**	0.0801	**0.5301**	0.1043	**0.4830**	0.0628	**0.5263**	0.0422
Further education compl., voc. adjustment	**0.6479**	0.2286	0.4613	0.4466	**0.6545**	0.0893	**0.5634**	0.0746
Job-preparative measure	-0.4764	1.0285	**2.6387**	0.5245	**1.1431**	0.4289	0.3364	0.5250
Job creation scheme	**2.1463**	0.0777	**3.0671**	0.1141	**1.7272**	0.0546	**1.5382**	0.0418
Rehabilitation measure	-0.0929	0.2706	**0.9368**	0.3406	0.4232	0.2273	0.3780	0.2720
Regional Context Variables[3]								
Cluster Ia	–	–	–	–	-0.1040	0.1291	0.1421	0.1238
Cluster Ib	–	–	–	–	*-0.3077*	0.1248	-0.0242	0.1210
Cluster Ic	–	–	–	–	*-0.2838*	0.1361	-0.1841	0.1292
Cluster II	**-0.2225**	0.0730	**-0.5666**	0.0960	Ref.		Ref.	
Cluster III	*-0.1841*	0.0722	**-0.4601**	0.0917	–	–	–	–
Cluster IV	-0.0080	0.1002	**-0.4530**	0.1423	–	–	–	–
Cluster V	Ref.		Ref.		–	–	–	–
No. of Part.	2,140		1,052		2,924		5,035	
No. of Nonpart.	44,095		34,227		64,788		76,512	

Bold letters indicate significance at the 1% level. *Italic* letters refer to the 5% level. Ref. denotes the reference category. – not included in the estimation/ no observation.

[1] DoR = Degree of restriction

[2] People with accepted degree of restriction, but no equalisation to other persons with the same DoR.

[3] See appendix C for further information.

Tab. 2: SOME QUALITY INDICATORS

	West Germany		East-Germany	
	Men	**Women**	**Men**	**Women**
Before Matching				
Observations[1]	46,235	35,271	67,712	81,505
Hit-Rate[2]	70.6	75.7	74.2	72.2
Pseudo R^2	0.1389	0.1775	0.1225	0.1144
F-Test	2,406.8 (41)	1,679.4 (40)	2,951.3 (41)	4,323.3 (40)
Mean of Standardised Bias (in percent)[3]	14.62	16.08	12.01	10.83
After Matching				
Observations[4]	4,246	1,960	5,846	10,054
Pseudo-R^2	0.006	0.009	0.004	0.003
F-Test	38.0 (41)	23.4 (40)	35.3 (41)	39.2 (40)
Mean of Standardised Bias (in percent)[3]	2.51	3.20	1.78	1.60

[1] Observations are the sum of participating and nonparticipating individuals.

[2] *Hit-rates* are computed as follows: If the estimated propensity score is larger than the sample proportion of persons taking treatment, i.e. $\hat{P}(X) > \bar{P}$, observations are classified as '1'. If $\hat{P}(X) \leq \bar{P}$ observations are classified as '0'.

[3] *Mean of Standardised Bias* calculated as mean of the single characteristics' standardised biases.

[4] Since we apply NN-matching without replacement and a caliper of 0.02 the number of treated individuals is reduced after matching by observations falling outside the region of common support. The numbers of the treated individuals can be calculated by dividing the number of observations by 2.

Tab. 3: DESCRIPTIVE STATISTICS FOR THE SELECTED TARGET GROUPS (PARTICIPANTS AND NONPARTICIPANTS)

West Germany	**Men**		**Women**	
	Part.	Nonpart.	Part.	Nonpart.
TARGET GROUP	SHARES IN PERCENT[1]			
Age $<$ 25 years	21.40	9.30	17.30	7.14
Age $>$ 50 years	16.12	37.27	15.30	35.21
Without professional training	62.62	49.12	45.25	49.94
Without work experience	12.76	7.44	15.11	7.44
Long-term unemployed (more than 52 weeks)[2]	39.16*	40.79*	39.16*	42.16*
More than 5 placement propositions	49.21	21.21	42.49	17.05
Vocational rehabilitation[3]	5.19	6.27	4.18	3.11
Placement restrictions[4]	16.54	21.58	14.07	17.51
Participation in ALMP before unemployment	28.55	10.05	33.17	8.86

East Germany	**Men**		**Women**	
	Part.	Nonpart.	Part.	Nonpart.
TARGET GROUP	SHARES IN PERCENT[1]			
Age $<$ 25 years	8.21	13.49	2.94	6.36
Age $>$ 50 years	38.06	31.05	30.69	35.71
Without professional training	28.63	23.10	22.26	25.85
Without work experience	10.02*	10.84*	9.89*	10.38*
Long-term unemployed (more than 52 weeks)[2]	37.55	30.75	49.45	48.89
More than 5 placement propositions	41.24	17.87	37.28	15.32
Vocational rehabilitation[3]	7.46*	7.48*	3.10	4.60
Placement restrictions[4]	13.47	16.16	7.47	11.92
Participation in ALMP before unemployment	47.16	17.08	57.28	27.85

* Denotes approximate equality of shares between treatment and comparison group (5% significance level).

[1] Shares are computed with respect to the number of participating/nonparticipating individuals in the according main group.

[2] Unemployment duration for participants and nonparticipants at end of January 2000.

[3] Persons in vocational rehabilitation are unable to work in their profession any longer and have to be qualified for a new profession.

[4] Placement restrictions refer to the assessment of the caseworker that health restrictions of the job-seeker reduce the number the job opportunities.

Tab. 4: EFFECTS FOR SELECTED TARGET GROUPS IN DECEMBER 2002

West Germany	**Men**			**Women**		
TARGET GROUP	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
Age $<$ 25 years	-0.0276	0.0326	440	-0.0679	0.0573	161
Age $>$ 50 years	0.0262	0.0241	344	**0.1267**	0.0562	159
Without professional training	-0.0046	0.0169	1,323	0.0425	0.0297	451
Without work experience	-0.0040	0.0414	256	-0.0703	0.0595	128
Long-term unemployed (more than 52 weeks)	**0.0503**	0.0169	832	**0.1125**	0.0326	403
More than 5 placement propositions	0.0300	0.0176	1,039	**0.0779**	0.0302	400
Vocational rehabilitation[1]	0.0300	0.0603	106	0.0571	0.0845	36
Placement restrictions[2]	0.0153	0.0287	335	0.1026	0.0562	130
Participation in ALMP before unemployment	-0.0323	0.0217	594	0.0541	0.0313	279

East Germany	**Men**			**Women**		
TARGET GROUP	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
Age $<$ 25 years	-0.0437	0.0503	240	0.0278	0.0589	148
Age $>$ 50 years	-0.0130	0.0079	1,109	-0.0020	0.0093	1,529
Without professional training	0.0120	0.0161	833	-0.0215	0.0156	1,119
Without work experience	0.0069	0.0349	292	0.0225	0.0220	495
Long-term unemployed (more than 52 weeks)	-0.0018	0.0093	1,097	**0.0245**	0.0080	2,487
More than 5 placement propositions	-0.0264	0.0145	1,201	-0.0054	0.0108	1,869
Vocational rehabilitation[1]	-0.0140	0.0369	217	-0.0068	0.0418	154
Placement restrictions[2]	0.0189	0.0254	394	-0.0166	0.0217	368
Participation in ALMP before unemployment	**-0.0336**	0.0114	1,378	-0.0028	0.0079	2,877

Effects are estimated using 1-NN matching without replacement and caliper of 0.02. **Bold** letters indicate significance on a 5% level. Standard errors calculated by bootstrapping with 50 replications.

[1] Persons in vocational rehabilitation are no more able to work in their profession and have to be qualified for a new profession.

[2] Placement restrictions refer to the assessment of the caseworker that health restrictions of the job-seeker reduce the number the job opportunities.

Tab. 5: ESTIMATED EFFECTS FOR THE TARGET SCORES[1] IN DECEMBER 2002

West Germany	**Men**			**Women**		
TARGET SCORE	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
0	0.0182	0.0850	55	-0.0133	0.0789	76
1	-0.0138	0.0363	295	0.0518	0.0401	208
2	-0.0180	0.0212	740	0.0316	0.0474	305
3	0.0256	0.0261	652	0.0276	0.0339	257
4	0.0199	0.0331	274	**0.1176**	0.0527	100
5 and more	**0.1449**	0.0591	84	0.0455	0.1033	32

East Germany	**Men**			**Women**		
TARGET SCORE	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
0	**-0.1014**	0.0484	141	**-0.0812**	0.0333	271
1	-0.0293	0.0198	581	-0.0064	0.0118	1,090
2	-0.0225	0.0155	937	-0.0093	0.0110	1,754
3	0.0013	0.0191	821	0.0112	0.0103	1,289
4	-0.0161	0.0213	322	0.0062	0.0159	508
5 and more	-0.0532	0.0448	94	0.0000	0.0393	106

Effects are estimated using 1-NN matching without replacement and caliper of 0.02. **Bold** letters indicate significance on a 5% level. Standard errors calculated by bootstrapping with 50 replications.

[1] Target Scores are calculated as the sum of the number of individual disadvantages from the selection of the target groups.

Tab. 6: RESULTS FOR STRATIFICATION MATCHING IN EAST GERMANY

Strata		Men				Women			
		No. of Obs.	p-value for H_A[1]	$E(Y_1)$, $E(Y_0)$	Δ	No. of Obs.	p-value for H_A[1]	$E(Y_1)$, $E(Y_0)$	Δ
1	Participants	146	0.0001	0.1781	-0.0585	251	0.0002	0.1355	0.0134
	Nonparticipants	16,171		0.2366		18,980		0.1221	
2	Participants	146	0.9303	0.1781	-0.0666	252	0.0168	0.1032	-0.0235
	Nonparticipants	9,532		0.2446		11,309		0.1267	
3	Participants	146	0.0218	0.1233	**-0.0897**	252	0.1633	0.1190	-0.0267
	Nonparticipants	7,657		0.2130		7,396		0.1458	
4	Participants	146	0.3283	0.1575	-0.0347	252	0.1581	0.0913	-0.0568
	Nonparticipants	5,529		0.1923		5,641		0.1480	
5	Participants	147	0.0537	0.0816	*-0.0772*	251	0.2593	0.1633	0.0137
	Nonparticipants	4,432		0.1588		5,098		0.1497	
6	Participants	146	0.2077	0.1233	-0.0245	252	0.1555	0.1111	-0.0245
	Nonparticipants	3,093		0.1478		4,298		0.1356	
7	Participants	146	0.9609	0.0822	-0.0476	252	0.5875	0.1627	0.0178
	Nonparticipants	2,727		0.1298		3,852		0.1449	
8	Participants	146	0.4523	0.0685	-0.0497	252	0.3221	0.1071	*-0.0494*
	Nonparticipants	2,640		0.1182		2,804		0.1566	
9	Participants	146	0.5098	0.1027	-0.0201	251	0.2600	0.1036	*-0.0609*
	Nonparticipants	2,116		0.1229		2,785		0.1645	
10	Participants	147	0.7602	0.1020	-0.0173	252	0.1690	0.0952	-0.0423
	Nonparticipants	2,037		0.1193		2,276		0.1375	
11	Participants	146	0.4703	0.0616	-0.0440	252	0.3124	0.1190	-0.0192
	Nonparticipants	1,448		0.1057		2,228		0.1382	
12	Participants	146	0.4960	0.0959	-0.0165	252	0.9466	0.1508	0.0133
	Nonparticipants	1,592		0.1124		1,665		0.1375	
13	Participants	146	0.3424	0.0411	**-0.0729**	251	0.9627	0.1036	-0.0151
	Nonparticipants	1,132		0.1140		1,651		0.1187	
14	Participants	146	0.8348	0.0616	-0.0373	252	0.0541	0.1310	0.0371
	Nonparticipants	980		0.0990		1,471		0.0938	
15	Participants	147	0.7724	0.1224	0.0296	252	0.2967	0.0992	0.0126
	Nonparticipants	948		0.0928		1,143		0.0866	
16	Participants	146	0.8285	0.0890	0.0152	252	0.9422	0.1071	0.0164
	Nonparticipants	772		0.0738		1,124		0.0907	
17	Participants	146	0.9521	0.0753	0.0253	251	0.3790	0.0797	-0.0071
	Nonparticipants	600		0.0500		910		0.0868	
18	Participants	146	0.4996	0.0822	*0.0403*	252	0.6872	0.0913	-0.0129
	Nonparticipants	645		0.0419		749		0.1041	
19	Participants	146	0.0053	0.0548	0.0193	252	0.7600	0.1349	0.0192
	Nonparticipants	479		0.0355		648		0.1157	
20	Participants	147	0.6655	0.0748	0.0244	252	0.6248	0.1548	0.0267
	Nonparticipants	258		0.0504		442		0.1281	
ATT:					**-0.0251**				-0.0084

Bold letters indicate significance at the 1% level. *Italic* letters refer to the 5% level. Subgroups are constructed using the estimated propensity score of the participants from the logit model reported in Table 1.

[1] Testing $H_0 : P(Z, D = 1) - P(Z, D = 0) = 0$. Corresponding H_A: $P(Z, D = 1) - P(Z, D = 0) \neq 0$ in stratum.

Tab. 7: RESULTS FOR STRATIFICATION MATCHING IN WEST GERMANY

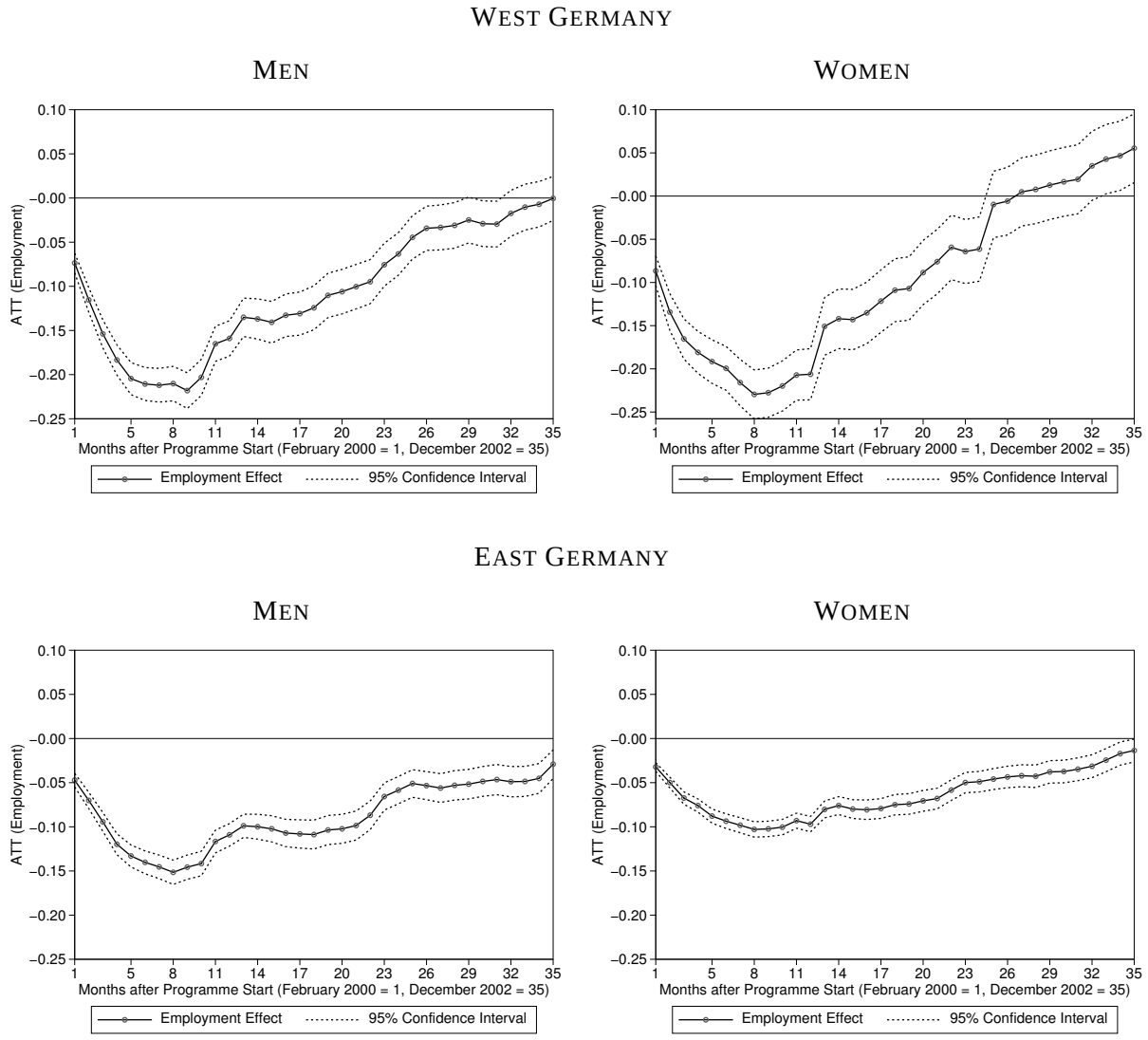
Strata		Men				Women			
		No. of Obs.	p-value for H_A[1]	$E(Y_1)$, $E(Y_0)$	Δ	No. of Obs.	p-value for H_A[1]	$E(Y_1)$, $E(Y_0)$	Δ
1	Participants	107	0.0000	0.1869	*0.0764*	52	0.0005	0.3846	**0.2649**
	Nonparticipants	14,220		0.1105		12,954		0.1197	
2	Participants	107	0.1905	0.1963	-0.0046	53	0.1774	0.3585	*0.1194*
	Nonparticipants	4,913		0.2009		4,119		0.2391	
3	Participants	107	0.2521	0.2336	0.0034	52	0.5364	0.3077	0.0201
	Nonparticipants	4,065		0.2303		2,754		0.2876	
4	Participants	107	0.8130	0.2150	-0.0355	53	0.7943	0.3962	0.1169
	Nonparticipants	3,522		0.2504		2,782		0.2793	
5	Participants	107	0.0430	0.2617	0.0278	53	0.6186	0.3019	-0.0110
	Nonparticipants	2,403		0.2339		1,742		0.3129	
6	Participants	107	0.5197	0.1682	*-0.0998*	52	0.7633	0.2692	-0.0341
	Nonparticipants	2,384		0.2680		1,556		0.3033	
7	Participants	107	0.0045	0.2056	-0.0484	53	0.9023	0.3585	0.0370
	Nonparticipants	2,331		0.2540		1,347		0.3215	
8	Participants	107	0.4353	0.2056	-0.0593	52	0.6411	0.2885	-0.0307
	Nonparticipants	1,748		0.2649		1,366		0.3192	
9	Participants	107	0.2616	0.2336	-0.0364	53	0.9991	0.2830	-0.0481
	Nonparticipants	1,533		0.2701		1,214		0.3311	
10	Participants	107	0.3627	0.2804	0.0005	53	0.6523	0.3396	-0.0242
	Nonparticipants	1,229		0.2799		841		0.3639	
11	Participants	107	0.1798	0.1963	-0.0831	52	0.8903	0.3269	-0.0184
	Nonparticipants	1,049		0.2793		611		0.3453	
12	Participants	107	0.5893	0.2991	0.0343	53	0.3965	0.2830	-0.0608
	Nonparticipants	929		0.2648		733		0.3438	
13	Participants	107	0.6554	0.2617	-0.0073	52	0.2097	0.3846	-0.0102
	Nonparticipants	751		0.2690		623		0.3949	
14	Participants	107	0.3683	0.2617	0.0088	53	0.3294	0.3208	-0.0260
	Nonparticipants	684		0.2529		571		0.3468	
15	Participants	107	0.5013	0.2056	-0.0667	53	0.2556	0.4340	0.1185
	Nonparticipants	661		0.2723		447		0.3154	
16	Participants	107	0.4412	0.2430	0.0452	52	0.0935	0.3077	0.0171
	Nonparticipants	551		0.1978		265		0.2906	
17	Participants	107	0.8646	0.1402	-0.0332	53	0.0282	0.3208	0.0615
	Nonparticipants	473		0.1734		108		0.2593	
18	Participants	107	0.0955	0.1308	0.0122	52	0.7560	0.3654	**0.1987**
	Nonparticipants	295		0.1186		78		0.1667	
19	Participants	107	0.4283	0.2617	**0.1413**	53	0.0389	0.3396	*0.1682*
	Nonparticipants	191		0.1204		70		0.1714	
20	Participants	107	0.0038	0.2710	**0.1606**	53	0.1637	0.3585	**0.2715**
	Nonparticipants	163		0.1104		38		0.0870	
ATT:					0.0018				**0.0565**

Bold letters indicate significance at the 1% level. *Italic* letters refer to the 5% level. Subgroups are constructed using the estimated propensity score of the participants from the logit model reported in Table 1.

[1] Testing $H_0 : P(Z, D = 1) - P(Z, D = 0) = 0$. Corresponding H_A: $P(Z, D = 1) - P(Z, D = 0) \neq 0$ in stratum.

Figures

Fig. 1: ATT (EMPLOYMENT) BETWEEN FEBRUARY 2000 AND DECEMBER 2002



A The Matching Estimator

The general form of matching estimators is given by:

$$\Delta^{MAT} = \frac{1}{N_1} \sum_{i \in I_1} [Y_i^1 - \sum_{j \in I_0} W_{N_0}(i,j) Y_j^0], \tag{5}$$

where N_0 is the number of observations in the comparison group I_0 and N_1 is the number of observations in the treatment group I_1. We estimate the effect of treatment for each treated observation $i \in I_1$ in the treatment group, by contrasting her outcome with treatment with a weighted average of comparison group observations $j \in I_0$. Matching estimators differ in the weights attached to the members of the comparison group (Heckman *et al.*, 1998), where $W_{N_0}(i,j)$ is the weight placed on the j-th individual from the comparison group in constructing the counterfactual for the i-th individual of the treatment group. The weights always satisfy $\sum_j W_{N_0}(i,j) = 1, \forall i$, that is the total weight of all comparisons sums up to one for each treated individual. Define a neighbourhood $C(P_i)$ for each i in the participant sample and denote as neighbours for i those nonparticipants $j \in I_0$ for whom $P_j \in C(P_i)$. Individuals matched to i are those people in the set A_i where $A_i = \{j \in I_0 | P_j \in C(P_i)\}$. Nearest neighbour (NN) matching sets

$$C^{NN}(P_i) = \min_j \|P_i - P_j\|, j \in N_0, \tag{6}$$

where $\|(.)\|$ is obtained through a distance metric. Doing so, the nonparticipant with the value of P_j that is closest to P_i is selected as the match, therefore:

$$W^{NN}_{N_0 N_1}(i,j) = \begin{cases} 1 & \text{if } \|P_i - P_j\| = \min_j \|P_i - P_j\| \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

Several variants of NN matching are proposed, e.g. NN matching 'with' and 'without replacement'. In the former case a nonparticipating individual can be used more than once as a match, whereas in the latter case it is considered only once. Matching with replacement involves a trade-off between bias and variance. If we allow replacement, the average quality of the matching will increase and the bias will decrease. NN matching faces the risk of bad matches if the closest neighbour is far away. This can be avoided by imposing a tolerance on the maximum distance $\|P_i - P_j\|$ allowed. This form of matching, caliper matching (Cochran and Rubin, 1973), imposes the condition:

$$\|P_i - P_j\| < \epsilon, j \in N_0, \tag{8}$$

where ϵ is a pre-specified level of tolerance. The weights for caliper matching (CM) are given by:

$$W^{CM}(i,j) = \begin{cases} 1 & \text{if } \|P_i - P_j\| = \min_j \|P_i - P_j\| \wedge \|P_i - P_j\| < \epsilon \\ 0 & \text{else} \end{cases}. \tag{9}$$

Treated observations for whom no matches within the neighbourhood $C(P_i) = \{P_j | \|P_i - P_j\| < \epsilon\}$ can be found are excluded from the analysis. Hence, caliper matching is one form of imposing a common support condition.

B Data Sources and Attributes

Table B.1 gives detailed information of the data sources and the included attributes. A selection of these attributes is used to estimate the participation probability.

Tab. B.1: Data Sources and Attributes

	Data Source	Attributes
MTG[1]	BewA and ST4[2]	a) **Socio-demographic**: age, gender, marital status, number of children, nationality, health restrictions b) **Qualification**: graduation, professional training, occupational group, position in last occupation, work experience, appraisal of qualification by the placement officer c) **Labour market history**: duration of unemployment, duration of last occupation, number of job offers, occupational rehabilitation, programme participation before unemployment
	ST11TN[3]	d) **Programme**: institution that receives subsidy, activity sector, time of qualification and/or practical training during programme, begin and end of programme (payment of the subsidy), entry and leave of the participant, duration of programme

[1] Programme participants master data set (Maßnahme-Teilnehmer-Gesamtdatei, MTG)
[2] Job-seekers data base (Bewerberangebotsdatei, BewA) and adjusted version for statistical purposes (ST4)
[3] Programme participants of subsidised employment data set (ST11TN)

C Regional Context Variables

The classification of the labour office districts was undertaken by a project group of the FEA. The aim of the project was to enhance the comparability of the labour office districts for a more efficient allocation of funds. The 181 labour office districts were split into twelve types of office districts with similar labour market circumstances. The comparability of the office districts is build upon several labour market characteristics. The most important criteria are the underemployment quota and the corrected population density. The underemployment quota is defined as the relation of the sum of unemployed individuals and participants in several ALMP programmes to the sum of all employed persons and these participants. The corrected population density is used to improve the comparability of rural labour office districts with metropolitan and city areas. In addition to that, the vacancy quota describing the relation of all reported vacancies at the labour office, the placement quota, that contains the number of placements to the number of employments, and the quota of people who achieve maintenance allowance in relation to the underemployment quota are used. Furthermore, an indicator for the tertiarisation level built on the number of employed persons in agricultural occupations and an indicator for the seasonal unemployment are considered.

The twelve types of comparable labour office districts can be summarised into five types for strategic purposes. Since almost all labour office districts in East Germany belong to the first of these five strategic types, we use the finer typing of three groups here. For West Germany we use the remaining four types for strategic purposes. Table C.1 presents the classification used in the analysis, containing a short description of the clusters and the number of labour offices in each clusters.

Tab. C.1: Classification of labour office districts in Germany

Cluster	Description	No.
Ia	East German labour office districts with worst labour market conditions	5
Ib	East German labour office districts with bad labour market conditions	23
Ic	East German labour office districts with high unemployment	5
II	Labour office districts dominated by large cities	21
III	West German labour office districts with rural elements, medium-sized industry and average unemployment	63
IV	West German centers with good labour market prospects	10
V	West German labour office districts with the best labour market prospects	47

No. describes the number of labour offices in cluster.
Source: Blien *et al.* (2004)