Effect Heterogeneity – How can we improve the Efficiency of Labour Market Policies?*

Marco Caliendo†, Reinhard Hujer‡ and Stephan L. Thomsen§

†*J.W.Goethe-University, Frankfurt and IZA, Bonn*

‡*J.W.Goethe-University, Frankfurt, IZA, Bonn, ZEW, Mannheim*

§*J.W.Goethe-University, Frankfurt*

Working Paper

Draft: Please do not quote!

September 9, 2004

Abstract Previous empirical studies of job creation schemes in Germany have shown that the average effects for the participating individuals is negative. However, we find that this is not true for all strata of the population. Identifying the personal 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. We use this information to deduce some easy implementable decision rules and discuss these rules in context of the actual allocation mechanisms. Finally, we run a small simulation study and show how the programme results can be enhanced by better allocation mechanisms.

Keywords: Evaluation – Profiling – Targeting – Efficient Allocation

JEL Classification: C13, H43, J68

---

*The authors thank Jeff Smith for valuable comments. Financial support of the Institute for Employment Research (IAB) within the project 'Effects of Job Creation and Structural Adjustment Schemes' is gratefully acknowledged. All remaining errors are our own.

†Marco Caliendo is Research Assistant at the Institute of Statistics and Econometrics, J.W.Goethe-University of Frankfurt and Research Affiliate of the IZA, Bonn, e-mail: caliendo@wiwi.uni-frankfurt.de.

‡Reinhard Hujer is Professor of Statistics and Econometrics at the J.W.Goethe-University of Frankfurt, 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, 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 vocational rehabilitation aspirants. ALMP was first introduced in Germany in the late 1960s. Since then, the labour market experienced several important changes, like the oil price shocks during the 1970s and the growth of the labour market after the German Unification in 1990. The set of programmes was gradually adjusted to these changes. Despite these reforms and large spendings on ALMP, the German labour market is still plagued by high and persistent unemployment. Therefore, the evaluation of ALMP has become a major topic and has been also legally anchored in the reformed legal basis in 1998, the Social Code III. This reform is characterised by the introduction of new instrument, a stronger decentralisation of competences and a more flexible allocation of funds and persons in need.

The requirement of evaluation is handled by the FEA utilising two different approaches. The first approach examines the share of programme participant, who are not registered unemployed six months after the end of the programmes ('Verbleibsquote'). Taking this information as a success criterion is problematic for several reasons. First, consideration of the unemployment status of only one specific month after the end of participation contains no information of the situation in other months after programme. Second, using gross-effects to draw policy relevant conclusions is problematic, because they do not answer the question, what would have happened to the participants, had they not participated. This question can only be answered by net-effects, which contrast the situation of the participants with the counterfactual situation of nonparticipation. Third, the criterion 'not registered unemployed' provides only limited information about the destination of the participants, since further participation, retirements and motherhood are also included. Hence, a valuable success criterion must refer to regular employment.

The latter problem is circumvented by the second measure of the FEA, which uses the proportion of participants, who are regular employed six months after the end of programmes ('Eingliederungsquote'). However, the two other points of criticism apply to this approach as well, making clear, that evaluation of the net-effects is an important task to draw policy relevant conclusions.

In this paper we focus on JCS, which have been one of the major instruments of ALMP in Germany over the last years. We use data of participants in JCS, who started their programmes in February 2000, and of 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 start. JCS 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 integration into the first labour market.[1] Recent empirical studies of job creation schemes for Germany have shown that the average effects for the participating individuals are negative (see for example Hujer, Caliendo, and Thomsen (2003)).[2] The reasons for these findings have to be analysed.

One possible explanation can be the poor quality of the 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) mention, negative mean impact results may be acceptable if most participants gain from the programme. Therefore, abandoning the 'common effect' assumption of treatment effects and identifying the individuals that benefit 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 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 might have been negative. However, knowing the sources of effect heterogeneity might help to do a better allocation job in the future, i.e. assign only older people to the programme in our example.

The aim of this paper is threefold: First, we will estimate the effects of JCS of the total population with respect to gender and region. Second, we will examine several sources of effect heterogeneity. Since we have a very informative dataset, we can use matching methods to estimate programme effects. Finally, we will use the results of the first two steps, to test possible allocation strategies that may improve the programme impacts. The remainder of the paper is organised in three parts. In the first part, we present the effects of JCS in Germany. To do so, we will first review some stylised facts on JCS in Germany, explain the data in use, present the methodological framework of our estimators and discuss the results of the average treatment effects on the treated by comparison of different matching estimators. The second part deals with effect heterogeneity. In a first step, we select target groups of the labour market according to the institutional settings and estimate their effects. Based on the individual number of disadvantages, we build a naive target score in a second step. If programmes are tailored to the needs of the most-disadvantaged, one would expect stronger effects for persons with a higher target score. The third step uses a stratification matching approach to answer the question, whether a higher participation probability correlates with a higher programme effect. In the last part of the paper, we discuss the possible role of the underlying

[1] Other enacted purposes like 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 an important 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 seems to be ineffective in terms of their aimed at 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.

allocation mechanisms for the effects. We start with a description of the contemporary allocation mechanisms with a distinction into non-statistical and statistical methods. This discussion will be followed by an out-of-sample simulation of a statistical model based allocation of participants and the estimation of the potential improvement of the impacts (not finished yet).

Part I. Effects of Job Creation Schemes in Germany

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

The legal basis for active labour market policies (ALMP) in Germany is the Social Code III ('Sozialgesetzbuch III'). The primary objective of ALMP in Germany is the permanent (re-) integration of unemployed people into regular employment. The main purpose of the employment promotion according to the Social Code III is to balance labour demand and supply. 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. Besides those explicit postulations of the legislator for the design of the labour market policy, the evaluation of the success of the instruments is now legally anchored. The analysis of the effects of ALMP is now a focus of labour market research in Germany. The purpose is a more contemporary evaluation of the different instruments, considering aspects like the net-effect on the employment chances for an individual, the identification of macroeconomic effects and cost-benefit analysis.

Although ALMP have a long tradition in Germany, their importance increased after the German Unification in 1990. Especially in Eastern Germany, 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 the German ALMP: First, VT that aim at a qualification transfer to circumvent and solve structural problems on the labour market. Second, JCS that aim at the stabilisation and qualification of unemployed workers for later re-integration into regular employment on the one hand, but also conduct an important relief function for the tense situation on the labour market. The higher emphasis on active labour market policies also reflects the postulations from other OECD countries of that time (see OECD (1994)). But, as recent reviews of the international experiences over the last decade show, active labour market policies did not achieve the expected outcomes.[3] Therefore, several countries, like Sweden and Switzerland, implemented a more 'activating' labour market policy that recommends the use of instruments with higher integration abilities. The effects of German ALMP are also below expectations and are reviewed to be gradually reformed ('Modern Services on the Labour Market', Bundesministerium für Wirtschaft und Arbeit (2003)) to incorporate more 'activating' elements.[4]

[3] See overviews by Martin and Grubb (2001) and Calmfors, Forslund, and Hemström (2002)
[4] Despite the 1998 reform, unemployment has raised over the last years. Therefore, the government proposes new laws

JCS[5] are the second most important instrument of ALMP in Germany in respect of the fiscal volume and the number of promoted individuals, although these figures have decreased in recent years.[6] JCS can be promoted if they support activities which are of value for the society and additional in nature. Furthermore individuals have to be employed, 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; now they can be accomplished with a special permission by the administration board of the local labour office to prevent substitution effects and windfall gains. Besides the social value and the additional benefit of the activities, participants in JCS in the private sector should be from special target groups of the labour market, e.g. young unemployed without professional training, and get educational supervision during occupation. Due to these special requirements the majority of activities is conducted in the public 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% of the costs are subsidies by the FEA and the rest is paid by the supporting institution (public or private legal entities, mainly municipalities), exceptions can be made in the direction of a higher subsidy-quota (up to 100%). The legal requirements for individuals to enter JCS are relaxed by the SGB 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. Additionally they had to fulfil the conditions for the entitlement of unemployment compensation.[7] In addition, the local placement officers were allowed to place up to five percent of the allocated individuals who do 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

called 'Modern Services on the Labour Market'. These laws contain a reshape of unemployment promotion to a more activating labour market policy with a particular emphasis on a personal contribution towards economic integration on the part of the unemployed. A gradually reduce and simplification of the legal regulations together with a greater decentralisation and the expansion of the budgetary competences in new employment offices should make it possible to place stronger emphasis on the regional programmes and integrate the activities of all those involved in economic and labour market policy in 'natural economic areas'. The activating measures in the context of promoting integration should be consistently geared to the needs of the job seekers and the companies in the respective regions. Also the unemployment insurance should be continually developed into what will be known as 'employment insurance'. These changes are on the one hand deserving, on the other hand there is a strong opposition in the population especially in East Germany.

[5] The legal basis for JCS is §§ 260–271, 416 Social Code III.

[6] For 2002 the number of promoted individuals in JCS amounts to 112,462 in East Germany and 52,229 in West Germany. These figures correspond to spendings from 1,639.5 million euro in East Germany and 693.5 million euro in West Germany.

[7] There are two kinds of unemployment compensation in Germany. The first kind are unemployment benefits (UB) that are paid dependent on the preceding duration of employment, the age and if the individual has children. To get UB, an individual must register unemployed at the local labour office, seek for a regular occupation and have worked in regular employment before. The UB amounts to 60% (67%) of the net-wage of the last occupation for unemployed without (with) children. The longest possible UB entitlement is 32 months. After expiration of the UB entitlement, unemployed can gain unemployment assistance (UA) if they are in need of further promotion. In analogy to the UB entitlement, the UA differs dependent on having or not having children. The amount of UA for persons without (with) children is 53% (57%) of the last net-wage. The UA is paid for one year at maximum, but can be prolonged by case-wise revision. For every following year the grants are paid on a p.a. 3% reduced last net-income basis. Participation in a job creation scheme prolongs the entitlement for UB in the same way as regular employment.

of unemployment) placed as tutors, and disabled who could be stabilised or qualified.

With the 2002 amendment, all unemployed individuals can enter a JCS independently of the preceding unemployment duration, but with the restriction that JCS is the only opportunity for occupation. In addition, the Five-Percent-Quota was augmented up to ten percent. The subsidy is normally paid for 12 months, but can be extended up to 24 or even 36 months, if it is followed by regular employment. Participants are allowed to do a practical training up to 40% of the time and a vocational training up to 20%, together no more than 50% of the programme duration. Priority should be given to projects which enhance the chances for permanent jobs, support structural improvement in social or environmental 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 active labour market policy are offered a place. JCS contain occupations in several sectors, like agriculture, office & services and community services. The responsible caseworker can cancel a running programme at any time, if the participant can be placed into regular employment. If an unemployed rejects the offer of a JCS or if a participant denies a career counselling by the placement officer, the labour office can stop the unemployment benefits for up to twelve weeks. Although the legal purpose of JCS is the re-integration of unemployed workers with labour market disadvantages, valuing these schemes in terms of efficiency might mask the strong equity efforts that result from the altruistic design.

3. Data Set and Selected Descriptives

3.1. Data Set

The data used for the empirical analysis contain information on all participants, who were placed in a JCS in February 2000, and a comparison group of nonparticipants, 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 data sets of the FEA. The central source for the information derived for the participants is a prototype version of the programme participants master data set ('Maßnahmeteilnehmergrunddatei', MTG). This data set merges 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). Due to this construction, the MTG contains on the one hand a large number of attributes to describe individuals aspects 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.[8] The information for the comparison group is derived from the BewA with the additional information of the

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

ST4. Therefore, almost all attributes in the analysis for the comparison as well as for the treatment group originate from the same data sets (see Appendix A for more details). The information is completed by a characterisation of the regional context by a classification of similar and comparable labour office districts by a project group of the FEA (see Appendix B).[9]

For the outcome variable we use information from the Employment Statistics Register ('Beschäftigtenstatistik', BSt), which includes information about the total population of all people who are registered in the social security system. These are all regular employed persons and participants of some ALMP programmes, but no self-employed or pensioners. To identify the spells or regular employment without further promotion, we complete the information on the outcome variable with the content of the final version of the MTG, since only regular employment is defined as a success. All kinds of subsidised employments or participations in ALMP 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.[10] We observe the labour market outcome for the treatment and comparison group until December, 2002. Our analyses in the following parts refer to this last month of the observation period. So all employment effects of JCS are estimated for December 2002, that is 35 months after programmes start. We exclude information on participants in Berlin.[11] Our final sample consists of 11,151 participants and 219,622 nonparticipants. Previous empirical findings have shown that the effects of JCS differ with regional and gender differences (Hujer, Caliendo, and Thomsen, 2003). Therefore, we separate our analyses by gender and region, i.e. we estimate separately for men and women in East and West Germany.

3.2. Selected Descriptives

To provide some more details on the participants and the comparison group in use, table 1 presents the numbers of persons and the mean and the median of selected characteristics for the four main groups. As mentioned in section 2., subsidies for JCS are normally paid for up to 12 months. As the figures show, that is also true for our sample. The mean duration of programmes lies between 278.04 days for men in West Germany and 333.27 days for women in East Germany. The median of the duration shows that more than 50% of the participants end their programmes after one year.

Comparing the participants and nonparticipants in terms of selected characteristics shows several notable differences. First, the mean and the median of the age of participants in West Germany are lower than in the comparison group, while this picture is the other way round for East Germany. This finding might indicate a slightly different purpose of JCS in the two regions. Placing on average

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

[10] 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.

[11] 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.

Tab. 1: Descriptive Statistics for Selected Attributes of Participants and Non-Participants

West Germany	**Men**				**Women**			
	Participants		Non-Participants		Participants		Non-Participants	
No. of obs.	2,140		44,095		1,052		34,227	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Programme Duration (days)	278.04	330	–	–	304.88	359	–	–
Unemployment Duration (weeks)[1]	61.90	35	73.04	36	60.05	37	76.44	40
Age (years)	37.21	38	43.22	44	37.82	39	43.33	44
Placement Propositions (number)	7.70	5	3.60	1	6.87	5	2.99	1
	Shares in %							
Work Experience	87.24		92.56		84.89		92.56	
Without Professional Training	62.62		49.12		45.25		49.94	

East Germany	**Men**				**Women**			
	Participants		Non-Participants		Participants		Non-Participants	
No. of obs.	2,924		64,788		5,035		76,512	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Programme Duration (days)	318.78	359	–	–	333.27	360	–	–
Unemployment Duration (weeks)[1]	50.00	38	47.25	27	65.77	52	82.24	50
Age (years)	44.51	47	41.73	42	43.86	45	44.01	45
Placement Propositions (number)	6.06	4	3.01	1	5.42	4	2.77	2
	Shares in %							
Work Experience	89.98		89.16		90.11		89.62	
Without Professional Training	28.63		23.10		22.26		25.85	

[1] Unemployment duration for participants and non-participants at end of January 2000.

younger persons into programmes might aim at enhancing their individual labour market prospects for later re-integration into non-subsidised employment. In East Germany, JCS are still used for older unemployed to relieve the labour market, prolong the entitlement for unemployment benefits of the participants and to function as a bridge to retirement. Second, since unemployed persons should be placed in JCS only if they have no other opportunities on the labour market, taking a look at the number of (unsuccessful) placement propositions by the local labour office shows, that this is true. While in West Germany participants have a median of five placement propositions, nonparticipants have a median of one. In East Germany the differences between participants and nonparticipants are similar. Participants have a median of four placement propositions, for nonparticipants these figures are one for males and two for females. Third, the average unemployment duration differs also between participants and nonparticipants. Considering the median is more informative here. In West Germany participants have a shorter median unemployment duration than the nonparticipants. In East Germany this picture is again the other way round. One last thing to mention are the different shares of persons without professional training. These figures are almost twice as high for West than for East Germany, but one has to bear in mind that almost everybody in the former GDR attended a professional training during or after school and received a certificate. Clearly, the value of these certificates in a market-economy is questionable.

4. First Empirical Evidence

4.1. Econometric Methodology

The Potential Outcome Framework and Selection Bias Since we work with non-experimental data, we have to deal with some identifying issues. As we consider only one specific programme compared to 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 then defined as the difference between her potential outcomes:

$$\Delta_i = Y_i^1 - Y_i^0. \tag{1}$$

Since there will be never an opportunity to estimate individual effects with confidence, we have to concentrate on population averages of gains from treatment. Two parameters are most frequently estimated in the literature. The first one is the (population) average treatment effect (ATE), which is simply the difference in the expected outcomes after participation and nonparticipation. This parameter answers the question which would be the outcome if individuals were randomly assigned to treatment. Heckman (1997) notes, that this estimate might not be of relevance to policy makers because it includes the effect on persons for whom the programme was never intended. For example, if a programme is specifically targeted at individuals with low family income, there is little interest in the effect of such a programme for a millionaire. Therefore, the most prominent evaluation parameter is the so called average treatment effect on the treated (ATT), which focusses explicitly on the effects on those for whom the programme is actually intended. It is given by:

$$ATT = E(\Delta \mid D = 1) = E(Y^1 \mid D = 1) - E(Y^0 \mid D = 1). \tag{2}$$

The expected value of the ATT is defined as the difference between the expected values of the outcome with and without treatment for those who actually participated in treatment. In the sense that this parameter focuses directly on actual treatment participants, it determines the realised gross gain from the programme and can be compared with its costs, helping to decide whether the programme is a success or not (Heckman, LaLonde, and Smith, 1999).

Given equation (2) the problem of selection bias is straightforward to see. Remember that the second term on the right hand side of equation (2) is unobservable as it describes the hypothetical outcome without treatment for those individuals who received treatment. If the condition $E(Y^0 \mid D = 1) = E(Y^0 \mid D = 0)$ holds, we can use the non-participants as an adequate control group. This identifying assumption is likely to hold only in randomised experiments, with nonexperimental data it will usually not hold. Consequently, estimating the ATT by the difference in 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, since

$$E(Y^1 \mid D = 1) - E(Y^0 \mid D = 0) = E(Y^1 - Y^0 \mid D = 1) + \{E(Y^0 \mid D = 1) - E(Y^0 \mid D = 0)\}. \tag{3}$$

The second last term in (3) is the actual average treatment effect on the treated, whereas the term in the squared brackets is the selection bias. Selection bias arises because participants and nonparticipants are selected groups that would have different outcomes, even in absence of the programme. This bias might come from observable factors like age or skill differences or unobservable factors like motivation. For both cases different estimation strategies are available.[12] If we are willing to assume that selection only occurs on observed characteristics, the matching estimator is an appealing choice. Its basic idea is to search from a large group of non-participants those individuals who are similar to the treated group in all relevant (observable) characteristics. See Imbens (2004) or Smith and Todd (2004) for a recent review regarding matching methods.

How Matching Solves the Evaluation Problem Matching is based on the identifying assumption that, conditional on some covariate X, the outcome Y is independent of D. Since we focus on the ATT, it is sufficient to assume that (in the notation of Dawid (1979)):

Assumption 1 *Unconfoundedness for Controls:*

$$Y^0 \amalg D \mid X, \tag{4}$$

where $\amalg$ denotes independence. If assumption 1 is true, then $F(y_0 \mid X, D = 1) = F(y_0 \mid X, D = 0)$ which means, that conditional on X the non-participant outcomes have the same distribution that participants would have experienced if they had not participated in the programme (Heckman, Ichimura, and Todd, 1997). Similar to randomization in a classical experiment, matching balances the distributions of all relevant[13] pre-treatment characteristics X in the treatment and control group, and thus achieves independence between the potential outcomes and the assignment to treatment. Hence, if the mean exists,

$$E(Y^0 \mid X, D = 1) = E(Y^0 \mid X, D = 0) = E(Y^0 \mid X)$$

and the missing counterfactual mean can be constructed from the outcomes of non-participants. In order for both sides of the equations to be well defined simultaneously for all X it is usually additionally assumed, that

Assumption 2 *Weak Overlap:*

$$Pr(D = 1 \mid X) < 1. \tag{5}$$

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 (2), because the moments

[12] See for example Heckman, LaLonde, and Smith (1999), Angrist and Krueger (1999) or Blundell and Costa Dias (2002).

[13] If we say relevant, we mean all those covariates that influence the assignment to treatment as well as the potential outcomes.

of the distribution of Y^1 for the treated are directly estimable. Under the above stated assumptions - either assumptions 1 and 2 - the mean impact of treatment on the treated can be written as:

$$\begin{aligned}\Delta_{ATT}^{MAT} &= E(Y^1 - Y^0|D=1) \\ &= E(Y^1|D=1) - E_X[E(Y^0|X, D=1)|D=1] \\ &= E(Y^1|D=1) - E_X[E(Y^0|X, D=0)|D=1],\end{aligned} \tag{6}$$

where the first term can be estimated from the treatment group and the second term from the mean outcomes of the matched comparison group and the outer expectation is taken over the distribution of X in the treated population. The method of matching can also be used to estimate the ATT at some points $X = x$, where x is a particular realisation of X:

$$ATT(X = x) = E(\Delta \mid X = x, D = 1) = E(Y^1 \mid X = x, D = 1) - E(Y^0 \mid X = x, D = 1). \tag{7}$$

This parameter measures the mean treatment effect for persons who were randomly drawn from the population of the treated given a specific realisation of certain characteristics X, e.g. if X is the educational level, one could define the expected impact for those with an university degree.

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, that is $X \amalg D|b(X)$. For participants and non-participants 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. The propensity score can be seen as the coarsest balancing score, and X is the finest balancing score (Rosenbaum and Rubin, 1983). The authors also show that if treatment assignment is strongly ignorable given X, it is also strongly ignorable given any balancing score. The unconfoundedness for controls in assumption 1 then extends to the use of the propensity score:

Assumption 3 *Unconfoundedness given the Propensity Score:*

$$Y^0 \amalg D|P(X). \tag{8}$$

Matching procedures Several matching procedures have been suggested and we will review them here shortly. A good overview can be found in Heckman, Ichimura, Smith, and Todd (1998) and Smith and Todd (2004). To introduce them a more general notation is needed: Let I_0 and I_1 denote the set of indices for nonparticipants and participants. 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 control group observations $j \in I_0$ in the following way:

$$\Delta^{MAT} = \frac{1}{N_1} \sum_{i \in I_1} [Y_i^1 - \sum_{j \in I_0} W_{N_0 N_1}(i,j) Y_j^0], \tag{9}$$

where N_0 is the number of observations in the control group I_0 and N_1 is the number of observations in the treatment group I_1. Matching estimators differ in the weights attached to the members of the comparison group (Heckman, Ichimura, Smith, and Todd (1998), where $W(i,j)$ 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 $\sum_j W(i,j) = 1, \forall i$, that is the total weight of all controls 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 nonparticipants $j \in I_0$ for whom $P_j \in C(P_i)$ for each i. The persons matched to i are those people in the set A_i where $A_i = \{j \in I_0 | P_j \in C(P_i)\}$. The matching estimators discussed in the following differ in how the neighbourhood is defines and the weights are constructed (Smith and Todd, 2004).

Nearest-Neighbour-Matching Nearest neighbour (NN) matching sets

$$C^{NN}(P_i) = \min_j \|P_i - P_j\|, j \in N_0, \tag{10}$$

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:

$$W^{NN}_{N_0 N_1}(i,j) = \begin{cases} 1 & \text{if } \|P_i - P_j\| = \min_j \|P_i - P_j\| \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Several variants of NN matching are proposed, e.g. NN matching 'with' and 'without replacement'. In the former case a non-participating 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. This is of particular interest with data where the distribution of the propensity score is very different in the treatment and the control group. For example if we have a lot of treated individuals with high propensity scores, but only few control individuals with high propensity scores, we get bad matches as some of the high-score participants will get matched to low-score non-participants. This can be overcome by allowing replacement, which in turn reduces the number of distinct participants used to construct the counterfactual outcome and thereby increases the variance of the estimator (Smith and Todd, 2004). Another problem which is related to NN matching without replacement is that the estimates depend on the order in which observations get matched.

It is also suggested to use more than one nearest neighbour ('oversampling'). This form of matching involves a trade-off between variance and bias, too. It trades reduced variance, resulting from using more information to construct the counterfactual from each participant, with increased bias that results from on average poorer matches (Smith and Todd, 2004). When using oversampling one has to decide, how many matching partners m should be chosen for each individual i and which weight should be assigned to them. One possibility is to use uniform weights, that is all the m control individuals within set A_i receive the weight $\frac{1}{m}$, whereas all other individuals from the control group receive the weight zero:

$$W^{NN0_1}(i,j) = \begin{cases} \dfrac{1}{m} & \text{if } j \in A_i \\ 0 & \text{else} \end{cases} \tag{12}$$

Another possibility is to use triangular weights like suggested by Davies and Kim (2001). To do so, the m individuals within set A_i have to be ranked, where $\rho = 1$ is the closest neighbour, $\rho = 2$ is the next closest neighbour and so on. The weights can then be written as:

$$W^{NN0_2}(i,j) = \begin{cases} \dfrac{2(m-\rho+1)}{m(m+1)} & \text{if } j \in A_i \\ 0 & \text{else} \end{cases} \tag{13}$$

Caliper and Radius Matching 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 (Cochrane and Rubin, 1973), imposes the condition:

$$\|P_i - P_j\| < \epsilon, j \in N_0, \tag{14}$$

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

$$W^{CM}(i,j) = \begin{cases} 1 & \text{if } \|P_i - P_j\| = \min_j \|P_i - P_j\| \wedge \|P_i - P_j\| < \epsilon \\ 0 & \text{else} \end{cases} \tag{15}$$

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. As Smith and Todd (2004) note, a possible drawback of caliper matching is that it is difficult to know a priori what choice for the tolerance level is reasonable.

Dehejia and Wahba (2002) suggest a variant of caliper matching which we will call radius matching. The basic idea of this variant is to use not only the nearest neighbour within each caliper but all of the comparison members within the caliper. A benefit of this approach is that it uses only as many comparison units as are available within the calipers and therefore allows for the use of extra (fewer) units when good matches are (not) available. Hence, it shares the attractive features of oversampling mentioned above, but avoids the risk of bad matches.

Stratification and Interval Matching The idea of stratification matching (SM) is to partition the common support of P into a set of intervals and to calculate the impact within each interval by taking the mean difference in outcomes between the treated and the control observations. This method is also known as interval matching, blocking and subclassification (Rosenbaum and Rubin, 1983).

To implement SM the (estimated) propensity score is used, to divide the full sample into M blocks of units of approximately equal probability of treatment. Let J_{im} be an indicator for unit i being in block m. One way of implementing this is to divide the unit interval into M blocks with boundary values equal to $\frac{m}{M}$ for $m = 1, ..., M-1$ so that (Imbens, 2004):

$$J_{im} = 1\left\{\frac{(m-1)}{M} < P(X_i) \leq \frac{m}{M}\right\}. \tag{16}$$

Within each block there are N_{1m} treated and N_{0m} untreated individuals, where $N_{1m} = \sum_i 1\{D_i = 1, J_{im} = 1\}$ and $N_{0m} = \sum_i 1\{D_i = 0, J_{im} = 1\}$. Hence, given these subgroups the average treatment effect within each block can be estimated as if random assignment was held:

$$\Delta_m^{SM} = \frac{1}{N_{1m}} \sum_{i=1}^{N} J_{im} D_i Y_i - \frac{1}{N_{0m}} \sum_{i=1}^{N} J_{im}(1-D_i)Y_i \tag{17}$$

The ATT is estimated by weighting the within-block average treatment effects by the number of treated units:

$$\Delta_{ATT}^{SM} = \sum_{m=1}^{M} \Delta_m^{SM} \cdot \frac{N_{1q}}{N_1}. \tag{18}$$

Clearly, one question to be answered is how many blocks should be used in empirical analysis. Cochrane (1968) shows that five subclasses are often enough to remove 95% of the bias associated with one single covariate. Since, as Imbens (2004) notes, all bias under unconfoundedness is associated with the propensity score, this suggests that under normality the use of five blocks removes most of the bias associated with all covariates. In a recent application (Aakvik, 2001) chooses ad-hoc twelve subgroups. One way to justify the choice of m is to check the balance of the propensity score (or the covariates) within each block. Most of the algorithm can be described in the following way: First, check if within a block the propensity score is balanced. If not, the blocks are too large and need to be split. If, conditional on the propensity score being balanced, the covariates are unbalanced, the specification of the propensity score is not adequate and has to be re-specified, e.g. through the addition of higher-order terms or interactions (Dehejia and Wahba, 1999).

4.2. First Results

Estimating the Propensity Score We 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 estimated separate models for the four main groups.[14] Our explanatory variables contain information on the socio-demographic background of the individuals, like age, marital status, the number of children, nationality and health restrictions. Furthermore, the individual qualification is described by the professional training, the occupational group, the professional rank and available work experience. The influence of the individual labour market history is given by the unemployment duration, the duration of the last occupation, the number of placement propositions by the labour office, the last contact to the personal caseworker, whether the

[14] We 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 mathing is reduced, i.e. the standardised bias (see below) is higher.

person is an aspirant for special vocational rehabilitation, special placement restraints due to health restrictions and information on a ALMP participation in the past. The regional context is explained by the classification of the FEA for comparable labour office districts (see Appendix B).

Table 2 presents the results of the estimation of the participation probability for JCS. Consideration of the socio-demographic characteristics makes it obvious, that German nationals have a higher participation probability than foreigners. Also persons with health restrictions are more likely to participate. The comparison of the coefficients of the marital status between East and West Germany shows differences. While married persons are more likely to participate in the East, married persons in West have a reduced participation probability. Referring to the age shows also differences. Older persons are less likely to participate in the West, while older persons in the East are more likely. The number of children has only a significant negative influence on the participation probability for men in West Germany.

The consideration of the characteristics that describe the individual qualification shows clear differences between genders and regions. Men in West Germany with an industrial training are for example less likely to participate than men in the same region without any professional training; for women in East Germany this finding is the other way round. An interesting point to note is that a higher education increases the participation probability for women in both regions. This finding might indicate that higher qualified women are better motivated to finish their unemployment even by taking an occupation in a JCS. The coefficient for the professional rank encourages this speculation.

The individual labour market history shows, that a longer foregoing employment has a negative influence on the participation probability, i.e. persons, who were employed for a longer period of time before unemployment, are less likely to participate. This indicates other re-integration opportunities for these persons. We included the duration of unemployment using three categories: Up to 13 weeks, between 13 and 52 weeks and more than 52 weeks. As expected, a higher unemployment duration increases the participation probability. The number of placement propositions by the labour office indicates the worse integration chances of the individual and has a positive impact on the participation probability. In contrast, persons with placement restraints due to health restrictions are less likely to participate. This is somewhat surprising since JCS should stabilise and qualify those groups for later re-integration.

Tab. 2: 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^2	*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	Reference		Reference		Reference		Reference	
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	Reference		Reference		Reference		Reference	
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	Reference		Reference		Reference		Reference	
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	Reference		Reference		Reference		Reference	
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	Reference		Reference		Reference		Reference	
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	Reference		Reference		Reference		Reference	
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								
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	Reference		Reference	
Cluster III	*-0.1841*	0.0722	**-0.4601**	0.0917				
Cluster IV	-0.0080	0.1002	**-0.4530**	0.1423				
Cluster V	Reference		Reference					

Bold letters indicate significance at the 1% level. *Italic* letters refer to the 5% level.

[1] DoR = degree of restriction

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

The coefficients for the regional context are not very intuitive. 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 plainest reduce of participation probability, while analogously for West German women and men living in labour office districts dominated by large cities with a 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 to participate.

Results and Sensitivity to Matching-Algorithm In section 4.1. we have presented several different matching algorithms. Since we have a very large group of control observations, it makes no sense to use kernel matching methods. However, even if we concentrate on nearest-neighbour matching, several choices have to be made when implementing the estimator. Each choice involves a trade-off between bias and efficiency. First, we have to choose how many nearest neighbours should be used for estimating the counterfactual outcome. When using more than one nearest neighbour we trade reduced variance, resulting from more information to construct the counterfactual outcome, with increased bias that results from on average poorer matches. NN-matching can also be done with a pre-specified tolerance level to ensure that only good matches (in terms of the distance in the propensity scores) are made. Finally, one has also to decide if matching should be done with replacement or not. Especially for data where only a few comparable control observations are available this choice matters a lot. The price to pay for matching with replacement and using control observations more than once is an increased variance of the estimator. In large samples each estimator should lead to the same result and therefore a good sensitivity test is to compare the results from different approaches.

Due to our very large number of observations in the comparison group, the likelihood of finding good matches is quite high. Hence, we will use a NN-matching without replacement and impose a caliper of 0.02. To test the sensitivity of our results with respect to the choice of the matching algorithm we also estimate NN-matching with replacement, 5-NN-matching and 10-NN-matching. The results of each matching estimator can be found in table 3.

First of all, it is worth mentioning that the effects are not very sensitive to the algorithm choice. For men in West Germany the effects of JCS in December 2002 are centered around zero, no significant effects could be established. For women in West Germany we get a significant positive effect between 4.59% and 6.10%, whereas in East Germany the effects are significantly negative for men (from -2.37% to -2.91%) and negative for women (from -1.01% to -1.93%). Clearly, from an economic standpoint this is a disappointing result, since only one of the four groups benefits from participation in JCS. From an econometric standpoint the small bandwidths of the different estimators are promising, since they show that the results are not sensitive to the chosen matching algorithms. Before we discuss the economic result in more detail, let us mention two points concerning the sensitivity. First, caliper matching without replacement reduces the number of matches (in comparison to NN-

Tab. 3: Comparison of the Estimated Effects by Different Matching-Estimators in December 2002

West Germany	**Men**			**Women**		
Method	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
NN without Replacement	-0.0005	0.0108	2,132	**0.0554**	0.0222	1,028
NN without Repl., Caliper 0.02	-0.0019	0.0098	2,123	**0.0459**	0.0208	980
NN with Replacement	0.0061	0.0142	2,140	**0.0504**	0.0216	1,052
5-NN	0.0011	0.0143	2,140	**0.0529**	0.0171	1,052
10-NN	-0.0003	0.0101	2,140	**0.0610**	0.0145	1,052

East Germany	**Men**			**Women**		
Method	Effect	Std. Err.	No. of Partici-pants	Effect	Std. Err.	No. of Partici-pants
NN without Replacement	**-0.0291**	0.0091	2,924	**-0.0135**	0.0058	5,032
NN without Repl., Caliper 0.02	**-0.0287**	0.0081	2,923	-0.0135	0.0070	5,027
NN with Replacement	**-0.0294**	0.0076	2,924	**-0.0193**	0.0083	5,035
5-NN	**-0.0237**	0.0074	2,924	-0.0101	0.0067	5,035
10-NN	**-0.0249**	0.0071	2,924	-0.0106	0.0058	5,035

Bold letters indicate significance on a 5% level. Standard errors calculated by bootstrapping with 50 replications.

matching without caliper but with replacement) by 17 men (0.79%) and 70 women (6.65%) in West Germany, but only for 1 men and 8 women in East Germany. This indicates that it is much easier to find comparable matches in the East than it is in the West. In other words, imposing a common support condition in form of caliper matching makes sense for West Germany, since we would face the risk of bad matches otherwise. Second, oversampling with 10-NN delivers - as expected - the lowest variance, even though this seems not to be of great importance, because there are no clear differences between the standard errors of the matching algorithms. Since we want to avoid the risk of bad matches, we choose to use the 1-NN-matching method with a caliper of 0.02 throughout the rest of the paper.

Quality of Propensity Score Estimation and Matching Our model specification for the propensity score estimation was based on specification tests to identify the relevant variables. One simple method to validate the ability of a good prediction is the computation of the out-of-sample hit-rate, i.e. the proportion of persons with a correct prediction of their status (participation and nonparticipation). As becomes obvious from table 4, these hit-rates lie between 70.6% for men in and 75.7% for women in West Germany and implies a quite accurate underlying model.

Since we do not condition on all covariates but on the propensity score, we have to check the ability of the matching procedure to balance the relevant covariates 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 the difference of the sample means in the treated and (matched) comparison subsamples as a percentage of the square root of the average of the sample variances in both groups. The SB before matching is given by

$$SB_{\text{before}} = 100 \cdot \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5 \cdot (V_1(X) + V_0(X))}}, \tag{19}$$

the SB after matching is given by

$$SB_{\text{after}} = 100 \cdot \frac{(\overline{X}_{1M} - \overline{X}_{0M})}{\sqrt{0.5 \cdot (V_{1M}(X) + V_{0M}(X))}}, \tag{20}$$

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}, $X_{0M}(V_{0M})$ are the corresponding values after matching. This is a common approach used in many evaluation studies, e.g. by Sianesi (2001). To abbreviate the documentation we calculated the means of the SB before and after matching for the four main groups (Table 4). 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.

Tab. 4: Some Quality Indicators

	West Germany		**East-Germany**	
	Men	**Women**	**Men**	**Women**
Before Matching				
Observations	46,235	35,271	67,712	81,505
Hit-Rate	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)
Log-Likelihood	-7462.4	-3891.5	-10572.4	-16733.2
Mean of Standardised Bias[1]	14.62	16.08	12.01	10.83
After Matching				
Observations	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)
Log-Likelihood	-2924.1	-1346.9	-4034.5	-6949.3
Mean of Standardised Bias[1]	2.51	3.20	1.78	1.60

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

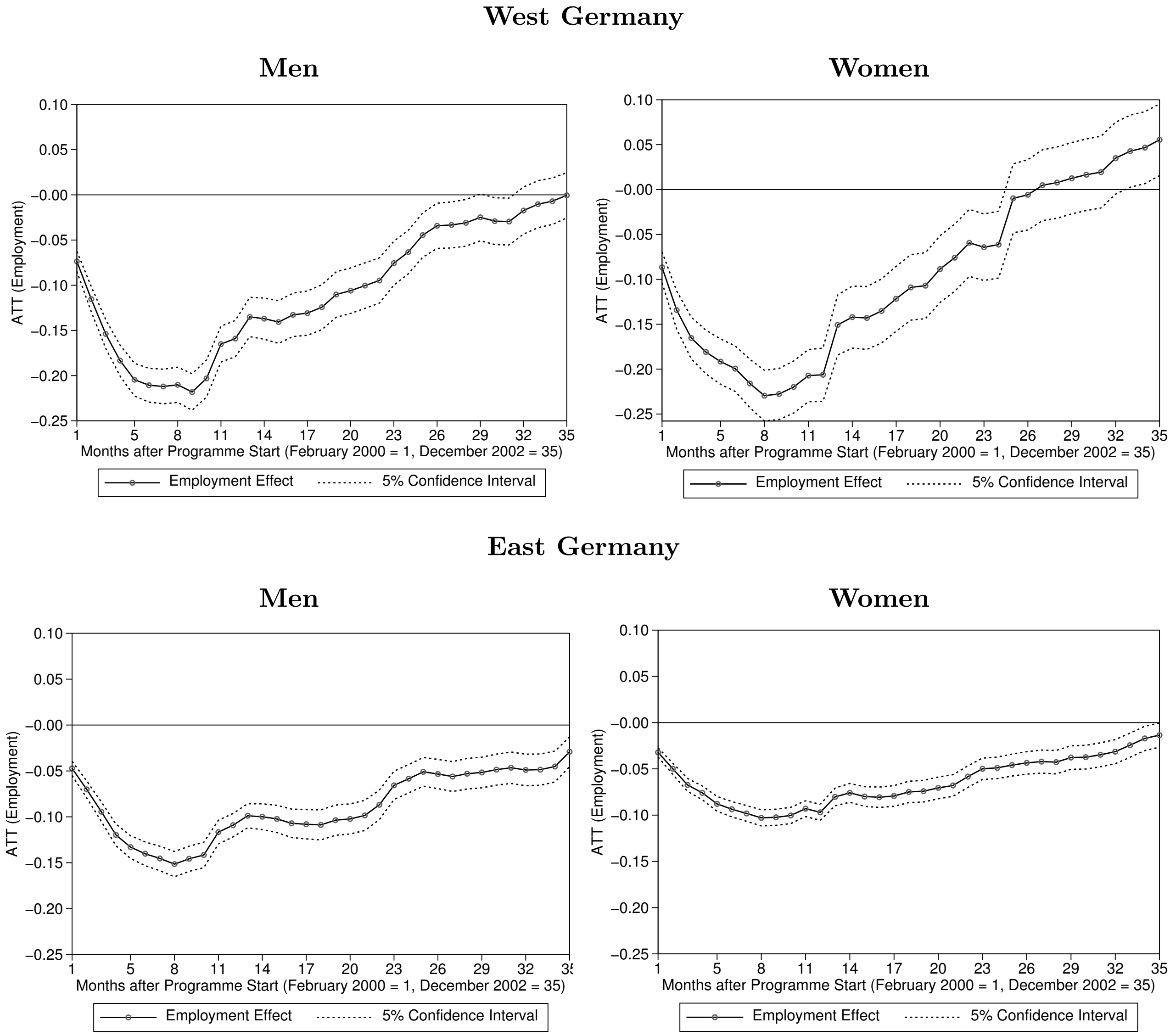
Sianesi (2004) suggests to re-estimate the propensity on the matched sample, that is only on participants and matched nonparticipants and compare the pseudo-R^2's before and after matching. The pseudo-R^2 indicates, how well the regressors X explain the participation probability. After matching there 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 4 show, this is true for our estimation. For the sake of completeness we also added the Log-Likelihood values, the results of the F-test (with degrees of freedom in brackets) and the number of observations before and after matching.

Effects over Time All estimated effects in this paper correspond to December 2002, the last month of our observation period. We are conscious, that consideration of only this month bears some shortcomings for a valuable interpretation of the programme effects. First, since December 2002 is almost three years after programmes start and for the majority of programmes nearly two years after the end, attributing all variation in the labour market status between participants and non-participants to the programme might oversimplify the situation, but comparability between participants and nonparticipants cannot be reviewed at this point of time. Second, since we only consider the effects of the first programme and further participation is regarded as an outcome of this first participation, the impact of multiple participation is not estimated separately. To give an idea of the time path of the effects, figure 1 presents the course of the estimated effects for the four main groups between February 2000 and December 2002.

At the beginning of the observation period, the programme effect is expected to be overlayed by the 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 work as regular employees and spend less time on job search for regular, but non-subsidised jobs. 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 for the re-integration into regular employment, we would expect increasing employment rates after the programmes end. 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 course of the effects for men in West Germany is also rising, but the effects are insignificant in the end, i.e. an increase in the employability by participation cannot be established. While the effects for West Germany are clearly rising, in East Germany we find a stepwise increase with relatively constant levels over one-year-periods. 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 non-participants.

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. For all other groups, JCS seem rather to harm the employability of the participants than to increase it. 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.

Fig. 1: ATT (Employment) between February 2000 and December 2002



Part II. Effect Heterogeneity

Clearly, as already mentioned, one possible explanation for the discouraging results in the previous section can 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. The central motivation in this context is that programme impacts are heterogeneous (Manski (1997) and (2000)). Therefore, the average effects from the above section must not apply to all strata of the population. Heckman, LaLonde, and Smith (1999) mentions, that negative mean impact results may be acceptable, if the majority of participants gains from the programme. Abandoning from 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. Especially, if we are able to identify the personal 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.

For this reason, we will present three approaches to identify potential sources of effect heterogeneity. 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 due to a better fit of the promotion. 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.

5. Effects for Target Groups

5.1. Selected Target Groups of Job Creation Schemes

Identifying groups, that benefit from programmes is a central purpose of programme evaluation. Recent evaluation studies of JCS in Germany (Hujer, Caliendo, and Thomsen, 2003) 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 lower employability than the average is a reasonable first approach to identify possible effect heterogeneity due to personal characteristics. The Social Code III targets several groups of individuals, who should be predominantly promoted by ALMP, like long-term unemployed, individuals with health restrictions or persons, who aspire to vocational rehabilitation[15]. Further target groups are young unemployed and the older as well as workers without any professional training. In addition, JCS in particular should be 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 unemployed with a duration of unemployment for more than one year, persons with special placement restrictions due to health restrictions and attendants in vocational rehabilitation. In addition to that, we select persons, who are hard to place indicated by more than five (unsuccessful) placement propositions by the local labour office, persons, who have already participated in an ALMP programme before unemployment, the group of persons without professional training and finally persons without work experience. Table 5 presents the estimated employment effects in December 2002 for these groups with the distinction for gender and region as above.

[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.

Tab. 5: Effects for Selected Target Groups in December 2002

West Germany	**Men**			**Women**		
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**		
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 should be stabilised and qualified in JCS for a new profession.
[2] Placement restrictions refer to the assessment of the caseworker, that the health restrictions of the participant reduce the number of job opportunities.

As the results show, the effects of the programmes are rather heterogeneous for the single target groups. While the results for the main groups in table 3 showed insignificant estimates for men and significant positive effects for women in West Germany, the effects for participants in East Germany were significant negative for men and insignificant for women. The insignificant effects for the whole group of men in West Germany split up into heterogeneous effects for the selected target groups as the results in table 5 show. One important finding here is, that particulary in West Germany, older unemployed women benefit from participation. The mean employment rates for West German women are 12.67% above the non-treatment state. Also participation of older men in this region tends to result in positive impacts, but the results are statistically not significant. Obviously, a tighter allocation of unemployed women in West Germany might raise the average outcome of the whole programme, too. For the groups in East Germany, no significant effects could be established. Taking a look at the groups of younger unemployed persons shows, that there are no significant findings. This should be interpreted as no effect on the employability by participation. But, apart from the significance, the results are rather negative than positive (except for women in East Germany). Referring to the purpose of the programmes, these results are plausible. Since occupations in JCS have to be additional in nature, i.e. they do no compete with regular occupations to avert substitution effects, the qualifying elements for market-competitive jobs have be assumed to be negligible. On the other hand, the stabilisation may be more important. Therefore, allocating young unemployed to improve their labour market chances by human capital generation is misleading. Furthermore, participation in JCS comes along with a stigma of the participant, i.e. potential employers suspect a reduced productivity. This stigma seems to be very important especially in the beginning of the labour market career. For older unemployed, this instance should only be of reduced importance. In contrast, for older unemployed participation may be seen as a motivation for changing the personal situation and taking almost every occupation.

Other results, that assist these assumptions, are the effects for long-term unemployed. They benefit from participation in West as well as in East Germany. Again, the effects are higher for participants in West Germany with 5.03% for men and 11.25% for women, but also women's situation in East Germany is improved by 2.45%. Since JCS are especially designed for persons with placement restrictions, we expect positive impacts for these groups. As the results show, there are no significant results. This is somewhat surprising, and the potential reason for this has to be enlightened. If the outside options for nonparticipants are better than after participation, this might be due to other kinds of promotion or even a very bad allocation of individuals to the programme. But, taking a look at the value for women in West Germany shows, that with a 10% significance level the employment rates after participation increase by 10.26%. This finding is a least a bit promising, that the purpose of the programmes is not totally missed.

Another interesting finding are the results for men in East Germany, who have already participated in an ALMP programme before unemployment. There is a significant negative impact of participation on the employability for regular employment. This finding assists the general supposition, that especially in the East so called 'programme careers' exist, i.e. persons, who are placed again and again in ALMP programmes without any positive results. For the other groups, except women in West Germany with more than five (unsuccessful) placement propositions, there are no significant differences in the labour market situation after participation. Besides the insignificant results for most of the groups, the set of clear findings should be regarded for further allocation. Especially the enacted target group of long-term unemployed persons seem to benefit from participation, whereas the employability of persons with programmes in the past is not increased. A tighter targeting to the group of long-term unemployed together with a reduced allocation of the other groups will improve the gains of JCS.

5.2. Target Score

The results of the estimation of the programme effects for the selected target groups show that most of the groups do not benefit from participation. Since these groups refer only to a single targeting criterion, the question arises whether individuals with more than one disadvantage on the labour market may benefit. To answer this question we build a simple index we would call 'target score' as the sum of the individual number of disadvantages from section 5.1. Without any particular weighting, each category adds one point to the target score. Persons, who do not belong to any of the categories in section 5.1., 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, he/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 assigned a target score of more than five, we summarised these persons in one group, i.e. target score five; the other categories refer to the actual number of restrictions. We estimate the programme effect on the employment rates in December 2002 within each target score class.

Tab. 6: Estimated Effects for the Target Scores 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	**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	1090
2	-0.0225	0.0155	937	-0.0093	0.0110	1754
3	0.0013	0.0191	821	0.0112	0.0103	1289
4	-0.0161	0.0213	322	0.0062	0.0159	508
5	-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.

If programmes are designed according to the needs of the most disadvantaged, one would expect higher target scores to indicate a higher need of assistance and a better outcome. The estimates of the effects in December 2002 are given in table 6. 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 1 or 2 are rather harmed, women with the same score seem to benefit. In East Germany, groups with a target score of less than 3 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 3.

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 choice of selected target groups is not complete. There are other characteristics, that increase or decrease the individual employment probability. 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. This issue can be analysed by building groups according to separately defined multiple disadvantages.

Taking a look at the significance of the results shows that our assumption cannot totally 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% after participation, for women with a target score of four by 11.76%. For the other groups, the estimates are not statistically significant, i.e. no clear increase nor 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% 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. Effects of 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 might benefit more since a high participation probability can to some extend 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 covariates X or respectively the propensity score into several strata. Within these strata, participants and non-participants 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 that 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. 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. Due to the large numbers of observations in our samples, using the whole range of the propensity scores of participants and nonparticipants leads to a skewed stratification. Referring to the propensity scores of the participants only reduces this skewness. The choice of twenty strata for each of the four groups emerged from some balancing tests of the propensity score among treated and comparison persons using a smaller number of blocks. As expected, the number of nonparticipants decreases over strata with higher participation probabilities. But, for all groups except women in West Germany this stratification leaves meaningful numbers of observations in each stratum.

The estimated treatment effects for each stratum are presented in table 7 for East Germany and in table 8 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 training 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 give the

Tab. 7: 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	Treatment	146	0.0001	0.1781	-0.0585	251	0.0002	0.1355	0.0134
	Comparison	16,171		0.2366		18,980		0.1221	
2	Treatment	146	0.9303	0.1781	-0.0666	252	0.0168	0.1032	-0.0235
	Comparison	9,532		0.2446		11,309		0.1267	
3	Treatment	146	0.0218	0.1233	**-0.0897**	252	0.1633	0.1190	-0.0267
	Comparison	7,657		0.2130		7,396		0.1458	
4	Treatment	146	0.3283	0.1575	-0.0347	252	0.1581	0.0913	-0.0568
	Comparison	5,529		0.1923		5,641		0.1480	
5	Treatment	147	0.0537	0.0816	*-0.0772*	251	0.2593	0.1633	0.0137
	Comparison	4,432		0.1588		5,098		0.1497	
6	Treatment	146	0.2077	0.1233	-0.0245	252	0.1555	0.1111	-0.0245
	Comparison	3,093		0.1478		4,298		0.1356	
7	Treatment	146	0.9609	0.0822	-0.0476	252	0.5875	0.1627	0.0178
	Comparison	2,727		0.1298		3,852		0.1449	
8	Treatment	146	0.4523	0.0685	-0.0497	252	0.3221	0.1071	*-0.0494*
	Comparison	2,640		0.1182		2,804		0.1566	
9	Treatment	146	0.5098	0.1027	-0.0201	251	0.2600	0.1036	*-0.0609*
	Comparison	2,116		0.1229		2,785		0.1645	
10	Treatment	147	0.7602	0.1020	-0.0173	252	0.1690	0.0952	-0.0423
	Comparison	2,037		0.1193		2,276		0.1375	
11	Treatment	146	0.4703	0.0616	-0.0440	252	0.3124	0.1190	-0.0192
	Comparison	1,448		0.1057		2,228		0.1382	
12	Treatment	146	0.4960	0.0959	-0.0165	252	0.9466	0.1508	0.0133
	Comparison	1,592		0.1124		1,665		0.1375	
13	Treatment	146	0.3424	0.0411	**-0.0729**	251	0.9627	0.1036	-0.0151
	Comparison	1,132		0.1140		1,651		0.1187	
14	Treatment	146	0.8348	0.0616	-0.0373	252	0.0541	0.1310	0.0371
	Comparison	980		0.0990		1,471		0.0938	
15	Treatment	147	0.7724	0.1224	0.0296	252	0.2967	0.0992	0.0126
	Comparison	948		0.0928		1,143		0.0866	
16	Treatment	146	0.8285	0.0890	0.0152	252	0.9422	0.1071	0.0164
	Comparison	772		0.0738		1,124		0.0907	
17	Treatment	146	0.9521	0.0753	0.0253	251	0.3790	0.0797	-0.0071
	Comparison	600		0.0500		910		0.0868	
18	Treatment	146	0.4996	0.0822	*0.0403*	252	0.6872	0.0913	-0.0129
	Comparison	645		0.0419		749		0.1041	
19	Treatment	146	0.0053	0.0548	0.0193	252	0.7600	0.1349	0.0192
	Comparison	479		0.0355		648		0.1157	
20	Treatment	147	0.6655	0.0748	0.0244	252	0.6248	0.1548	0.0267
	Comparison	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 2.

[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.

average treatment effect on the treated. Obviously, these effects are similar to those estimated with the other matching estimators in section 4.2. In addition to the mean outcomes and the effects, the

Tab. 8: 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	Treatment	107	0.0000	0.1869	*0.0764*	52	0.0005	0.3846	**0.2649**
	Comparison	14,220		0.1105		12,954		0.1197	
2	Treatment	107	0.1905	0.1963	-0.0046	53	0.1774	0.3585	*0.1194*
	Comparison	4,913		0.2009		4,119		0.2391	
3	Treatment	107	0.2521	0.2336	0.0034	52	0.5364	0.3077	0.0201
	Comparison	4,065		0.2303		2,754		0.2876	
4	Treatment	107	0.8130	0.2150	-0.0355	53	0.7943	0.3962	0.1169
	Comparison	3,522		0.2504		2,782		0.2793	
5	Treatment	107	0.0430	0.2617	0.0278	53	0.6186	0.3019	-0.0110
	Comparison	2,403		0.2339		1,742		0.3129	
6	Treatment	107	0.5197	0.1682	*-0.0998*	52	0.7633	0.2692	-0.0341
	Comparison	2,384		0.2680		1,556		0.3033	
7	Treatment	107	0.0045	0.2056	-0.0484	53	0.9023	0.3585	0.0370
	Comparison	2,331		0.2540		1,347		0.3215	
8	Treatment	107	0.4353	0.2056	-0.0593	52	0.6411	0.2885	-0.0307
	Comparison	1,748		0.2649		1,366		0.3192	
9	Treatment	107	0.2616	0.2336	-0.0364	53	0.9991	0.2830	-0.0481
	Comparison	1,533		0.2701		1,214		0.3311	
10	Treatment	107	0.3627	0.2804	0.0005	53	0.6523	0.3396	-0.0242
	Comparison	1,229		0.2799		841		0.3639	
11	Treatment	107	0.1798	0.1963	-0.0831	52	0.8903	0.3269	-0.0184
	Comparison	1,049		0.2793		611		0.3453	
12	Treatment	107	0.5893	0.2991	0.0343	53	0.3965	0.2830	-0.0608
	Comparison	929		0.2648		733		0.3438	
13	Treatment	107	0.6554	0.2617	-0.0073	52	0.2097	0.3846	-0.0102
	Comparison	751		0.2690		623		0.3949	
14	Treatment	107	0.3683	0.2617	0.0088	53	0.3294	0.3208	-0.0260
	Comparison	684		0.2529		571		0.3468	
15	Treatment	107	0.5013	0.2056	-0.0667	53	0.2556	0.4340	0.1185
	Comparison	661		0.2723		447		0.3154	
16	Treatment	107	0.4412	0.2430	0.0452	52	0.0935	0.3077	0.0171
	Comparison	551		0.1978		265		0.2906	
17	Treatment	107	0.8646	0.1402	-0.0332	53	0.0282	0.3208	0.0615
	Comparison	473		0.1734		108		0.2593	
18	Treatment	107	0.0955	0.1308	0.0122	52	0.7560	0.3654	**0.1987**
	Comparison	295		0.1186		78		0.1667	
19	Treatment	107	0.4283	0.2617	**0.1413**	53	0.0389	0.3396	*0.1682*
	Comparison	191		0.1204		70		0.1714	
20	Treatment	107	0.0038	0.2710	**0.1606**	53	0.1637	0.3585	**0.2715**
	Comparison	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 2.

[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.

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. We also checked the balancing property of stratification by comparing the means of the incorporated variables in the logit models of participants and nonparticipants within strata as suggested by Rosenbaum and Rubin (1983). See Appendix C for details.

The results of the hypothesis tests show that the division into twenty strata stratification provides approximately equal propensity scores for most groups. Only for the groups at the borders of the propensity score range, this equality is hampered. For men in West Germany, strata 1, 5, 7 and 20 are imbalanced, for women in the same region strata 1, 17 and 19. In East Germany the strata with lower participation probabilities are more problematic. 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 7), 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 employability 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 around zero. Also for participants in West Germany (table 8) we do not find a consistent picture to our hypothesis. One can loosely see that higher participation probabilities correlate with higher impacts, but these findings may be inconsistent as the tests above show. It seems that the participation probability is no adequate measure for effect heterogeneity here and successful integration in the labour market depends on an other composition of the individual characteristics than does selection into programmes.

Part III. Allocation Mechanisms (not finished yet)

7. Overview

The potential improvement of allocation mechanisms is a much discussed topic in the recent evaluation literature (see for example Lechner and Smith (2003), Frölich, Lechner, and Steiger (2003) and Frölich (2001)). Due to the dissatisfying effects of ALMP, the possible misallocation of participants to programmes must be reviewed. 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). The success of the different allocation mechanisms has to be considered conditional on the underlying target. We follow Frölich, Lechner, and Steiger (2003) and summarise caseworker discretion, deterministic and random assignment under the categroy 'non-statistical allocation mechanisms', whereas profiling and targeting are summarised under 'statistical allocation mechanisms' and discuss both categories subsequently.

7.1. Non-Statistical Allocation Mechanisms

The most common allocation mechanism is caseworker discretion. 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, 2003). There is only a small literature that examines the quality of caseworker allocation in Europe. Frölich (2001) analyses the effects of caseworker allocation in Sweden; Lechner and Smith (2003) and Frölich, Lechner, and Steiger (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 might be the lack of knowledge of the 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, Lechner, and Steiger, 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 regard for how much different groups might have benefited from programmes (see for example Bell and Orr (2002)).

However, the evaluation of caseworker allocations might give some guidance for the implementation of an allocation algorithm based on statistical methods. Lechner and Smith (2003) state that caseworkers in Switzerland are making use of the flexibility available to them to assign unemployed persons in large numbers to all of the treatment types. Furthermore, they do not allocate persons at random with respect to their observed characteristics, but the allocation shows evidence for systematic, reasonable patterns. This result can also be derived from the study of Hujer, Caliendo, and Thomsen (2003), who analyse the effects of job creation schemes in Germany regarding potential programme heterogeneity. The composition of participants in the different occupations in job creation schemes show systematic differences with respect to the foregoing mean unemployment duration, mean age of the participants, and the mean individual qualification. For example, unemployed persons with a university or college degree are more often allocated to programmes in the office and services occupations, while persons without any professional training are primarily placed in the agricultural and construction sectors. Although the empirical results of caseworker allocation are not encouraging, caseworker assignment has got the virtue that it allows for idiosyncratic information about the unemployed person and about the institutional environment to affect the allocation process, that may be difficult to implement in a statistical decision system or to encode in a deterministic rule (Plesca and Smith, 2002). Moreover, caseworkers fulfil a set of additional functions besides allocation like monitoring the unemployed, encouraging them to look for work or training, networking with employers to develop opportunities for subsidised employment and keeping abreast of local training opportunities (Lechner and Smith, 2003).

As mentioned above, additionally or alternatively to caseworker allocation deterministic or random assignment mechanisms could be used.[16] While random assignment avoids a selectivity bias in allocation, it is not able to control for effective placement without further restrictions. Random assignment mechanisms are used in experimental design, where based on the population of eligible persons a sample is allocated to services while another is not. Due to ethical and financial restrictions in many countries, there is only a limited number of social experiments. A prominent example is the National JTPA study of Bloom, Orr, Cave, Bell, and Doolittle (1993) that implemented an experimental design after caseworkers determined the applicant's eligibility and allocated two thirds of the population to the treatment group and one third to the control group.

Deterministic allocation provides only a small potential to improve the effectiveness of placement, since programme effects are heterogenous with respect to individual characteristics and also for persons with the same set of (observable) characteristics. The virtue of deterministic rules are the simplicity and equity, in the sense of treating observationally equivalent cases in the same way (Plesca and Smith, 2002). An example for a deterministic allocation mechanism in Germany is that every first-time job-seeker gets an invitation to an individual counselling at the local labour office.

7.2. Statistical Allocation Mechanisms

Statistical allocation mechanisms should on the one hand help to allocate persons to programmes they will benefit from. On the other hand it should avoid allocation of persons to programmes that are harmful in terms of the expected goals (equity of efficiency, see above). The individual utility of participation in a specific programme is estimated using an statistical method or an econometric model. 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'. Using this definition the 'target score' in section 5.2. and the stratification approach in section 6. are suggestions for an allocation by profiling. Frölich, Lechner, and Steiger (2003) amend this definition by 'targeting', that they define as allocation by a statistical treatment rule based on predicted impacts. They use

[16] Experimental designs have been increasingly used to evaluate North American employment and training programmes during the late 1980s and 1990s. They are less common in Europe, though a small number of experiments have been conducted in Germany (Dann, Kirchmann, Spermann, and Volkert, 2002), Britain, Norway and Sweden (Heckman, LaLonde, and Smith, 1999).

the term 'profiling' as allocation based on predicted outcomes. In the North American literature these two are used interchangeably. We follow this distinction. Profiling tries to identify persons with the highest need of assistance estimating some kind of score, whereas targeting is the more sophisticated method and assesses hypothetical or predicted impacts of the individuals to support the allocation decision.

Profiling systems are typically implemented as indicators of the person's need for more intensive assistance. Possible indicators are whether the individuals are hard-to-place or at risk to become long-term unemployed (OECD, 2002). Profiling is a one-dimensional method since only a single score for every person is computed and persons are allocated to programmes according to this score. The success of profiling mechanisms depends on two crucial features. First, the quality of the score-prediction is fundamental since it determines the allocation of the unemployed individuals. Second, the profiling approach implicitly assumes a positive correlation between the effectiveness of programmes and the computed score, i.e. the individual's need for further assistance. That is, for individuals with a high score, participation in any programme should per se have positive impacts independently of the programme. This strong simplification of the allocation mechanism should be kept in mind (Frölich, Lechner, and Steiger, 2003).

Targeting systems try to overcome the problems of profiling systems by including the hypothetical impacts of alternative programmes (and/or nonparticipation) in the allocation mechanism. In the binary case (participation vs. nonparticipation) there are two potential outcomes to consider. If more than one programme alternative is evaluated, the approaches are multi-dimensional and require more sophisticated econometric methods (Frölich, Lechner, and Steiger, 2003). Analogously to the profiling system, targeting can be seen as a two-step estimation. In the first step, the potential outcomes for the treatments (including nonparticipation) are computed to calculate the impacts. In dependence of the underlying outcome variable of interest, different econometric methods can be used, e.g. a probit or logit model if the outcome variable is the binary employment status. The outcome is predicted conditional on the individual set of characteristics. Since the targeting system does not rely on the assumption of profiling (positive correlation between effectiveness of programmes and the profiling score), but on estimation of the hypothetical impacts for participation and nonparticipation, it provides a larger potential to improve efficiency. Nevertheless, targeting systems might face several problems in the empirical implementation. First, as the experiences from Canada have shown (see OECD (2002), Frölich, Lechner, and Steiger (2003)), if the introduction of such a system coincides with a lay-off of service delivery staff, it might cause systematic denial of the service staff. A second problem might be violation of the privacy rules of the unemployed, since accurate prediction of labour market outcomes requires a lot of information on the individuals. The ignorance of possible scale effects, if the targeting system allocates persons on a large extent to certain programmes, may also be problematic. Furthermore, the choice of a reasonable outcome variable for the broad variety of programmes may be difficult, since programmes have different duration and labour market purposes.

Plesca and Smith (2002) discuss the required data for implementing targeting respectively profiling.

The selection of the targeting (profiling) variable depends not only on the goals of the allocation rule, but also on the available data. If the goal is to assign the neediest persons, then a targeting (profiling) variable that correlates positively with need will be used. If efficiency is the goal, then an obvious variable is the predicted net impact. Both approaches require information on the observable characteristics in the estimated model for everyone who is to be profiled, as well as data on the sample that will be used to estimate the profiling model. To estimate expected impacts the ideal is to have experimental data so that the experimental impacts can be estimated without a bias as a function of observable characteristics.

8. Simulation

8.1. Allocation by Predicted Outcomes

< to be completed >

8.2. Allocation by Predicted Impacts

< to be completed >

References

AAKVIK, A. (2001): "Bounding a matching estimator: the case of a Norwegian training program," *Oxford Bulletin of Economics and Statistics*, 63(1), 115–143.

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.

BLOOM, H., L. ORR, G. CAVE, S. BELL, AND F. DOOLITTLE (1993): *The National JTPA Study: Title II-A Impacts on Earning and Employment at 18 Months*. Abt Associates, Bethesda.

BLUNDELL, R., AND M. COSTA DIAS (2002): "Alternative Approaches to Evaluation in Empirical Microeconomics," *Portuguese Economic Journal*, 1(2), 91–115.

Bundesministerium für Wirtschaft und Arbeit (2003): "Moderne Dienstleistungen am Arbeitsmarkt," http://www.bmwa.bund.de/Navigation/Arbeit/Arbeitsmarktpolitik/moderne-dienstleistungen-am-arbeitsmarkt,did=10144.html.

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.

Cochrane, W. (1968): "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society, Series A*, 128, 234–255.

Cochrane, W., and D. Rubin (1973): "Controlling Bias in Observational Studies," *Sankyha*, 35, 417–446.

Dann, S., A. Kirchmann, A. Spermann, and J. Volkert (2002): "Einstiegsgeld in Baden-Württemberg," Schlussbericht, Sozialministerium Baden-Württemberg, Stuttgart.

Davies, R., and S. S. Kim (2001): "Matching and the Estimated Impact of Interlisting," Isma Discussion Papers in Finance 2001-11, ISMA Centre, University of Reading.

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

Dehejia, R. H., and S. Wahba (1999): "Causal effects in Nonexperimaental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448), 1053–1062.

——— (2002): "Propensity Score Matching Methods for Nonexperimental Causal Studies," *The Review of Economics and Statistics*, 84(1), 151–161.

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," Working Paper, University of St.Gallen.

Heckman, J. (1997): "Instrumental Variables - A Study of the Implicit Behavioral Assumptions Used in Making Program Evaluations," *The Journal of Human Resources*, 32(3), 441–462.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.

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

Heckman, J., 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. Ashenfelter, and D. Card, pp. 1865–2097. Elsevier, Amsterdam.

Hujer, R., M. Caliendo, and S. Thomsen (2003): "New Evidence on the Effects of Job Creation Schemes in Germany - A Matching Approach with Threefold Heterogeneity," Discussion Paper No. 750, IZA, Bonn.

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 (2003): "What is the Value Added by Caseworkers," Discussion Paper No. 728, IZA.

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 Apr 2004.

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

Manski, C. (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 (1994): "The OECD Jobs Strategy: Enhancing the Effectiveness of Active Labour Market Policies," Report, Paris.

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," *Biometrica*, 70, 41–50.

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

Sianesi, B. (2001): "An Evaluation of the Active Labour Market Programmes in Sweden," Working Paper No. 2001:5, IFAU - Office of Labour Market Policy Evaluation.

——— (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 (2004): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estiamtors?," *Journal of Econometrics*.

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

A Data Sources and Attributes

Table A.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. A.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 (Maßnahmeteilnehmergrunddatei, MTG)
[2] Job-seekers data base (Bewerberangebotsdatei, BewA) and adjusted version for statistical purposes (ST4)
[3] Programme participants of subsidized employment data set (ST11TN)

B 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 B.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. B.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.

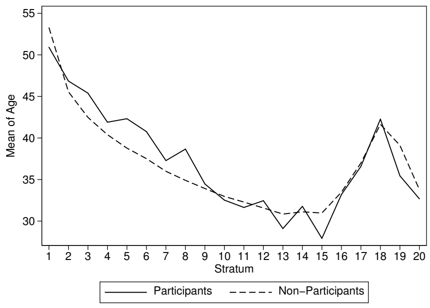
C Balancing of X within Strata

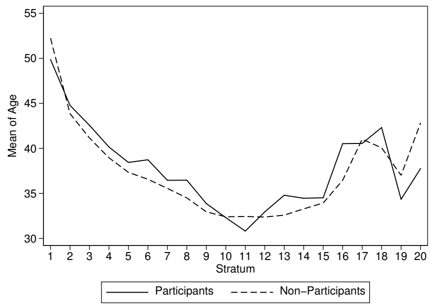
To provide valid estimates of the treatment effects for the treated with stratification matching, it has to be checked whether the covariates among the treatment and comparison groups within the strata are balanced. The following figures present the within-strata means of a selection of the incorporated covariates in the propensity score estimation from table 2 for participants and nonparticipants. The results for the other covariates are available on request by the authors.

Fig. C.1: Balancing of AGE within Strata

West Germany

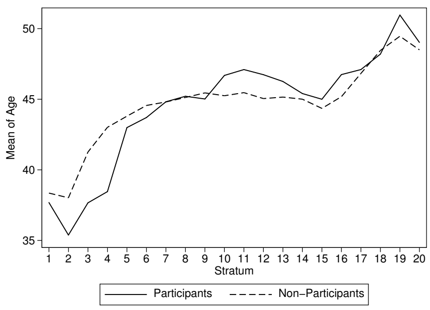
Men **Women**





East Germany

Men **Women**



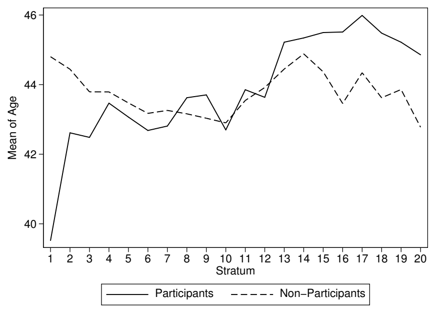
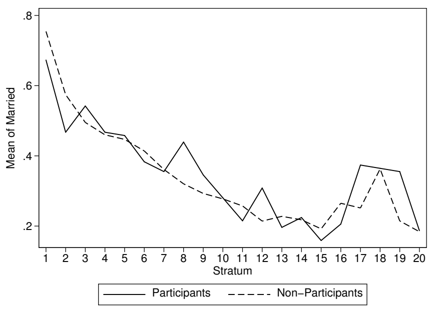
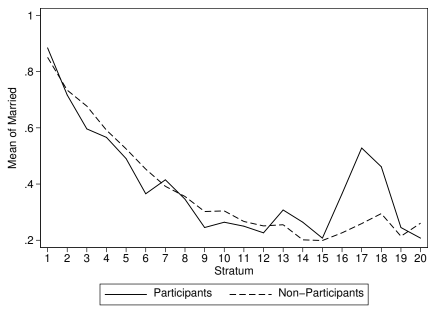


Fig. C.2: Balancing of MARITAL STATUS within Strata

West Germany

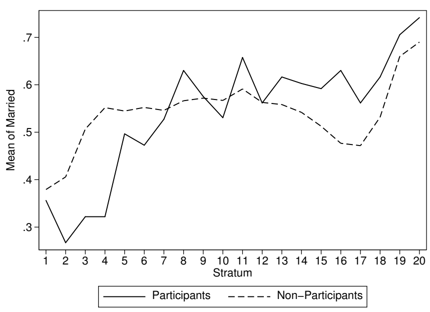
Men **Women**





East Germany

Men **Women**



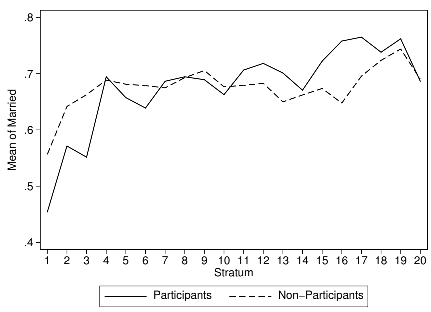
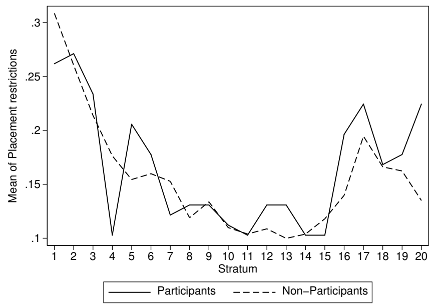


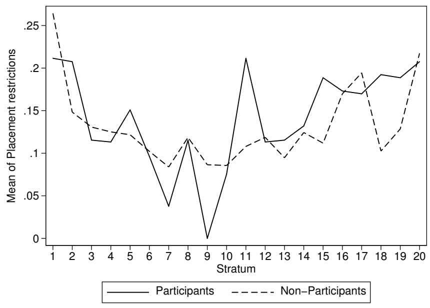
Fig. C.3: Balancing of PLACEMENT RESTRICTIONS within Strata

West Germany

Men

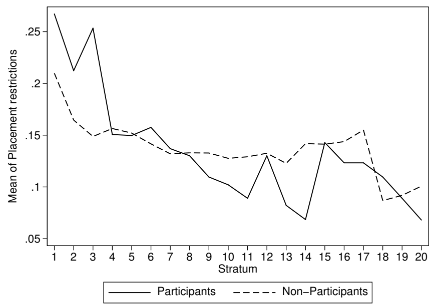


Women



East Germany

Men



Women

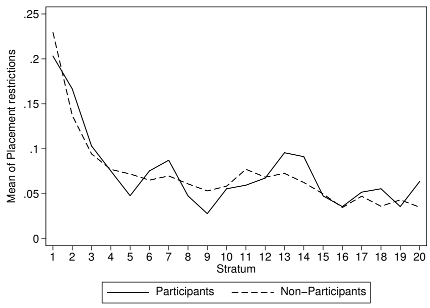
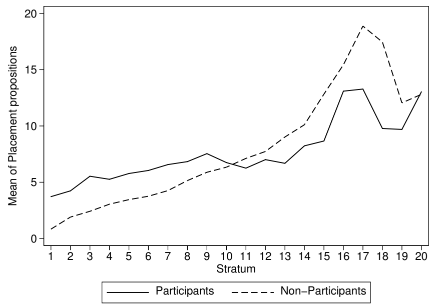


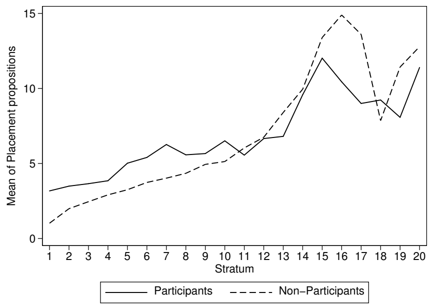
Fig. C.4: Balancing of NO OF PLACEMENT PROPOSITIONS within Strata

West Germany

Men

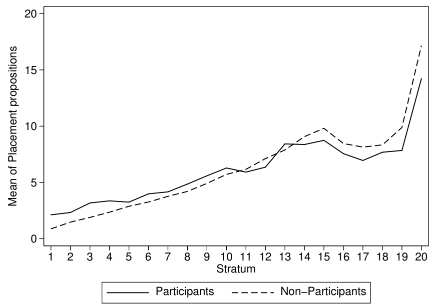


Women



East Germany

Men



Women

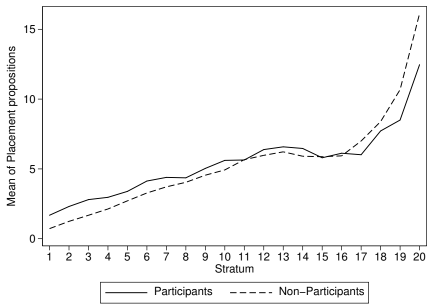
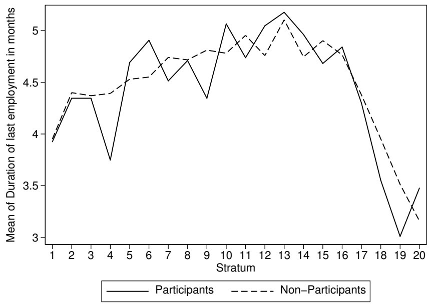


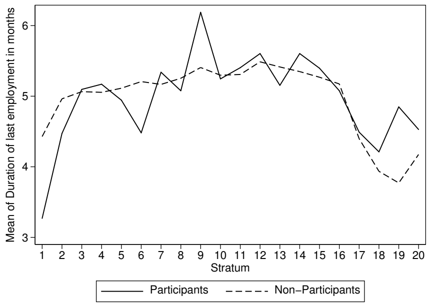
Fig. C.5: Balancing of DURATION OF LAST EMPLOYMENT within Strata

West Germany

Men

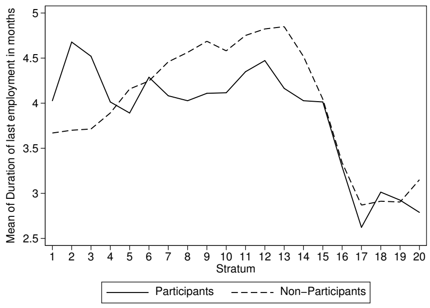


Women



East Germany

Men



Women

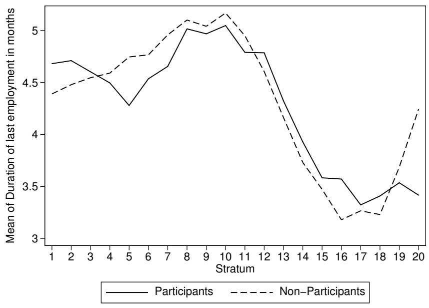
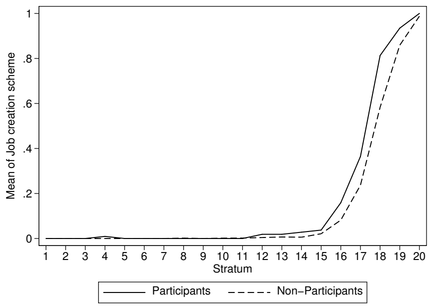


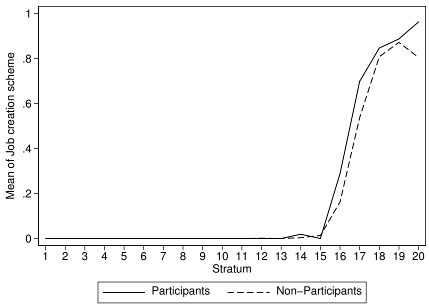
Fig. C.6: Balancing of JOB CREATION SCHEME within Strata

West Germany

Men

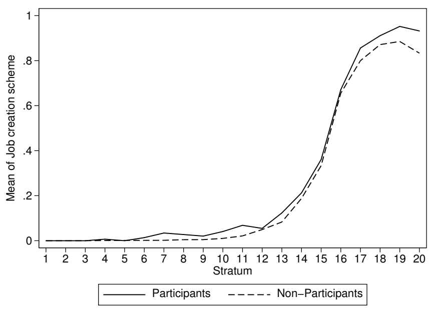


Women



East Germany

Men



Women