# Evaluation of Active Labour Market Policy: Methodological Concepts and Empirical Estimates

Reinhard Hujer and Marco Caliendo*

J.W.Goethe-University of Frankfurt/M.

Discussion Paper

**Abstract** Persistently high unemployment, tight government budgets and the growing scepticism regarding the effects of active labour market policies (ALMP) are the basis for a growing interest in evaluating these measures. This paper intends to explain the need for evaluation on the micro- and macroeconomic level, introduce the fundamental evaluation problem and solutions to it, give an overview of the newer developments in evaluation literature and finally take a look on empirical estimations of ALMP effects.

**Keywords:** Active Labour Market Policy, Evaluation, Matching, Difference-in-differences

**JEL Classification:** C14, C33, H43, J64, J68

*Reinhard Hujer is Professor of Econometrics and Statistics, Marco Caliendo is Research Assistant, University of Frankfurt. Corresponding Author: Reinhard Hujer, Department of Economics and Business Administration, Johann Wolfgang Goethe-University, Mertonstr.17, 60054 Frankfurt, Germany, hujer@wiwi.uni-frankfurt.de. The authors thank Christoph M. Schmidt for valuable comments. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

# Contents

# 1. Introduction

Most OECD countries have been plagued by high and persistent unemployment since the early 1970s. In recent years there has been a growing interest in active labour market policies as a means of fighting this unacceptable situation. This is easy to understand in view of the disillusionment with more aggregate policies. The traditional demand stimulation has been discredited because it faces the risk of increasing inflation with only small effects on employment and furthermore supply-side structural reforms, aimed to remove various labour market rigidities are difficult to implement or appear to produce results rather slowly.

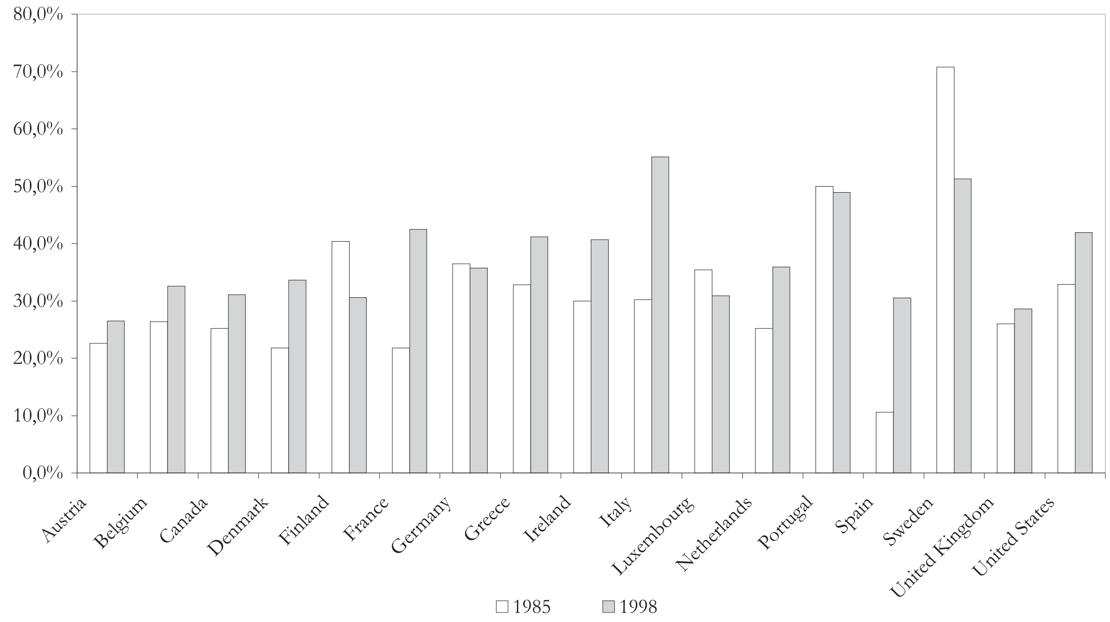
In this situation, like Calmfors (1994) notes, the active labour market policies (ALMPs) are regarded by many as the deus ex machina that will provide the solution to the unemployment problem. Not only that they provide a more efficient outcome on the labour market, they also equip individuals with higher skills and therefore lower the risk of poverty. In this sense ALMPs are capable to meet efficiency and equity goals at the same time (OECD (1991)). A distinction between active and passive programmes is best done by looking at the aims of the programmes. The goal of passive programmes is mainly to bridge the income shortage caused by unemployment ("Transferzahlungen"). The main categories in this sector are unemployment compensation and early retirements. In contrast to this, active programmes are intended to fight the structural problems of the labour market and to enhance the re-employment probability of (long-term) unemployed, respectively to avoid people to drift off in this group. Therefore ALMPs include public employment services and administration, labour market training, youth measures, subsidized employment and measures for the disabled.[1]

Alongside with the general tendency to deregulate markets and to promote work incentives, it has become a common theme in the political debate that governments should shift the balance of public spending on labour market policies away from passive income support towards more active measures designed to get the unemployed back into work. Especially Anglo-Saxon policy makers favour the idea of tying the right of welfare to the duty of work. Welfare then becomes workfare (Card (2000)). This should manifest itself in a higher relative importance of ALMPs. Figure 1 compares for some OECD countries the percentage of active policies on total spending for labour market policies in 1985 and 1998. There is a clear tendency of a rising share of active measure in nearly all the countries showing their growing importance. Nevertheless only Italy and Sweden directed more than 50 per cent of their total spending towards active measures in 1998. One obvious reason for the limited success in switching resources into active measures is the rising trend of unemployment in many countries. As unemployment benefits are entitlement programmes, i.e. rising unemployment automatically increases public spending on passive income support, most of the active labour market programmes are discretionary in nature and therefore easier disposable in a situation of tight budgets (Martin (1998)). The relation in Germany remains relatively constant. The german government spent about 45.3 bn DM on ALMPs, which is 33.5% of the total spending on labour market policies in 1999. Compared to 1985 with 36.5%, this is a slight fall, which might be due to the relatively generous passive labour market policies and the special situation regarding the situation in East Germany after reunification.

The most important measures have been vocational training ("Förderung der beruflichen Wei-

[1] See OECD (1993) for some standardized categories and sub-categories for labour market policies.

Figure 1: **Percentage of Active Policies on Total Spending 1985, 1998**$^{(a)}$




(a) Data for Germany from 1999, for the United Kingdom from 1997, for Denmark and Portugal from 1986.
*Source*: Staat (1997), OECD (1999)

terbildung", VT) which amounts to 9.8% of the total spending and traditional job creation schemes ("Arbeitsbeschaffungsmaßnahmen", JCS) amounting to 5.8%. These programmes absorb significant shares of national resources, which are than unavailable for alternative programmes or private expenditure. In an era of tight government budgets and a growing disbelief regarding the positive effects of ALMPs, evaluation of these policies becomes imperative. The ideal evaluation process can be looked at as a series of three steps. First, the impacts of the programme on the individual should be estimated. Second, it should be examined if the impacts are large enough to yield net social gains. Finally, it should be answered if this is the best outcome that could have been achieved for the money spent (Fay (1996)). However, evaluation in most OECD countries focuses on the first two steps, namely the microeconometric and the macroeconometric evaluation, and so will we.

Empirical microeconometric evaluation is conducted with individual data. The main question is if the interesting outcome variable for an individual is affected by the participation in an ALMP programme. Relevant outcome variables could be the future employment probability or the future earnings. We would like to know the difference between the value of the participants outcome in the actual situation and the value of the outcome if he or she had not participated in the programme. The fundamental evaluation problem arises because we never observe both states (participation and non-participation) for the same individual at the same time, i.e. one of the states is counterfactual. Therefore finding an adequate control group is necessary to make a comparison possible. This is not easy because the participants in programmes usually differ in more aspects than just participation from the non-participants. Taking simply the difference between their outcomes after training will not reveal the true training impact, i.e. will lead to a biased estimate. The literature of the solution to this problem is dominated by two points of view.

Some analysts like LaLonde (1986) or Ashenfelter and Card (1985) view social experiments as the only valid evaluation method, whereas a second group of researchers like Heckman and Hotz (1989) or Lechner (1998) believe that it is possible to construct a comparison group using non-experimental data and using econometric and statistical methods to solve the fundamental evaluation problem. In nonexperimental or observational studies, the data are not derived in a process that is completely under the control of the researcher. Instead one has to rely on information how individuals actually performed after the intervention, that is we observe the outcome with treatment for participants and the outcome without treatment for nonparticipants. The objective of observational studies is to use this information to restore the comparability of the both groups by design. To do so, more or less plausible identification assumptions have to be imposed. There are several approaches differing with respect to the methods applied for this problem. Some studies control for observables as part of parametric evaluation models, others construct matched samples. Furthermore some authors think that conditioning on observables is not enough and one has to take into account unobservables, too.

Besides the microeconometric analysis, there is also a growing amount of empirical studies using macroeconomic data, like Fay (1996), Bellmann and Jackman (1996a), Puhani (1999) or Hagen and Steiner (2000). This approach discusses the impact of active labour market policies not on particular individuals but on aggregate economic variables. Instead of looking at the effect on individuals' performance, the question is whether ALMPs represent a net gain to the whole economy. Most of the empirical studies on macroeconomic evaluation are concerned with estimating augmented matching functions, that is, relating outflows from unemployment to jobs to the stocks of vacancies and unemployed persons and thereby controlling for the level of ALMP intervention.[2] Although this is a very appealing approach because it takes into account the net impact of these schemes, some studies go further beyond this and present more numerous channels of operation. The model of Layard and Nickell (1986) has become a standard model for labour market analysis. In a slightly modified version it serves as a framework for the analysis of the effects of ALMPs on a number of economic variables or processes, that influence the aggregate employment and unemployment rates.

The remainder of this paper is organized as follows. In chapter 2 we will give a short overview of the legal basis and the evolution of ALMP in Germany. Chapter 3 will be concerned with the microeconometric evaluation approach, whereas chapter 4 considers the macroeconometric one. Finally in chapter 5 empirical estimates will be presented. Chapter 6 concludes.

## 2. Overview - ALMP in Germany

Labour market policies in Germany are organized by the Federal Employment Office ("Bundesanstalt für Arbeit"). Up to 1998 the legal basis for the labour market policy in Germany has been the work support act ("Arbeitsförderungsgesetz", AFG), founded in 1969. From there on, the new Social Code SGB III ("Sozialgesetzbuch") plays this role. Changes have been made not only in the

[2] To avoid confusion of terms, we would like to stress that the statistical matching method is in no way related to the macroeconomic concept of the augmented matching function.

objectives, like a more intensive focus on problem groups of the labour market, but also in the institutional organization of labour market policy, leading to decentralization and more flexibility in the regional allocation of resources to different measures. As a result of these changes, the comparability of the different measures has suffered a little bit. As most of the empirical studies which we present in chapter 5 are concerned with measures under the AFG, we will mainly focus on that act.

Staat (1997) gives a brief overview of AFG's historical evolution. The main goals of the AFG have been: (a) securance of a high employment ratio, (b) avoidance of low-quality employment, (c) improvement of the structure of the labour force, (d) promotion of mobility, (e) social goals and (f) promotion of target groups.[3]

The improvement of the labour force structure, i.e. the adjustment of the labour supply to the changing labour demand has been the primary goal in the early years. It was aimed to attend the continual growth of the economy, that changed the labour market conditions permanently, with a continuous adjustment of the labour force structure to fulfill the new requirements. In detail, a short supply in jobs with specific (high-level) skills as well as an excess supply in jobs requiring low skills only was to be avoided. This goals had to be revised quite soon. At the end of 1973 the sharp rising unemployment rate caused by the first oil price shock drove attention to the fight against this development. This becomes clear when we look at the participation structure of the measures. In the early 1970s less than 15 per cent of all participants have been unemployed before participation, whereas in the 1980s this was the case for almost 80 per cent.

After some innovations and amendments the AFG has been replaced by the SGB III in 1998. A good overview of the most relevant reforms can be found in Fitzenberger and Speckesser (2000). Sell (1998) presents an extensive discussion of the new SGB III, regarding especially the self-responsibility of employees for their own labour market success. Fertig and Schmidt (2000) explain and classify the different measures of employment promotion and distinguish explicitly between non-discretionary and discretionary measures.

Whereas the AFG has been implemented, under full employment conditions, the SGB III has been born in a rougher economic situation, where labour market policy is not so much dictated by the Minister of Labour but the Minister of Finance. Some of the AFG's objectives, like the securance of a high employment ratio and the avoidance of low-quality employment are dropped. The prominent goal is the (re-)integration of problem groups in the first labour market. As the government sees itself in a promoting role only, the SGB III places particular emphasis on the fact, that employees have to act on their own authority regarding their labour market success. This comes together with a tightening of the reasonableness-clause ("Zumutbarkeitsklausel"), which for example makes it harder for unemployed to turn down job offers.

Besides the change of the objectives there have been organizational changes, too, increasing the flexibility of ALMP on a regional and local level. The local employment offices are allowed to allocate their budgets relatively freely to different measures. This decentralization allows an adjustment to the situation on the local labour markets. Furthermore, 10% of the budget can be used for 'free promotion' ("Freie Förderung", §10, SGB III), allowing a more individualized support. Another promising feature is the so-called "Eingliederungsplan" to avoid long-term unemployment. Under this new plan, the local Labour Exchange and the unemployed have to establish any later

[3] See §§1,2 in the AFG for details.

than six month after the begin of the unemployment spell, which active measures or which action from the unemployed will help to avoid a drifting off into (long-term)unemployment.

Table 1: **Spending on Labour Market Policies in 1999**

| | Germany | | West | | East | |
|---|---|---|---|---|---|---|
| | Spending | | Spending | | Spending | |
| | in bn DM | % of total | in bn DM | % of total | in bn DM | % of total |
| Spending Federal Employment Office | 101.10 | | 61.89 | | 39.22 | |
| Spending Government(Bund) | 34.18 | | 21.36 | | 12.82 | |
| Total Spending | 135.29 | | 83.25 | | 52.04 | |
| A) Passive labour market policies | 81.19 | 60.0 | 53.31 | 64.0 | 27.88 | 53.6 |
| B) Active labour market policies | 45.30 | 33.5 | 22.98 | 27.6 | 22.32 | 42.9 |
| Support of vocational training | 13.20 | 9.8 | 7.77 | 9.3 | 5.43 | 10.4 |
| Traditional job creation schemes (ABM) | 7.81 | 5.8 | 2.14 | 2.6 | 5.66 | 10.9 |
| Structural adjustment schemes (SAM) | 5.05 | 3.7 | 0.39 | 0.5 | 4.66 | 9.0 |
| - SAM-East for private firms (SAMOfW) | 3.57 | 2.64 | 0.14 | 0.16 | 3.43 | 6.59 |
| Rehabilitation measures | 4.50 | 3.3 | 3.28 | 3.9 | 1.22 | 2.3 |
| Support of professional training | 2.76 | 2.0 | 1.66 | 2.0 | 1.10 | 2.1 |
| Crash programme against youth unemployment | 1.90 | 1.4 | 1.13 | 1.4 | 0.78 | 1.5 |
| Free Support | 1.09 | 0.8 | 0.50 | 0.6 | 0.59 | 1.1 |

*Source*: Bundesanstalt für Arbeit (1999), Arbeitsmarkt.

Table 1 shows the spending on labour market policies in Germany in 1999. 33.5% of the total spending have been dedicated to active measures, the relation in East Germany (42.9%) being much higher than in the West (27.6%). The most important measures have been the support of vocational training with 13.2 bn DM and subsidized employment, consisting of traditional job creation schemes with 7.8 bn DM and structural adjustment schemes ("Strukturanpassungsmaßnahmen", SAM) with 5.05 bn DM.

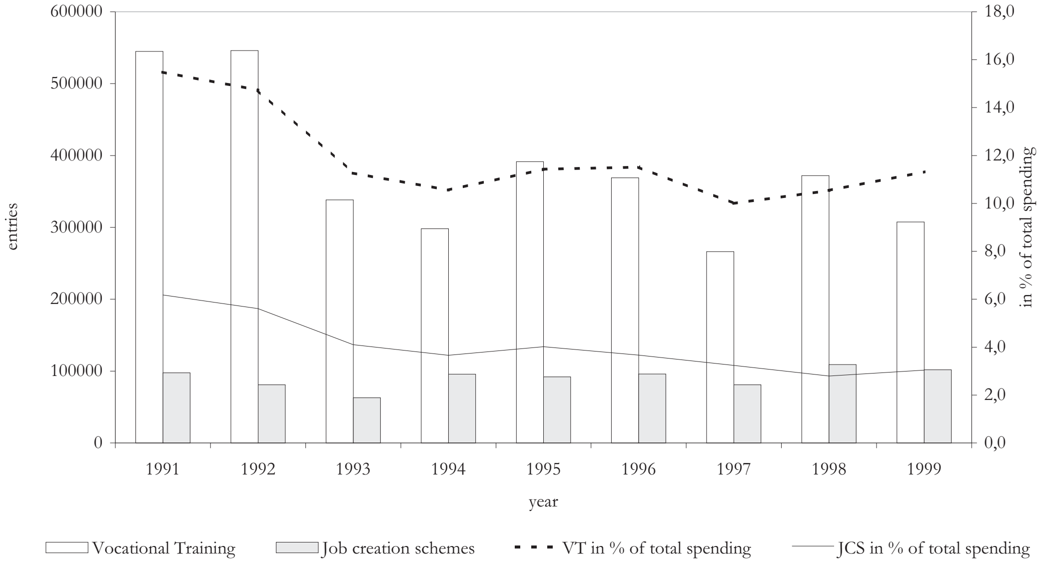
Hujer, Maurer, and Wellner (1999b) and Hujer and Wellner (2000b) provide an overview of vocational training under the AFG. The Federal Employment Office pays the costs of the training measures and a subsistence allowance ("Unterhaltsgeld") to the participants, which amounts to 60 or 67 per cent of the previous net wages (equal to unemployment benefit). In principle public vocational training comprises of three types of training measures. The first type is further training ("Fortbildung") in an occupation the participant is already trained in (§§41-46 AFG). A special focus is to improve the labour market conditions for disadvantaged individuals. Moreover further training is used to avoid supply shortages of labour with specific skills and to enhance advancement in an occupation ("Aufstiegsfortbildung"). On top of that further training includes short courses for the unemployed which aim to improve job search skills. The second type is retraining ("Umschulung") for a new occupation (§47 AFG). This is addressed to people whose qualifications have become useless, due to changing labour market conditions, e.g. technological changes. In a certain way they should help to improve the occupational mobility. The third and final type is training to familiarize with a new occupation ("Einarbeitung"). In this case, the labour offices support employers for providing training to those employees who need a long time to familiarize themselves with a new job (§49 AFG).

A good overview of subsidized employment programmes, consisting of traditional job creation schemes (JCS) and structural adjustment schemes (SAM) can be found in Bergemann, Fitzenberger, Schultz, and Speckesser (2000). JCS is the more important programme in West and East Germany. It includes limited employment for long-term unemployed in projects, which are of value for the society and additional in nature, that is without the subsidy they could not be executed. Even though JCS should be co-financed measures where between 30% and 75% of the costs are subsidies by the Federal Employment Office and the rest is paid by the implementing body (public or private legal entities, mainly municipalities), exceptions can be made, in the direction of an higher subsidy-quota. Priority is given to projects, which improve the chances for permanent jobs, that support structural improvement in social or environmental services or that aim at the integration of extremely hard-to-place individuals.

Especially in East Germany SAM play a prominent role. Their goal is analogous to JCS the integration into regular employment, but less severe eligibility criteria apply to participants, so that individuals who are only threatened by unemployment may participate already. The SAM consist of a wage subsidy equal to the amount of unemployment allowance or assistance the individual would have received if unemployment had continued. In East Germany the SAM may be implemented by public institutions and private companies ("SAM Ost für Wirtschaftsunternehmen", SAMOfW), whereas in West Germany only the first is possible.

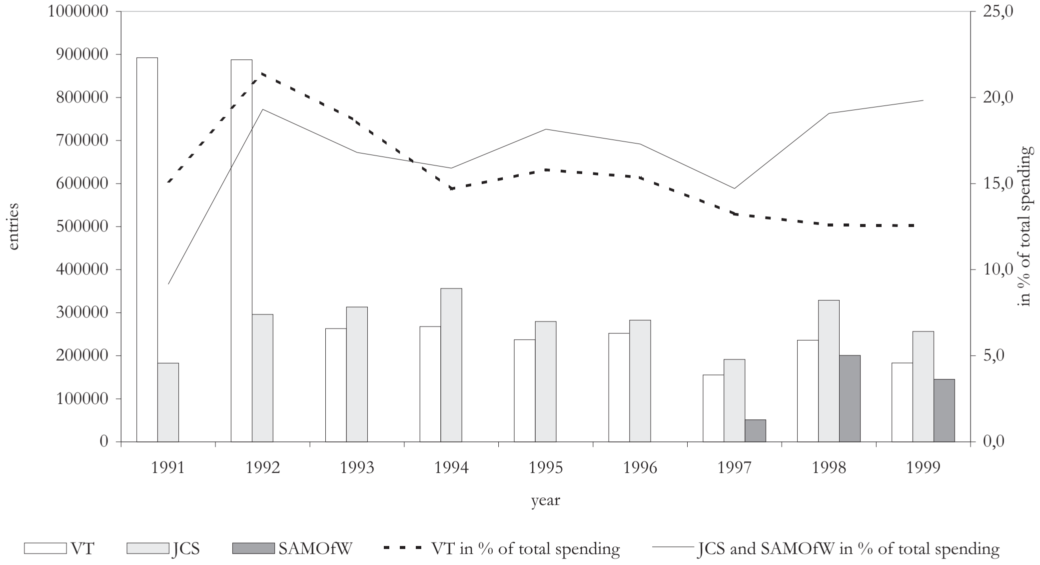
Figures 2 and 3 show the entries into vocational training and job creation schemes in West and East Germany from 1991 to 1999, as well as the spending on these measures as a percentage of the total spending. 1991 and 1992 have been the years with the most entries into vocational training. Consequently the relative spending has reached its highest level, too. The sharp decline in 1993 in entries as well as in spending mainly results from a reform of the AFG in 1993 which led to financial restrictions and institutional changes, so that the Federal Labour Office ("Bundesanstalt für Arbeit", BA) no longer supported training courses with very short duration. An interesting fact can be seen, if one compares the spending in the East and the West. The Federal Labour Office spent 1992 more than DM 10.8 billion in Eastern, but only DM 8.2 billion in Western Germany on vocational training, although the labour force in East Germany was nearly 75 per cent smaller. This was due to the special situation in East Germany which is well described in Fitzenberger and Prey (1997)) and Hujer and Wellner (2000b). After the reunification in 1990, the large effects of the transformation process on the labour market made public vocational training play an even more important role in the eastern part of Germany than in the western. The sudden exposure to a western-style economical environment and the loss of the main trading partners in the East led to a sharp reduction in production which changed the relative prices dramatically. Due to political pressure to reduce the disparity in living standards between both parts of Germany and to avoid massive East-West migration, the growth rate of gross earnings were pushed above which was considered economically to be the market-clearing levels, resulting in further growing unemployment. To combat this unfavourable situation the Federal Employment Office (BA) spent

Figure 2: **Entries in and Spending on Vocational Training and Job Creation Schemes in West Germany**$^{(a)}$




(a) Job creation schemes include traditional job creation and structural adjustment schemes as well as SAM-East for private firms.
*Source*: Bundesanstalt für Arbeit, Arbeitsmarkt, various issues.

Figure 3: **Entries in and Spending on Vocational Training and Job Creation Schemes in East Germany**$^{(a),(b)}$




(a) Job creation schemes include traditional job creation and structural adjustment schemes.
(b) SAMOfW stands for SAM-East for private firms.
*Source*: Bundesanstalt für Arbeit, Arbeitsmarkt, various issues.

an enormous amount of money, leading to the somehow strange situation described above.

For the time period from 1991-1999 vocational training has clearly been the most important programme in West Germany. For East Germany this is true only up to 1993. From that time onwards the number of entries into job creation schemes and the new created structural adjustment schemes outnumbered the entries into vocational training. As Kraus, Puhani, and Steiner (2000) note, JCS and SAM have not only been used as a means to keep people off the dole and to avoid social hardship associated with long-term unemployment, but also as an investment in the East German industrial infrastructure, such as the removal of environmental damages. Therefore it is easy to understand that these kinds of measures have been very important in East Germany. After a peak in 1994 the number of entries decreased due to policy changes and financial restrictions as did the entries into all important active labour market policies. The decline in the number of participants is not equally reflected in the development of expenditures on JCS and SAM. This is due to the fact, that the programmes became more cost-intensive when large scale programmes, so called 'Mega-JCS' and Societies for Employment Promotion and Structural Development ("ABS-Gesellschaften") were established (Kraus, Puhani, and Steiner (2000)). With the introduction of the SGB III in 1998, SAM could additionally be used by private firms (SAMOfW), leading to an even higher importance.

## 3. Microeconometric Evaluation Approach

Whereas macroeconometric approaches to ALMP evaluation are of fairly recent date, microeconometric evaluations have a slightly longer tradition, especially in the United States where evaluation has mostly focused on the earning effects of training programmes (Riddell (1991)). Although the budget for ALMPs in the United States has been relatively small compared to Europe, the evaluation efforts have been much higher. But the situation in Germany has improved in the last years. As a consequence of tight government budgets and the growing importance of ALMPs in the face of high and persistent unemployment the opinion gained acceptance that evaluation of these measures is imperative. Therefore, a lot of studies trying to evaluate labour market policies in Germany have evolved. In our empirical presentation of microeconometric evaluations we will focus on studies of vocational training and job creation schemes, because these are the most important measures. Before we do so we will present some methodological considerations. In microeconometric evaluation we use individual data and usually a large number of observations. This makes it easier to address the sample selection problem in microeconometric studies which arises from the fact that ALMP participants are a selected group and are thus not directly comparable to the non-participants. Therefore one has to find an adequate control group to solve the fundamental evaluation problem and make a comparison possible. In order to do so, it is necessary to account for any kind of selection bias. The recent evaluation studies differ with respect to the methods applied to solve this problem, especially regarding the question of considering observable and unobservable characteristics. Some studies apply parametric models to control for this, others construct matched samples non-parametrically. Bergemann, Fitzenberger, Schultz, and Speckesser (2000) doubt that controlling for observable characteristics is enough to solve the problem. There might be unobservables which influence both the selection process and the outcome variable, too. Therefore they suggest a 'conditional difference-in-differences method' which combines maching with the traditional difference-in-differences approach. Before we will present both methods, we discuss very briefly the suitable choice of an outcome measure and present the fundamental evaluation problem as well as solutions to it with experimental data.

## 3.1. Outcome Measures

The first step and one major element of programme evaluation is the suitable choice of an outcome measure, that is one has to clarify, what outcomes are to be considered, i.e. what should be defined as a success. Whereas evaluation in the United States has mainly focused on the earnings after training participation, in Europe the training effect on the employment situation plays a dominant role. This is not surprising as the unemployment rates in Europe are much higher and more persistent than in the United States. But the same qualitative outcome, say improvement in the employment situation, may plausibly be measured in several ways, for instance by hours per week in the new job or by a simple distinction between employment and unemployment. Another way, suggested by Hujer, Maurer, and Wellner (1999b), would be to look at the duration of unemployment (or employment) after the intervention has taken place. Schmidt (1999) notes, that there are some problems arising with the choice of an appropriate outcome measure. Outcomes may not be comparable across interventions. Therefore a policy maker who has to decide which measure to implement will normally try to translate the gains of a programme into money terms or to carry out a so called cost-utility analysis. Still this is not easy because new problems like time or group preferences emerge.

## 3.2. The Fundamental Evaluation Problem

Inference about the impact of a treatment on the outcome of an individual, involves speculation about how this individual would have responded, had he not received the treatment.[4] The framework serving as a guideline for the empirical analysis of this problem is the potential outcome approach, variously attributed to Fisher (1935), Neyman (1935), Roy (1951), Quandt (1972, 1988) or Rubin (1974), but most often it is just called the Roy-Rubin-model.

The main building blocks of this model are individuals, treatment (participating in a programme or not) and potential outcomes, that are also called responses. In the basic model there are two potential outcomes $(Y^T, Y^C)$ for each individual, where $Y^T$ indicates a situation with training and $Y^C$ without, i.e. the individual is then in the comparison group. To complete the notation we additionally denote variables that are unaffected by treatments - called attributes by Holland (1986) - by $X$. Attributes are exogenous in the sense that their potential values for the different treatment states coincide. Furthermore we define a binary assignment indicator $D$, indicating whether an individual actually participated in training ($D = 1$) or not ($D = 0$) (Hujer and Wellner (2000b), Lechner (2000)). The treatment effect for each individual is then defined as the difference between his/her potential outcomes:

$$\Delta = Y^T - Y^C \tag{1}$$

[4]This is clearly different from asking whether there is an empirical association between training and the outcome (Lechner (2000)). See Holland (1986) for an extensive discussion of concepts of causality in statistics, econometrics and other fields.

The fundamental problem of evaluating this individual treatment effect arises because the observed outcome for each individual is given by:

$$Y = D \cdot Y^T + (1 - D) \cdot Y^C. \tag{2}$$

Unfortunately we can never observe $Y^T$ and $Y^C$ for the same individual simultaneously. The unobservable component in (1) is called the counterfactual outcome, so that for individuals who participated in the training measure ($D = 1$), $Y^C$ is the counterfactual outcome, and for those who did not it is $Y^T$. In this sense the problem of evaluating the individual treatment effect is a missing data problem because for any given individual the counterfactual outcome cannot be estimated. The concentration on a single individual requires that the effect of the intervention on each individual is not affected by the participation decision of any other individual, i.e. the treatment effect $\Delta$ for each person is independent of the treatment of other individuals. In the statistical literature (Rubin (1980)) this is referred to as the stable unit treatment value assumption (SUTVA) and guarantees that average treatment effects can be estimated independently of the size and composition of the treatment population.[5] Note that there will be no opportunity to ever estimate individual gains with confidence. Therefore we have to concentrate on the population average of gains from treatment. The most prominent evaluation parameter is the so-called average treatment effect on the treated:

$$E(\Delta \mid D = 1) = E(Y^T \mid D = 1) - E(Y^C \mid D = 1) \tag{3}$$

Like Hujer and Wellner (2000b) note, this parameter gives an answer to the following question: "What is the expected, or mean outcome gain to individuals who received treatment to the hypothetical situation had they not received it?" This question focuses directly on actual training participants, so that it determines the realized gross gain from the training programme and can be compared with its costs. This will help to decide whether the programme is a success or not (Heckman, Ichimura, Todd (1997, 1998), Heckman, LaLonde, Smith (1999)). Despite the fact that most evaluation research focuses on average outcomes, partly because most statistical techniques focus on mean effects, there is also a growing interest regarding effects of policy variables on distributional outcomes. Examples where distributional consequences matter for welfare analysis include subsidized training programmes (LaLonde (1995)) or minimum wages (DiNardo, Fortin, and Lemieux (1996)). Koenker and Bilias (2000) show that quantile regression methods can play a constructive role in the analysis of duration (survival) data, too. They describe the link between quantile regression and the transformation model formulation of survival analysis, offering a more flexible analysis than conventional methods.

Nevertheless we will focus on the average treatment effect on the treated in this paper. The second term on the right side in equation (3) is unobservable as it describes the hypothetical outcome without treatment for those people who received treatment. If the condition

$$E(Y^C \mid \ D = 1) = E(Y^C \mid \ D = 0) \tag{4}$$

[5] Among other things SUTVA excludes cross-effects or general equilibrium effects. Its validity facilitates a manageable formal setup; nevertheless in practical applications it is frequently questionable whether it holds.

holds, we can use the non-participants as an adequate control group. In other words we would take the mean outcome of non-participants as a proxy for the counterfactual outcome of participants. This identifying assumption is definitely valid in social experiments. The key concept here is the randomized assignment of individuals into treatment and control groups. Individuals who are eligible to participate in training are randomly assigned to a treatment group which participates in the programme and a control group that does not. This assignment mechanism is a process that is completely beyond the workers' or the administrators' control. If the sample size is sufficiently large, randomization will generate a complete balancing of all relevant observable and unobservable characteristics across treatment and control groups. Therefore the comparability between experimental treatment and control groups is facilitated enormously. On average, the two groups do not systematically differ except for having participated in training. As a result the difference in the outcome after training is supposed to be solely induced by the programme itself, i.e. the impact of training is isolated and there should be no selection bias. Formally, random assignment ensures, that the potential outcomes are independent of the assignment to the training programme. We write:

$$Y^T, Y^C \amalg D, \tag{5}$$

$\amalg$ denoting independence. When assignment to treatment is completely random it follows that:

$$E(Y^C \mid D=1) = E(Y^C \mid D=0) \;\; and \;\; E(Y^T \mid D=1) = E(Y^T \mid D=0) \tag{6}$$

Therefore, treatment assignment becomes ignorable (Rubin (1974)) and we get an unbiased estimate of $E(\Delta)$, i.e. the randomly generated group of non-participants can be used as an adequate control group to consistently estimate the counterfactual term $E(Y^T \mid D=0)$ and thus the causal training effect $E(\Delta \mid D=1)$. Although this approach seems to be very appealing in providing a simple solution to the fundamental evaluation problem, there are also some problems associated with it. In practice, a randomized experiment may suffer from similar problems, that affect behavioral studies. Bijwaard and Ridder (2000) investigate the problem of non-compliance to the assigned intervention, that is, when members of the treatment sample drop out from the programme and members of the control group participate. If the non-compliance is selective, i.e. correlated with the outcome variable, then the difference of the average outcomes is a biased estimate of the effect of the intervention, and correction methods have to be applied, too. Besides relatively high costs and ethical issues concerning the use of experiments, further methodological problems might arise, like a substitution or randomization bias, which make the use of experiments questionable. For an extensive discussion of these topics the interested reader should refer to Burtless (1995), Burtless and Orr (1986) and Heckman and Smith (1995).

More important for practical application is in fact that in most European countries experiments are, out of several reasons, not conducted and researchers have to work with nonexperimental data anyway. In nonexperimental data, equation (4) will normally not hold:

$$E(Y^C \mid D=1) \neq E(Y^C \mid D=0) \tag{7}$$

The use of the non-participants as a control group will therefore lead to a selection bias. Heckman and Hotz (1989) point out that selection might occur on observables or unobservables.

We will present in the following sub-chapters for each case an estimation approach that tries to estimate the unobserved counterfactual term using the observed outcome information of the non-participants. We will start with the matching approach, before we will present a conditional difference-in-differences estimator, suggested by Heckman, LaLonde, and Smith (1999) and recently applied for East Germany by Bergemann, Fitzenberger, Schultz, and Speckesser (2000).

## 3.3. Selection on Observables - Matching Approach

Matching is in one of the most appealing non-experimental approaches to solve the fundamental evaluation problem. That is because it shows a close link to the experimental context. The basic idea underlying the matching approach is to search from a large group of non-participants those individuals who are similar to the participants in all relevant pre-training characteristics. That being done, the differences in the outcomes between the well selected and thus adequate control group and the trainees can then be attributed to the programme. The matching approach originated in the statistical literature (see Rubin (1974), (1977), (1979), Rosenbaum and Rubin (1983), (1985a), (1985b) or Lechner (1998)).

Of course matching is first of all plagued by the same problem as all non-experimental estimators, which means that assumption (4) cannot be expected to hold when treatment assignment is not random. However, following Rubin (1977) treatment assignment may be random given a set of covariates. The construction of a valid control group via matching is based on the identifying assumption that conditional on all relevant pre-training covariates ($Z$), the potential outcomes ($Y^T, Y^C$) are independent of the assignment to training. This so called conditional independence assumption (CIA) can be written formally as:

$$Y^T, Y^C \amalg D \mid Z.^6 \tag{8}$$

If assumption (8) is fulfilled we get:

$$E(Y^C \mid Z, D=1) = E(Y^C \mid Z, D=0) = E(Y^C \mid Z) \tag{9}$$

Similar to randomization in a classical experiment, the role of matching is to balance the distributions of all relevant pre-treatment characteristics $Z$ in the treatment and control group, and thus to achieve independence between the potential outcomes and the assignment to treatment, resulting in an unbiased estimate.[7] The implementation of conditioning on all relevant covariates is, however, limited in case of a high dimensional vector $Z$. For instance, if $Z$ contains $n$ covariates which are all dichotomous, the number of possible matches will be $2^n$. In this case cell matching, that is exact matching on $Z$, is mostly not possible, because an increase in the number of variables increases the number of matching cells exponentially (Hujer and Wellner (2000b)). To deal with this dimensionality problem, Rosenbaum and Rubin (1983) suggest the use of balancing scores $b(Z)$, i.e. functions of the relevant observed covariates $Z$ such that the conditional distribution of $Z$ given $b(Z)$ is independent of the assignment to treatment, that is $Z \amalg D \mid b(Z)$ holds.

[6] For the purpose of estimating the mean effect of treatment on the treated the assumption of conditional independence of $Y^C$ is sufficient, because we like to infer estimates of $Y^C$ for persons with $D=1$ from data on persons with $D=0$ (Heckman, Ichimura, and Todd (1997)).

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

For trainees and non-trainees with the same balancing score, the distributions of the covariates $Z$ are the same, i.e. they are balanced across the groups. Moreover Rosenbaum and Rubin (1983) show that if the treatment assignment is strongly ignorable when $Z$ is given, it is also strongly ignorable given any balancing score.[8] The propensity score, i.e. the probability of participating in a programme is one possible balancing score. It summarizes the information of the observed covariates $Z$ into a single index function. Rosenbaum and Rubin (1983) show how the conditional independence assumption extends to the use of the propensity score so that

$$Y^C \amalg D \mid P(Z) \tag{10}$$

Therefore we get:

$$E(Y^C \mid P(Z), D=1) = E(Y^C \mid P(Z), D=0) = E(Y^C \mid P(Z)), \tag{11}$$

which allows us to rewrite the crucial term in the average treatment effect (3) as:

$$E(Y^C \mid D=1) = E_{P(Z)}[E(Y^C \mid P(Z), D=0 \mid D=1] \tag{12}$$

Hujer and Wellner (2000b) note that the outer expectation is taken over the distribution of the propensity score in the treated population. The major advantage of the identifying assumption (10) is that it turns the estimation problem into a much easier task since one only has to condition on a univariate scale, i.e. on the propensity score. When $P(Z)$ is known the problem of dimensionality can be eliminated. The evaluation of the counterfactual term via matching on the basis of the group of non-participants then only requires us to pair participants with non-participants which have the same propensity score. This insures a balanced distribution of $Z$ across both groups. Unfortunately $P(Z)$ will not be known a priori so it has to be replaced by an estimate. This can be achieved by any number of standard probability models. Dehejia and Wahba (1995) decide to use a logistic model to estimate the propensity score, whereas Hujer, Maurer, and Wellner (1999b) apply a panel probit model.[9] Clearly the empirical power of matching to reduce the problem of selection bias relies crucially on the quality of the estimate of the propensity score on the one hand and on the existence of comparison persons that have equal propensity scores as the treated persons. If the latter is not ensured we face the risk of incomplete matching with biased estimates.[10]

## 3.4. Selection on Unobservables - Difference-in-Differences

Bergemann, Fitzenberger, Schultz, and Speckesser (2000) claim, that controlling for selection on observables may not be sufficient since remaining unobservable differences might still lead to a biased estimation of treatment effects. These differences might arise from differences in the benefits

[8]Strongly ignorable means that assumption (8) holds and: $0 < p(Z) \equiv P(D=1 \mid Z) < 1$. The latter ensures, that there are no characteristics in $Z$ for which the propensity score is zero or one. Proofs go beyond the scope of this work and can be found in Rosenbaum and Rubin (1983).

[9]The approach of Hujer, Maurer, and Wellner (1999b) is a special case because it accounts for the possibility of time-varying covariates.

[10]Matching was much discussed in the recent econometric literature. Heckman and his colleagues reconsidered and further developed the identifying assumptions of matching stated by Rubin (1977) and Rosenbaum and Rubin (1983). It turns out that the new identifying assumptions are weaker compared to the old which brings along some advantages. Presenting these ideas goes beyond the scope of this work. The interested reader should refer to Heckman, Ichimura, Smith and Todd (1996, 1998), Heckman, Ichimura and Todd (1997,1998) and Heckman and Smith (1995).

which individuals expect from participation in a treatment that might influence their decision to participate. Furthermore some groups might exhibit bad labour market prospects or differences in motivation. These things are unobservable to a researcher and might cause a selection bias.

To account for selection on unobservables, Heckman, LaLonde, and Smith (1999) suggest econometric selection models and difference-in-differences (DiD) estimators. The DiD-estimator requires access to longitudinal data and can be seen as an extension to the classical before-after estimator (BAE). Whereas the BAE compares the outcomes of participants after they participate in the programme with their outcomes before they participate, the DiD-estimator eliminates common time trends by substracting the before-after change in nonparticipant outcomes from the before-after change for participant outcomes. The DiD-estimator is based on the assumption of time-invariant linear selection effects. The critical identifying assumption of this method is, that conditional on individual characteristics $X$, the biases are the same on average in different time periods before and after the period of participation in the programme, so that differencing the differences between participants and nonparticipants eliminates the bias (Heckman, Ichimura, Smith, and Todd (1998)). Let $t$ be a post-programme period and $t'$ a pre-programme period. The method identifies the mean effect of treatment on the treated if:

$$B_t - B_{t'} = 0, \tag{13}$$

where $B_t$ denotes the bias in time $t$ and is defined as:

$$B = E(Y^C \mid X, D = 1) - E(Y^C \mid X, D = 0). \tag{14}$$

Bergemann, Fitzenberger, Schultz, and Speckesser (2000) apply a combination of the matching- and the DiD-estimator in a recent paper, by implementing a 'conditional difference-in-differences estimator', conditional meaning that treatment and control group are already partly comparable conditional on $Z$. The change for the treated individuals is contrasted to the change for comparable non-treated individuals and any effect of selection on unobservables can be differenced out. The DiD-estimator can be written as:

$$[\mathrm{Y}^T_{i,t'} - \mathrm{Y}^C_{i,t}] - [\mathrm{Y}^C_{j,t'} - \mathrm{Y}^C_{j,t}] \tag{15}$$

Its validity relies crucially on assumption (13), that is the time-invariant nature of the selection effect. Only if it remains constant and can be cancelled out an unbiased estimate will be achieved. A good example of a case in which this assumption is likely to be violated can be found in Ashenfelter (1978). He observed that participants experience a decline in their earnings prior to the enrollment in a training programme. Later research demonstrates that the so-called 'Ashenfelters's Dip' is a common feature of the pre-programme earnings and applies to employment chances, too (Heckman, LaLonde, and Smith (1999)). If the 'dip' is transitory and the dip is eventually restored even in the absence of participation in the programme, the bias will not average out. Therefore Kluve, Lehmann, and Schmidt (1999) suggest to use the pre-treatment (labour market) histories of the individuals as an important variable in the matching process, so that only individuals with identical pre-treatment histories are compared.

### 3.5. Heterogenous and Multiple Treatment

The standard model of the microeconometric evaluation literature that we have presented so far allows an individual to choose between two states only, like participation in a training programme or non-participation. However, when evaluating labour market programmes a more complex framework seams to be necessary. Our presentation of ALMPs in Germany has shown that a wide variety of different measures exists. These measures often differ with respect to their target population, their contents and duration, their selection rules as well as with respect to their effects. The standard models do not account for this. Lechner (1999a, 1999b) and Imbens (1999) proposed an extension of the two-potential-outcome model to the case of heterogeneous states or treatments, that extends the actual choice set of individuals to more than just two options. Instead of just two potential outcomes $\{Y^C, Y^T\}$, that is treatment or not, now $(M+1)$ different mutually exclusive treatments are considered, denoted by $\{Y^0, Y^1, Y^2, \ldots, Y^M\}$. This method takes into account the heterogeneity of the different programmes and is therefore a valuable extension of the standard approach.

A further problem which might arise in the evaluation context might be multiple treatment, in the sense that several treatments take place during a short period of time. Bergemann, Fitzenberger, Schultz, and Speckesser (2000) found this effect to be relevant especially in East Germany. They expect that the effectiveness of each participation might vary depending on the participants' history in ALMP programmes. They estimate the causal effect for each separate treatment using conditional DiD estimators.

## 4. Macroeconometric Evaluation Approach

In this chapter we will discuss the impact of active labour market policies not on particular individuals but on aggregate economic variables. Instead of looking at the effect on individuals' performance we would like to know if the ALMPs represent a net gain to the whole economy. If the total number of jobs is not affected by labour market policies, the effects will be distributional only. This might be desirable, e.g. if work is shifted from the old to the young, but can hardly justify the substantial fiscal costs of the ALMPs. Another situation is given if one thinks that ALMPs can alter the overall number of jobs in the economy and hence affect both unemployment and output. Following this view there would be scope for aggregate economic benefits and fiscal payoffs, e.g. arising from increased tax revenues or decreasing spending for unemployment benefits, allowing a direct comparison between costs and benefits of the programmes. Clearly, like Bellmann and Jackman (1996a) note, from an economic standpoint it is possible for ALMPs to increase output only if one starts from a position where there are market imperfections. Departing from the traditional Keynesian perspective, according to which the level of unemployment is simply determined by the level of aggregate demand in the economy, they discuss the search, the job-rationing and the structural model, trying to identify the nature of imperfections in order to assess whether interventions can counteract them. Following their discussion is behind the scope of this paper. For an extensive discussion of these topics the reader is referred to Layard, Nickell, and Jackman (1991), or Franz (1999).

Bellmann and Jackman (1996b) note, that empirical work on the macroeconomic effects of ALMPs is rare and they conclude, that one obstacle is the absence of an obvious theoretical framework within which to couch the analysis. Leaving the traditional way of "cheating the Phillips curve", i.e. improving the inflation-unemployment trade-off and thereby reducing the nonaccelerating inflation rate of unemployment (Baily and Tobin (1977)), aside, the Layard and Nickell (1986) model has become the standard framework for labour market analysis. Calmfors (1994) shows how it can be developed to allow an analysis of ALMPs. He presents the 'revised Layard-Nickell model' as a basic framework for the analysis of the effects of ALMPs on a number of economic variables and processes that influence the aggregate employment and unemployment rates. A major advantage of this model is, that it distinguishes explicitly between regular employed people, openly unemployed people and participants in labour market programmes. This allows to separate the gross or bookkeeping effect of increased programme participation, which is a reduction in open unemployment, from the net effect. The latter takes into account to what extent the gross impact on open unemployment is reinforced or counteracted by indirect effects. This may include effects on the process of wage setting, the regular labour demand and the effective labour force (Calmfors (1994)). We will focus on the most prominent indirect effects, which are deadweight losses, substitution and displacement effects.

## 4.1. Deadweight Losses, Substitution and Displacement Effects

The so-called deadweight losses and substitution effects have received substantial attention in the literature (Layard, Nickell, and Jackman (1991), or OECD (1993)), mainly in the context of job creation schemes. If the outcome of the programme is no different from what would have happened in its absence we talk about a deadweight loss. A common example is the hiring from the target group that would have occurred also without the programme. If a worker is taken on by a firm in a subsidized job instead of an unsubsidized worker who would have been hired we talk about a substitution effect. The net short-term employment effect in this case is zero. Calmfors defines "the substitution effect as the extent to which jobs created for a certain category of workers simply replace jobs for other categories, because relative wage costs are changed."[11] Such effects are likely in the case of subsidies for private-sector work. There is always a risk that the employers hold back ordinary job creation in order to be able to take advantage of the subsidies. In order to minimize this danger, a principle of additionality may be imposed.

Another problem might be, that active labour market programmes may crowd out regular employment. This can be seen as a generalization of the so called displacement effect. This effect typically refers to displacement in the product market, e.g. if firms with subsidized workers may increase output, but displace (reduce) output among firms who do not have subsidized workers. Clearly, these effects have to be taken into account, if one likes to make statements about the net effect of ALMP.

## 4.2. The Matching Function Approach

Many macroeconometric studies have been based on matching functions, that is relating the hiring of workers to a minimum set of determining variables (Puhani (1999)). Usually it is distinguished between ordinary and augmented matching functions, whereat the latter takes into account the heterogeneity of the pool of unemployed people. Often attention is drawn to the returns of scale in

[11] Calmfors (1994), p.17

the matching process. Constant returns to scale in the matching function mean that if the labour market were twice as large in terms of unemployed people and vacancies, the numbers of matches would be twice as large, too. Increasing (decreasing) returns to scale would mean that there are positive (negative) externalities with respect to the size of the labour market, which increase or decrease the matching efficiency. Formally the matching function can be written as:

$$h = \alpha \cdot m(U, V), \tag{16}$$

$h$ being new hiring (matches, outflows from unemployment into employment), $U$ and $V$ being the stocks of unemployed persons and vacancies. $m$ is a continuous non-constant function, whereas $\alpha$ is the mismatch parameter (the higher $\alpha$, the smaller the mismatch).[12] A better match between labour demand and supply leads, ceteris paribus, that is with the same level of unemployed and vacancies, to higher outflows from unemployment into employment. In a recent study, Hagen and Steiner (2000) take different labour market programmes into account and estimate the augmented matching function in the following way (log-linear notation):

$$ln(h) = \beta_1 \cdot ln(U) + \beta_2 \cdot ln(V) + \beta_3 \cdot ln(JCS) + \beta_4 \cdot ln(SAM) + \beta_5 \cdot ln(VT) + \epsilon.^{13} \tag{17}$$

A theoretical justification for this approach can be seen in the search efficiency theory, where the expected effect of active labour market policy is an increase in the search efficiency of unemployed people, e.g. by improving their job-applications or supporting their mobility. This shortens the search duration and the matching efficiency rises (Bellmann and Jackman (1996a)). Most matching approaches are based on a similar regression equation and some empirical examples will be presented in chapter 5.3.

A major weakness of the matching function approach is that it does not consider job-to-job transitions, i.e. it assumes that all vacancies are filled only with unemployed people. Furthermore the available data does not allow to make a difference between outflows from unemployment into employment or into out-of-the-labour force.

But these are not the only problems which are likely to arise in macroeconomic assessment of active labour market policies. As we work with aggregated data the results tend to get vague and less robust. Regarding the high degree of heterogeneity in the conception, design and implementation of active labour market programmes we would need more detailed data to evaluate the effects properly. Instead we have to deal in many cases with relatively crude data, making the use of proxy variables necessary.

A further major problem for macroeconomic empirical analysis is the endogeneity or simultaneity problem. Given that governments react to rising unemployment or other labour market problems with increased policy effort, it becomes very hard to disentangle the effect of policy on the labour market. Basically, the expenditures on ALMPs can affect the unemployment rate, and it might be equally the case that the level of unemployment affects the spending on ALMPs (OECD

[12] $\alpha$ may capture a variety of factors like occupational or geographical mismatch and the efficiency of labour market institutions.

[13] The dependent variable in their model is either the outflow from or the inflow in unemployment. Furthermore they allow for seasonal effects as well as a time trend and they take into account only lagged values of the programme participation (to reduce the simultaneity problem, which will be discussed below). $\epsilon$ covers all the unobserved influence factors. Their estimation is based on monthly data of 176 local labour office districts ("Arbeitsamtbezirke").

(1993)). A solution to this problem might be to invoke some behavioural assumptions for the government concerning the spending on ALMPs. These assumptions have to be tested then. If a study does not account for the simultaneity problem, results have to be treated with great care.

# 5. Selected Empirical Results

In this chapter we will present micro- and macroeconometric evaluation studies of vocational training and job creation schemes for East and West Germany. Whereas the number of microeconometric evaluation studies of vocational training has been growing fast in the last years, the evaluations of job creation schemes are relatively small and so are the macroeconometric studies, too. Similar overviews of evaluation studies can be found in Fitzenberger and Speckesser (2000) and Hagen and Steiner (2000). We try to order the studies according to their results, although this is not always easy because some studies find contrary effects for different sub-groups.

## 5.1. Microeconometric Evaluations of Vocational Training

All evaluation studies of vocational training for West Germany are based on the German Socio-Economic Panel (GSOEP)-West and can be found in Table 3 in the appendix. The period under consideration ranges from 1984 to 1994. With respect to the design of the programmes, measures on-the-job and off-the-job are as well considered as measures with and without income maintenance. The applied evaluation methodologies include discrete hazard rate models, matching, IV-estimation and simultaneous dynamic models. Beside unemployment duration and the re-employment probability, hourly wages and the employment stability are considered as outcome variables. Hujer, Maurer, and Wellner (1998) and Hujer, Maurer, and Wellner (1999c) find that participation in vocational training has significant reducing effects on the unemployment duration. In further studies, Hujer, Maurer, and Wellner (1999b) and Hujer and Wellner (2000a) discover positive effects only for short courses ($<$ 6 months), whereas long courses do not have (significant) positive effects. Hujer, Maurer, and Wellner (1999a) get the result, that on-the-job training has no significant effects on the unemployment duration, whereas off-the-job training reduces it in the short-run and has no significant effects in the long-run. This finding corresponds to Pannenberg (1995) who observes that participation in off-the-job training increases the re-employment probability in the short run. In opposite to this positive findings, the following studies get rather negative results. Prey (1997) examines vocational training with income maintenance and finds negative (no) effects for men (women) on the employment probability. In a further study Prey (1999) gets negative (no) effects on the employment probability for measures with (without) income maintenance and no effects on the wages. Staat (1997) studies public sector training with income maintenance. His results indicate positive effects on the search duration only for sub-groups, no significant effects on the employment probability but positive effects on the wages.

The studies for East Germany, presented in Table 4, are either based on the East German Labour Market Monitor (EGLMM), the GSOEP-East or the Labour Market Monitor Sachsen-Anhalt, a panel based on the population of the state of Sachsen-Anhalt. Similar to the studies for West Germany training on- and off-the-job as well as with and without income maintenance is analysed. The period under consideration ranges from 1989 to 1998. As outcome variables the (un)employment duration, (stable and unstable) employment (probabilities), job search, working time and wages are examined. Fitzenberger and Prey (1997) find that training outside the firm shows considerable positive effects on the employment probabilities, whereas training inside has negative effects. Hübler (1998) gets the result that on-the-job training increases the job security, whereas off-the-job training leads to higher earnings. But this is true only for privately financed training. Publicly financed training has only positive effects in the short-run. Fitzenberger and Prey (2000) inquire training supported by public income maintenance outside a firm and discover positive effects for employment and earnings (but only a few significant in the long-run). Pannenberg (1995, 1996) ascertains positive effects on the re-employment probability and the income for vocational training on- and off-the-job. Lechner (1999b) examines enterprise-related continuous VT and gets positive income effects, too, but no effects on the unemployment probability. Kraus, Puhani, and Steiner (1999) observe for a sub-period of their study positive effects on (stable) employment for on- and off-the-job training. Hübler (1994) examines on-the-job qualifying measures and gets the result, that training induces search activities and reduces the effective hours of work.

A lot of studies do not find any clear positive or negative effects. Bergemann, Fitzenberger, Schultz, and Speckesser (2000) examine multiple participation in further training. They do not discover positive effects for first training programmes, and the additional effects of a second participation are on average not different from zero. Hujer and Wellner (2000b) find no significant effects for public-sector sponsored VT on the unemployment duration, but very weak hints that short courses seem to be more effective reducing the duration. Lechner (1999a) investigates off-the-job training (publicly financed and enterprise related). He can not establish robust positive effects on either the employment probability or the earnings. Staat (1997) examines public sector training with income maintenance and finds no effects neither on the search duration, the employment stability or the level of hourly wages.

The results of the following studies tend to be negative. Fitzenberger and Prey (1998) examine the effects of training within and outside the firm and with and without public income maintenance on employment and wages. Most often they get negative or no effects, differing with respect to the specification. Lechner (2000) finds no positive long term effects on the employment probabilities or the earnings for public-sector-sponsored VT. He gets negative results, regarding the risk of unemployment in the short-run.

## 5.2. Microeconometric Evaluations of Job Creation Schemes

Table 2 contains the few existing evaluation studies of job creation schemes for East Germany. All studies are either based on the EGLMM or the Labour Market Monitor of Sachsen-Anhalt. As evaluation methodologies discrete hazard rate models, nonparametric DiD-methods IV-estimators, linear control functions and the matching approach are used. The outcome variables considered include unemployment probabilities, transition rates to regular employment as well as (re)employment probabilities and (un)stable employment. The only study which finds clearly positive effects of public employment programmes is the one from Eichler and Lechner (1999). Their results indicate a substantially reduced unemployment risk for participants. Kraus and Steiner (1995) observe positive effects on the transition rates of regular employment only for men and only in the long-run, whereas the effects for women are negative. Bergemann, Fitzenberger, Schultz, and Speckesser (2000) discover that the long-run effects of JCS are at best insignificant. This corresponds to a study from Hübler (1997) which gets the result of no positive effects of public work programmes. Contrary to this, Kraus, Puhani, and Steiner (2000) ascertain clearly negative effects on the re-employment probabilities for the participants of public work programmes.

### 5.3. Macroeconometric Evaluations

Table 5 contains the macroeconometric evaluations for West and East Germany. Büttner and Prey (1998) use the data from 74 planning regions of West Germany to evaluate training programmes and public sector job creation. They find that job creation schemes reduce the mismatch, whereas training programmes do not have any significant effects. Prey (1999) extends this work by additionally controlling for the regional age structure and recipients of social assistance and estimating separately for men and women. She finds that VT increases (decreases) the mismatch for women (men), whereas JCS decreases the mismatch for men.

Pannenberg and Schwarze (1998) use the data from 35 local labour office districts to evaluate training programmes in East Germany. They find that the programmes have negative effects on the regional wages. Schmid, Speckesser, and Hilbert (2000) use the data from 142 local labour office districts to estimate the effects of further training, retraining, public sector job creation and wage subsidies on long term unemployment (LTU) and the unemployment outflow. They find that JCS reduces only 'short' LTU, VT reduces LTU and wage subsidies help only the very long-term unemployed. Steiner, Wolf, Egeln, Almus, Schrumpf, and Feldotto (1998) examine the effects of vocational training on the labour market mismatch, using the data from 35 local labour office districts in East Germany. They get observe only very small effects on the matching efficiency which disappear in the long-run. Hagen and Steiner (2000) evaluate VT, JCS and SAM for East and West Germany using the data from local labour office districts. The estimated net-effects are not very promising as all measures increase unemployment in West Germany. Only SAM reduces the unemployment rate slightly in East Germany, whereas JCS and VT increase it, too.

## 6. Conclusion

Persistently high unemployment, tight government budgets and the growing scepticism regarding the effects of active labour market policies (ALMP) are the basis for a growing interest in evaluating these measures. We explained in this paper the need for evaluation on the micro- and macroeconomic level, introduced the fundamental evaluation problem as well as solutions to it, gave an overview of the newer developments in evaluation literature and finally took a look on empirical estimations of ALMP effects. Our presentation of microeconometric evaluations of vocational training and job creation schemes in Germany allows no definite conclusions, as there is significant variation regarding the estimated effects of these measures. Our overview has shown that the different studies apply a wide variety of evaluation methodologies, so that some of the occurring variation might be due to this. But what seems to be the major problem of evaluation in Germany is the fact that the available data is rather crude. Therefore the participants- and the programme-heterogeneity could not be taken properly into account. Especially regarding the latter, improvements have to be made and better data has to be made accessible, to allow a distinction of the courses by their contents and their aims or target groups. Furthermore, more data on the individuals' characteristics before and after the training measure has taken place, would allow researchers to estimate more precisely the true training impact.

## References


Ashenfelter, O. (1978): "Estimating the Effects of Training Programs on Earnings," *Review of Economics and Statistics*, 60, 47–57.

Ashenfelter, O., and D. Card (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programmes," *Review of Economics and Statistics*, 66, 648–660.

Baily, M., and J. Tobin (1977): "Macroeconomic Effects of Selective Public Employment and Wage Subsidies," *Brooking Papers on Economic Activity*, 2.

Bellmann, L., and R. Jackman (1996a): "Aggregate Impact Analysis," in *International Handbook of Labour Market Policy and Evaluation*, ed. by G. Schmid, J. O'Reilly, and K. Schömann, pp. 143–162. Edward Elgar.

——— (1996b): "The Impact of Labour Market Policy on Wages, Employment and Labour Market Mismatch," in *International Handbook of Labour Market Policy and Evaluation*, ed. by G. Schmid, J. O'Reilly, and K. Schömann, pp. 725–746. Edward Elgar.

Bergemann, A., B. Fitzenberger, B. Schultz, and S. Speckesser (2000): "Multiple Active Labor Market Policy Participation in East Germany: An Assesment of Outcomes," Working Paper, Institute for Economic Research Halle, University of Mannheim.

Bijwaard, G., and G. Ridder (2000): "Correcting for Selective Compliance in a Re-employment Bonus Experiment," Working Paper, John Hopkins University, Baltimore.

Büttner, T., and H. Prey (1998): "Does Active Labour-Market Policy Affect Structural Unemployment? An Empirical Investigation for West German Regions, 1986-1993," *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 118, 389–413.

Bundesanstalt für Arbeit (1994): *Arbeitsmarkt 1993*. Nürnberg.

——— (1997): *Arbeitsmarkt 1996*. Nürnberg.

——— (2000): *Arbeitsmarkt 1999*. Nürnberg.

Burtless, G. (1995): "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, 9, 63–84.

Burtless, G., and L. Orr (1986): "Are Classical Experiments Needed for Manpower Policy?," *The Journal of Human Resources*, 21, 606–640.

Calmfors, L. (1994): "Active Labour Market Policy and Unemployment - A framework for the analysis of crucial design features," *OECD Economic Studies*, 22, 7–47.

Card, D. (2000): "Reforming the Financial Incentives of the Welfare System," Discussion Paper No.172, IZA.

Dehejia, R., and S. Wahba (1995): "Matching Approach for Estimating Causal Effects in Non-Experimental Studies," Working Paper, Department of Economics, Harvard University.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1045.

Eichler, M., and M. Lechner (1999): "An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt," Discussion Paper No.94, IZA.

Fay, R. (1996): "Enhancing the Effectiveness of Active Labor Market Policies: Evidence from Programme Evaluations in OECD Countries," Labour Market and Social Policy Occasional Papers, OECD.

Fertig, M., and C. Schmidt (2000): “Discretionary Measures of Active Labour Market Policy - The German Emplyoment Promotion Reform in Perspective,” Working Paper, Department of Economics (Econometrics), University of Heidelberg.

Fisher, R. (1935): *Design of Experiments.* Hafner, New York.

Fitzenberger, B., and H. Prey (1997): “Assessing the Impact of Training on Employment,” *ifo-Studien*, 43, 71–116.

——— (1998): “Beschäftigungs- und Verdienstwirkungen von Weiterbildungsmaßnahmen im ostdeutschen Transformationsprozeß: Eine Methodenkritik.,” in *Qualifikation, Weiterbildung und Arbeitsmarkterfolg. ZEW-Wirtschaftsanalysen Band 31*, ed. by F. Pfeiffer, and W. Pohlmeier. Nomos-Verlag.

——— (2000): “Evaluating public sector sponsored training in East Germany,” forthcoming, Oxford Economic Papers.

Fitzenberger, B., and S. Speckesser (2000): “Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland,” Discussion Paper No. 00-06, ZEW, Mannheim.

Franz, W. (1999): *Arbeitsmarktökonomik.* Springer-Verlag, Berlin, Heidelberg, New York.

Hagen, T., and V. Steiner (2000): *Von der Finanzierung der Arbeitslosigkeit zur Förderung von Arbeit - Analysen und Empfehlungen zur Arbeitsmarktpolitik in Deutschland.* Nomos Verlagsgesellschaft, Baden-Baden.

Heckman, J., and J. Hotz (1989): “Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training,” *Journal of the American Statistical Association*, 84, 862–880.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1996): “Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching As A Program Evaluation Method,” *Proceedings of the National Academy od Sciences*, 93, 13416–13420.

——— (1998): “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017–1098.

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

——— (1998): “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies*, 65, 261–294.

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.

Heckman, J., and J. Smith (1995): “Assessing the Case for Social Experiments,” *Journal of Economic Perspectives*, 9, 85–110.

Holland, P. (1986): “Statistics and Causal Inference,” *Journal of the American Statistical Association*, 81, 945–970.

Hübler, O. (1994): “Weiterbildung, Arbeitsplatzsuche und individueller Beschäftigungsumfang - Eine ökonometrische Untersuchung für Ostdeutschland,” *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 114, 419–447.

——— (1997): “Evaluation beschäftigungspolitischer Maßnahmen in Ostdeutschland,” *Jahrbücher für Nationalökonomie und Statistik*, 216, 21–44.

——— (1998): "Berufliche Weiterbildung und Umschulung in Ostdeutschland - Erfahrungen und Perspektiven," in *Qualifikation, Weiterbildung und Arbeitsmarkterfolg. ZEW-Wirtschaftsanalysen Band 31*, ed. by F. Pfeiffer, and W. Pohlmeier. Nomos-Verlag.

Hujer, R., K. Maurer, and M. Wellner (1998): "Kurz- und langfristige Effekte von Weiterbildungsmaßnahmen auf die Arbeitslosigkeitsdauer in Westdeutschland," in *Qualifikation, Weiterbildung und Arbeitsmarkterfolg. ZEW-Wirtschaftsanalysen Band 31*, ed. by F. Pfeiffer, and W. Pohlmeier, pp. 197–221. Nomos-Verlag.

——— (1999a): "Analyzing the Effects of On-the-Job vs. Off-the-Job Training on Unemployment Duration in West Germany," in *Panelanalysen zu Lohnstruktur, Qualifikation und Beschäftigungsdynamik*, ed. by L. Bellmann, and V. Steiner, pp. 203–237. IAB - Beiträge zur Arbeitsmarkt- und Berufsforschung 229.

——— (1999b): "The Effects of Public Sector Sponsored Training on Unemployment Duration in West Germany - A Discrete Hazard Rate Model Based on a Matched Sample," *ifo-Studien*, 3/1999, 45 Jhg., 371–410.

——— (1999c): "Estimating the Effect of Vocational Training on Unemployment Duration in West Germany - A Discrete Hazard Rate Model with Instrumental Variables," *Jahrbücher für Nationalökonomie und Statistik*, 218/5+6, 619–646.

Hujer, R., and M. Wellner (2000a): "Berufliche Weiterbildung und individuelle Arbeitslosigkeitsdauer in West- und Ostdeutschland: Eine mikroökonometrische Analyse," Discussion Paper.

——— (2000b): "The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany," Discussion Paper No.141, IZA.

Imbens, G. (1999): "The Role of the Propensity Score in Estimating Dose-Response Functions," Technical Working Paper, NBER.

Kluve, J., H. Lehmann, and C. Schmidt (1999): "Active Labour Market Policies in Poland: Human Capital Enhancement, Stigmatization, or Benefit Churning?," *Journal of Comparative Economics*, 27, 61–89.

Koenker, R., and Y. Bilias (2000): "Quantile Regression For Duration Data:A Reappraisal Of The Pennsylvania Reemployment Bonus Experiments," Working Paper, University of Illinois.

Kraus, F., P. Puhani, and V. Steiner (1999): "Employment Effects of Publicly Financed Training Programs - The East German Experience," *Jahrbücher für Nationalökonomie und Statistik*, 219, 216–248.

——— (2000): "Do Public Works Programs Work? Some Unpleasant Results from the East German Experience," in *Research in Labour Economics*, ed. by S. Polachek. JAI Press.

Kraus, F., and V. Steiner (1995): "Haben Teilnehmer an Arbeitsbeschaffungsmaßnahmen in Ostdeutschland bessere Wiederbeschäftigungschancen als Arbeitslose?," in *Mikroökonomik des Arbeitsmarktes*, ed. by L. Bellmann, and V. Steiner, pp. 387–423. IAB - Beiträge zur Arbeitsmarkt- und Berufsforschung 192.

LaLonde, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 76, 604–620.

——— (1995): "The Promise of Public-Sector Sponsored Training Programs," *Journal of Economic Perspectives*, 9, 149–168.

Layard, R., and S. Nickell (1986): "Unemployment in Britain," *Economica*, Suppl.53, 121–170.

Layard, R., S. Nickell, and R. Jackman (1991): *Unemployment - Macroeconomic Performance and the Labour Market.* Oxford University Press, New York.

Lechner, M. (1998): "Mikroökonometrische Evaluationsstuidien: Anmerkungen zu Theorie und Praxis," in *Qualifikation, Weiterbildung und Arbeitsmarkterfolg. ZEW-Wirtschaftsanalysen Band 31*, ed. by F. Pfeiffer, and W. Pohlmeier. Nomos-Verlag.

——— (1999a): "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany After Unification," *Journal of Business Economic Statistics*, 17, 74–90.

——— (1999b): "The Effects of Enterprise-related Continuous Vocational Training in East Germany on Individual Employment and Earnings," *Annals d'Economie et de Statistique*, 55-56, 97–128.

——— (1999c): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption," Discussion Paper No. 99/08, University of St. Gallen.

——— (1999d): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies," Discussion Paper No. 2000/01, University of St. Gallen.

——— (2000): "An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany," *Journal of Human Resources*, Spring, 347–375.

Martin, J. (1998): "What Works Among Active Labour Market Polices: Evidence From OECD Countries' Experience," Occasional Papers, OECD.

Neyman, J. (1935): "Statistical problems in agricultural experiments," *The Journal of the Royal Statistical Society*, 2, 107–180.

OECD (1991): *Employment Outlook.* Paris.

——— (1993): *Employment Outlook.* Paris.

——— (1999): *Employment Outlook.* Paris.

Pannenberg, M. (1995): *Weiterbildungsaktivitäten und Erwerbsbiographie - Eine empirische Analyse für Deutschland.* Campus-Verlag, Frankfurt, New York.

——— (1996): "Zur Evaluation staatlicher Qualifizierungsmaßnahmen in Ostdeutschland: Das Instrument Fortbildung und Umschulung (FuU)," Discussion Paper No.38, Institut für Wirtschaftsforschung Halle.

Pannenberg, M., and J. Schwarze (1998): "Labor market slack and the wage curve," *economics letters*, 58, 351–354.

Prey, H. (1997): "Wirkungen von Maßnahmen staatlicher Arbeitsmarkt- und Beschäftigungspolitik," Cile Discussion Paper No. 45, University of Konstanz.

——— (1999): *Wirkungen staatlicher Qualifizierungsmaßnahmen. Eine empirische Untersuchung für die Bundesrepublik Deutschland.* Paul Haupt-Verlag, Bern, Stuttgart, Wien.

Puhani, P. (1999): *Evaluating Active Labour market Policies - Empirical Evidence for Poland During Transition.ZEW Economic Studies, Vol.5.* Physica-Verlag, Heidelberg, New York.

Quandt, R. (1972): "Methods for estimating switching regressions," *Journal of the American Statistical Association*, 67 (338), 306–310.

——— (1988): *The economics of disequilibrium.* Basil Blackwell, Oxford.

RIDDELL, C. (1991): “Evaluation of Manpower and Training Programmes: The North American Experience,” in *Evaluating Labour Market and Social Programmes, The State of a Complex Art*, ed. by OECD, pp. 43–72. OECD.

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

——— (1985a): “The Bias due to incomplete Matching,” *Bioemetrics*, 41, 103–116.

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

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

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

——— (1977): “Assignment to Treatment Group on the Basis of a Covariate,” *Journal of Educational Studies*, 2, 1–26.

——— (1979): “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies,” *Journal of the American Statistical Association*, 74, 318–328.

——— (1980): “Comment on Basu, D. - Randomization Analysis of Experimental Data: The Fisher Randomization Test,” *Journal of the American Statistical Association*, 75, 591–593.

SCHMID, G., S. SPECKESSER, AND C. HILBERT (2000): “Does Active Labour Market Policy Matter? An Aggregate Impact Analysis for Germany,” in *Labour Market Policy and Unemployment. Evaluation of Active Measures in France, Germany, The Netherlands, Spain and Sweden (im Erscheinen)*. Edward Elgar.

SCHMIDT, C. (1999): “Knowing What Works - The Case for Rigorous Program Evaluation,” Discussion Paper No.77, IZA.

SELL, S. (1998): “Entwicklung und Reform des Arbeitsförderungsgesetzes als Anpassung des Sozialrechts an felxible Erwerbsformen,” *Mitteilungen aus der Arbeitsmarkt und Berufsforschung*, 3, 532–549.

STAAT, M. (1997): *Empirische Evaluation von Fortbildung und Umschulung - Schriftenreihe des ZEW 21*. Nomos Verlagsgesellschaft, Baden-Baden.

STEINER, V., E. WOLF, J. EGELN, M. ALMUS, H. SCHRUMPF, AND P. FELDOTTO (1998): *Strukturanalyse der Arbeitsmarktentwicklung in den neuen Bundesländern*. Nomos Verlagsgesellschaft, Baden-Baden.

## A Tables

Table 2: **Microeconometric Evaluations of Job Creation Schemes for East Germany**

| **Authors** | **Data Period** | **No.of obs.** | **Programme** | **Evaluation Methodology** | **Outcome Variable** | **Results** |
|---|---|---|---|---|---|---|
| Bergemann et al. (2000) | LMM Sachsen-Anhalt 1990-1998 | T:615 C:3137 | Job creation schemes | Matching, 'conditional' DiD-estimator, regarding multiple participation | Employment | Long-run effects of job creation schemes are at best insignificant |
| Hübler (1997) | EGLMM 1990-1994 | T:varies C:2886 | Public works, training programmes on-TJT and off-TJT, short-time work | Linear control functions; IV-estimator, random-effects-estimator, matching approach | Employment (opportunities) | No clearly effects can be detected for training (negative in the short-run, positive in the long-run), public work and short time-work have no positive effects |
| Eichler and Lechner (1999) | LMM Sachsen-Anhalt 1992-1997 | T:1299 C:16334 | Public employment programmes | Nonparametric DiD methods, matching | Unemployment probabilities | Unemployment risk for participants is reduced substantially |
| Kraus and Steiner (1995) | EGLMM 1990-1992 | T:582 C:2179 | Job creation schemes | Discrete hazard rate model (multinomial logit), different sub-groups | Transition rates to regular employment | Negative effects for women, positive effects for men after 12 months, only slight effects before |
| Kraus, Puhani, and Steiner (2000) | EGLMM 1990-1992 1992-1994 | T:718 C:3503 | Public work programmes | Discrete hazard rate model with unobserved heterogeneity | Re-employment probabilities, (un)stable employment | Negative effects on re-employment probabilities |

LMM: Labour Market Monitor, EGLMM: East German Labour Market Monitor
On-TJT: On-the-job training, Off-TJT: Off-the-job training, DiD: Difference-in-differences

Table 3: **Microeconometric Evaluations of Vocational Training for West Germany**

| **Authors** | **Data Period** | **No.of obs.** | **Programme** | **Evaluation Methodology** | **Outcome Variable** | **Results** |
|---|---|---|---|---|---|---|
| Hujer, Maurer, and Wellner (1998) | GSOEP-West 1986-1993 | T:219 C:715 | Vocational Training | Matching (Discrete hazard rate model, gamma distributed unobserved heterogeneity) | Unemployment duration | Significant reducing effect in the short run, insignificant in the long-run |
| Hujer, Maurer, and Wellner (1999a) | GSOEP-West 1986-1993 | T:219 C:715 | Vocational training (on- and off-TJT) | Matching | Unemployment duration | On-TJT: No significant effect. Off-TJT: Significant negative impact on unemployment duration in the short run, insignificant in the long-run |
| Hujer, Maurer, and Wellner (1999b) | GSOEP-West 1985-1993 | T:162 C:1851 | Public-sector sponsored vocational training | Matching (Panel-Probit) | Unemployment duration | Significant positive effects for short courses (<6months), longer courses show no positive effects |
| Hujer, Maurer, and Wellner (1999c) | GSOEP-West 1986-1993 | T:100 C:1996 | Vocational Training | IV-estimator (Discrete hazard rate model, gamma distributed unobserved heterogeneity) | Unemployment duration | Participation has significant negative effect on unemployment duration |
| Hujer and Wellner (2000a) | GSOEP-West 1984-1994 | T:162 C:1851 | Vocational training | Matching, Panel-Probit | (Un)employment duration | Short courses (< 6 months) tend to be positive, but the effects diminish, no significant effects for long courses |
| Pannenberg (1995) | GSOEP-West 1984-1991 | T:26-308 C:689-1657 | Vocational Training (on- and off-TJT) | Discrete hazard rate, logit and probit estimates | Re-employment probability | Off-TJT increases the re-employment probability in the short-run |
| Prey (1997) | GSOEP-West 1984-1993 | | Vocational training with income maintenance | Simultaneous dynamic random-effects probit model, pre-programme test | Employment Probability | No effect for women, negative effect for men |
| Prey (1999) | GSOEP-West 1985-1994 | T:42-159 C:469-614 | Qualifying measures with and without income maintenance | Simultaneous dynamic probit (participation, employment) and tobit (wages) model, pre-programme test | Employment probability, hourly wages | Negative (no) effects on the employment probability for measures with (without) income maintenance, no effects on the wages |
| Staat (1997) | GSOEP-West 1984-1994 | T:312 C:1548 | Public sector training with income maintenance | Ordered-probit model, IV estimation | Job search duration; employment stability; wages | Positive effects on search duration only for sub-groups, no significant effects on the employment stability, search duration for women is reduced (positive effects on the wages) |

GSOEP: German Socio-Economic Panel, IV: Instrument-variable
On-TJT: On-the-job training, Off-TJT: Off-the-job training

Table 4: **Microeconometric Evaluations of Vocational Training for East Germany**

| **Authors** | **Data Period** | **No.of obs.** | **Programme** | **Evaluation Methodology** | **Outcome Variable** | **Results** |
|---|---|---|---|---|---|---|
| Bergemann et al. (2000) | LMM Sachsen-Anhalt 1990-1998 | T:931 C:3137 | Further training | Matching, 'conditional' DiD-estimator, regarding multiple participation | Employment | No positive effects for first training programmes, additional effects of a second participation on average not different from zero |
| Fitzenberger and Prey (1997) | EGLMM 1990-1992 | T:270-307 C:3775 | Training within and outside of the firm | Simultaneous probit equations model of sample attrition, training participation and labour force status | Employment probabilities | Training outside (inside) the firm shows considerable positive (negative) effects |
| Fitzenberger and Prey (1998) | EGLMM 1990-1994 1993-1994 | Women: T:67-79 C:3415 Men: T:27-88 C:3680 | Training within and outside of the firm, with and without public income maintenance | Random-effects estimation, pre-programme test, difference-in-differences, Matching | Employment, wages | Results differing with respect to the specification, most often negative or no effects |
| Fitzenberger and Prey (2000) | EGLMM 1990-1994 | Women T:325 C:2048 Men T:146 C:2268 | Training supported by public income maintenance outside a firm | Simultaneous random-effects probit and tobit model for treatment, employment and wages, difference-in-differences | Employment, wages | Positive effects for employment and earnings (but only a few significant in the long-run) |
| Hübler (1994) | EGLMM 1990 | T:1276 C:4679 | Qualifying measures (only on-TJT) | Simultaneous probit model for participation and search behaviour | Job search, working time | Training activities induce search activities and reduce the effective hours of work |
| Hübler (1998) | EGLMM 1990-1994 | T:206-934 C:922 | Vocational training and retraining, publicly and privately financed, on-TJT and off-TJT | Different estimation methods are compared (matched samples, OLS, Logit-ML estimates) | Earnings, working time, employment probability | Priv.financed: Pos. effects on the job security for on-TJT, higher earnings for off-TJT. Publ.financed: Only in the short-run pos. effects, than reduced earnings, shorter working time |
| Hujer and Wellner (2000b) | GSOEP-East 1990-1993 | T:231 C:1401 | Public-sector sponsored vocational training | Matching (discrete time hazard rate), panel-probit | (Un)employment duration | No significant effects, very weak hint that short courses seem to be more effective reducing unemployment duration |

Table 4 (contd.)

| **Authors** | **Data Period** | **No.of obs.** | **Programme** | **Evaluation Methodology** | **Outcome Variable** | **Results** |
|---|---|---|---|---|---|---|
| Kraus, Puhani, and Steiner (1999) | EGLMM 1989-1992 1992-1994 | T:1744 C:3503 | Publicly-financed training and retraining, on-TJT, off-TJT | Linear control function estimator, multinomial logit duration models with unobserved individual heterogeneity | (Un)Stable employment | No positive effects in the first period, positive effects of on-TJT and off-TJT in the second period |
| Lechner (1999a) | GSOEP-East 1990-1994 | T:131 C:1105 | Off-TJT, publicly financed and enterprise-related | Matching | Employment probabilities, earnings | No robust positive effects of off-TJT were found |
| Lechner (1999b) | GSOEP-East 1990-1994 | T:222 C:980 | Enterprise-related continuous vocational training | Matching | Employment probabilities, earnings | Positive earnings effects, no effects on unemployment probability |
| Lechner (2000) | GSOEP-East 1990-1996 | T:131 C:1280 | Public-sector-sponsored vocational training and retraining | Matching | Employment probabilities, earnings | No positive long term effects on the employment probabilities or the earnings. Negative effects on risk of unemployment in the short-run |
| Pannenberg (1995) | GSOEP-East 1990-1992 | T:269 (On) T:112 (Off) C:2037 | Vocational Training (on- and off-TJT) | Discrete hazard rate, logit and probit estimates | Re-employment probability, wages | Positive effects on the re-employment probability and the income |
| Pannenberg (1996) | GSOEP-East 1990-1994 | T:90 (Empl.) C:985 T:55 (Wages) C:606 | Off-TJT, publicly financed | Discrete hazard rate model, linear control function estimator, pre-programme test | Re-employment probability, earnings | Positive effects on re-employment probability and on earnings |
| Staat (1997) | GSOEP-East 1992-1994 | (Empl.) T:316 C:804 | Public sector training with income maintenance | Ordered-probit model, IV estimation | Job search duration, employment stability, wages | No effects on the search duration, the employment stability or the level of hourly wages |

EGLMM: East German Labour Market Monitor, GSOEP: German Socio-Economic Panel
DiD: Difference-in-differences, On-TJT: On-the-job training, Off-TJT: Off-the-job training
IV: Instrument-variable, OLS: Ordinary Least Square, ML: Maximum-Likelihood

Table 5: **Macroeconometric Evaluations**

| **Authors** | **Data Period** | **No. of obs.** | **Programme** | **Evaluation Methodology** | **Outcome Variable** | **Results** |
|---|---|---|---|---|---|---|
| Büttner and Prey (1998) | Planning regions of West Germany 1986-1993 | N:74 | Training programmes, public job creation | Disequilibrium approach, estimation of the matching efficiency (static and dynamic) with OLS, LSDV and 2SLS, controlling for endogeneity | Labour market mismatch | No significant effects of training programmes, job creation schemes reduce mismatch |
| Hagen and Steiner (2000) | Local labour office districts 1990-1999 (and sub-periods) | N:142 (West) N:35 (East) | Vocational Training; job-creation schemes, structural adjustment schemes | Augmented matching function, fixed-effects model | Inflow in / outflow of unemployment | Net-effects East: JCS and VT increase the unemployment rate, SAM reduces it slightly. West: All measures increase unemployment |
| Pannenberg and Schwarze (1998) | Local labour office districts 1992-1994 | N:35 | Training programmes | OLS with regional fixed effects; FGLS with random indiv. And fixed regional effects; 2SLS estimator with fixed regional effects | Hourly (aggregate) wages | ALMPs have negative effects on regional wages |
| Prey (1999) [extension of Büttner/Prey (1998)] | Planning regions of West Germany 1986-1993 | N: 74 | Training programmes, public job creation | See Büttner and Prey (1998). Additional: controlling for regional age structure and recipients of social assistance, separately for men and women | Labour market mismatch | VT increases (decreases) the mismatch for women (men), JCS decreases the mismatch for men |
| Schmid, Speckesser, and Hilbert (2000) | Local labour office districts 1994-1997 | N:142 | Further training, retraining, public job creation and wage subsidies | Pooled cross-section analysis, fixed-effects model; augmented matching model | Long-term unemployment (LTU); unemployment outflows | JCS reduces only short LTU (<6 months), VT reduce LTU (not all significant), wage subsidies help only the very long-term unemployed |
| Steiner et al. (1998) | Local labour office districts 1992-1997 (and sub-periods) | N:35 | Vocational Training | Pooled cross-section analysis, augmented matching function | Outflow of unemployment | No (very small) effects on the matching efficiency (in the short run) |

OLS: Ordinary Least Squares, LSDV: Least Squares Dummy Variables
2SLS: Two-Stage Least Squares, FGLS: Feasible Generalized Least Squares