

Discretionary Measures of Active Labor Market Policy

The German Employment Promotion Reform in Perspective*

By Michael Fertig and Christoph M. Schmidt

Abstract

This paper provides a preliminary assessment of recent reforms of German employment promotion policy. While several recent studies analyze the impact of measures of employment promotion for the case of Germany, no comparable study exists on the aggregate level, thus precluding any assessment of the overall impact of such measures. The reform of 1998 recognized for the first time in the history of German employment promotion that evaluation should be an integral part of the delivery of policy interventions. Yet, while this has led to an important first step – the systematic collection of data – evaluation itself has mistakenly been understood by the legislator as a pure accounting exercise. We discuss this problem from the perspective of the recent evaluation literature in economics and statistics, thus providing a frame of reference for performing credible scientific evaluation of employment promotion policy.

Zusammenfassung

Dieses Papier bietet eine erste Bestandsaufnahme der jüngsten Reform im Arbeitsförderungsrecht der Bundesrepublik Deutschland. Das Arbeitsförderungsreformgesetz von 1998 erkennt erstmals in der Geschichte des deutschen Arbeitsförderungsrechts die Notwendigkeit der Evaluierung solcher Maßnahmen explizit an. Während dies zu einem wichtigen ersten Schritt hin zu einer wissenschaftlichen Evaluierung geführt hat, nämlich der systematischen Erhebung von Datenmaterial, wurde die Notwendigkeit zur Evaluierung selbst als ein rein buchhalterisches Unterfangen der Arbeitsämter missverstanden. Darüber hinaus existieren in der wissenschaftlichen Literatur einige Studien, die die Effektivität von Maßnahmen aktiver Arbeitsförderung in der Bundesrepublik Deutschland auf die individuelle Arbeitsmarktsituation der Maßnahmeteilnehmer analysieren, es existiert jedoch keine vergleichbare Studie auf aggregierter Ebene. In diesem Papier wird dieses Problem aus der Perspektive der modernen Evaluierungsliteratur mit dem Ziel diskutiert, einen einheitlichen Rahmen

* The authors are grateful to Boris Augurzky, Klaus F. Zimmermann and three anonymous referees for helpful comments and Maria Ntokoutsi for research assistance.

für die Evaluierung von Maßnahmen der Arbeitsförderung in der Bundesrepublik Deutschland auf aggregierter Ebene zu entwickeln.

JEL-Classification: J68, H43, R23.

1. Introduction

This paper provides a preliminary assessment of German employment promotion policy and its reform, most notably of the shift of decision authority towards the local employment offices, from the perspective of contemporaneous economic research. Every year Germany spends several billions of Euros on active measures of employment promotion with the explicit intention to contribute to the reduction of unemployment. Yet, unemployment has been a persistent problem throughout the last two decades, raising the question as to the actual effects of employment promotion. In line with a general tendency to redefine and modernize all kinds of different administrative organizations, a major reform of the policy of employment promotion was launched in 1998, combined with the requirement that the effects of any implemented measure be evaluated systematically.

On an individual level, inspired by a growing body of international evaluation literature (cf. the pioneering work of Rubin (1974) and (1986)), several recent studies evaluate the labor market impact of some measures of active labor market policy (mainly training measures) for Germany¹. The evidence of these studies concerning the efficacy of active labor market policy (ALMP) interventions is mixed. Most of them, as well as the majority of international studies, show a rather small, if any, effect of employment promotion measures on the individual level. However, individual level studies can at best be indicative of the overall program impact, since indirect effects are not addressed at all.

No comparable study exists on the aggregate level, although – as we will argue below – potentially informative data material has been collected as a consequence of reform. Most importantly, while the legislator raises the requirement for evaluation, no clear guidance is given on how any scientific evaluation should proceed. Rather, what is required of the local employment offices is a pure accounting exercise, leaving open the question if favorable outcomes have to be attributed to the specific design of the programs, to the particular selection of target individuals within a region, or even to the state of the regional labor market. What is needed, therefore, is methodological

¹ See e.g. Fitzenberger and Prey (2000), Hübler (1997), Hujer et. al. (1997), Lechner (1998) and (1999).

guidance as to how a credible evaluation study trying to assess the German policy of employment promotion has to proceed.

However, for this endeavor it is imperative to have a clear understanding of the *institutional* framework of German employment promotion policy because this allows one to lift at least some of the fog around the decision processes of the labor offices. This institutional aspect is usually neglected in the received literature on the evaluation of labor market programs. This paper, therefore, contributes to the literature in two important directions. Firstly, it describes and organizes the institutional and legal framework of German employment promotion policy along the lines of economically interesting and meaningful categories. Secondly, it provides an overview on the principles one has to adhere to if one wishes to seriously evaluate these measures on a semi-aggregate level. These two issues are intimately connected and can, therefore, hardly be discussed separately.

The setup of the paper is as follows. The second section introduces the main characteristics of the regime change in 1998, especially the shift of budget authority towards the local employment offices. To conceptualize the ample variety of labor market measures currently available, section 3 explains and classifies the different measures of employment promotion. We explicitly distinguish between *non-discretionary* and *discretionary measures* and summarize the latter into four broad, economically interesting categories in deviation from the wording of the law. In section 4 the problem of evaluating the impact of this reform is explained. We will demonstrate that the *semi-aggregate* nature of the regional data, in contrast to the usual individual-level data, needs to be addressed carefully in any evaluation attempt. Moreover, it will become clear that the problem of evaluating the implementation of this reform cannot be solved by the input accounting approach demanded by the new law. Finally, section 5 offers some conclusions and provides an agenda for future research.

2. Decentralization of Employment Promotion

One of the declared objectives of the German *Federal Ministry of Labor and Social Affairs* (*Bundesministerium für Arbeit und Sozialordnung*) has always been the promotion of employment. While a specific employment promotion law provided the legal basis for employment promotion policies throughout much of (West) German post-war history, this legal framework has been integrated into the *Social Code* in January 1998 (*Sozialgesetzbuch* (SGB), *Drittes Buch* (III) *Arbeitsförderung*). Actual employment promotion policies themselves are implemented by the *Federal Employment Services*

Agency (*Bundesanstalt für Arbeit*) in Nürnberg and its regional (*Landesarbeitsämter*) and local employment offices (*Arbeitsämter*).

Approximately three quarters of the annual budget disbursed for employment promotion policies are unambiguously determined by eligibility criteria alone, while the remaining quarter, in 1998 almost 25 billion DM (on average, 4.28 million workers were unemployed during 1998), are earmarked for discretionary measures of employment promotion (*Ermessensentscheidungen*). Funds out of this part of the overall budget are granted at the individual level on a case-by-case basis by administrators at the local employment offices.

Until the end of 1997, a central advisory board (*Verwaltungsrat*) at the Federal Employment Services Agency determined how the overall budget being available for discretionary measures was allocated to local employment offices. Moreover the same board determined the budget shares to be received by individual measures of employment promotion. It was the major task of the staff in the local employment offices to disburse the funds according to their designated purpose, irrespective of the actual severity or composition of unemployment in the local labor market compared to other local labor markets. In particular, once budget shares were determined there was no possibility to cross-subsidize policy measures.

The year 1998 introduced a major regime switch. The reform of the employment promotion law (*Arbeitsförderungs-Reformgesetz* (AFRG)) on March 24, 1997 had several important implications for employment promotion in Germany. One very important change was that the largest part of the funds earmarked for discretionary measures were pooled (*Eingliederungstitel*)². Moreover, the administrative boards of the *local* employment offices were awarded discretion on how to allocate these funds to the individual policy measures.

The AFRG which came into force on *January 01, 1998* replaced the *Arbeitsförderungsgesetz* (AFG) from 1969. With this reform the employment promotion law is now the third part of the *Social Code* (Sozialgesetzbuch III – *Arbeitsförderung*; short: SGB III). Its main task is the regulation of active employment promotion; its structure is similar to that of its historical predecessor. The AFG from 1969 was the first German law to put the main emphasis on active labor market policies, especially on the promotion of qualification and on job creation schemes (*Arbeitsbeschaffungsmassnahmen*).

Moreover, it introduced several forms of benefit schemes, e.g. the insolvency allowance and a longer duration of unemployment benefits. The aims

² Some funds for discretionary measures are separated from the *Eingliederungstitel*.

of this law were very ambitious. § 1 AFG stated that the main purposes of the law were (i) to achieve a high level of employment, (ii) to enhance the employment structure, and (iii) to promote economic growth (Gagel (1999)). Yet, there was no systematic attempt whatsoever at gauging the success of this law. As unemployment rose during the years it became transparent that the simultaneous achievement of all these ambitious goals was quite unrealistic. Over time approximately 100 amendment laws were passed, still without any plans for impact evaluation. At the end of the last decade, the legislator considered the AFG in need of reform. The declared objective of this reform (Deutscher Bundestag, Drucksache 13/4941 vom 18. 06. 1996; short: BT-Drucksache 13/4941) was the provision of a unified framework for employment promotion measures to make them more transparent and more comprehensible.

The SGB III, established in a situation of high public budget deficits, explicitly chooses a starting point different to the AFG. The objectives of this reform as declared by the legislator (Clever (1998) and BT-Drucksache 13/4941) was to avoid that a detailed catalog of aims would lead to increasing claims for further benefit payments. Therefore, the main goals of the reform were formulated much more generally as (i) to increase the chances of the unemployed to find a new job and to reduce the probability of the employed to lose their jobs, (ii) to improve the traditional labor law and the applicability of its regulations, (iii) to increase the effectiveness and efficacy of the Federal Labor Office (*Bundesanstalt für Arbeit*), (iv) to prevent illegal work and the illegal claim of benefit payments, and (v) to reduce the overall cost of labor market policy.

Furthermore, in the discussion accompanying the draft of the new law the legislator expressed his concern that the widespread belief in active labor market measures as a way to create many new jobs may be quite unrealistic (BT-Drucksache 13/4941). Furthermore, it was recognized that there may also be the possibility of endangering existing jobs by employment promotion measures. Therefore, the SGB III emphasizes the special responsibility of employers and employees in creating new jobs as well as in maintaining existing ones (§ 2 SGB III) and puts the main emphasis of the law on assisting the unemployed to find a job as soon as possible by moderating between employers which offer jobs and the unemployed (§ 1,(1) SGB III).

At the operational level two major changes were introduced by the new law. First, the responsibility for the implementation of the measures of active labor market policy were delegated to the local employment agencies (§ 9 SGB III, § 71b and § 71c SGB IV). These agencies cover regions varying in size and economic structure, in particular regarding the labor market. As a consequence of this reform, the local agencies now directly determine the

amount of money spent for the discretionary measures of active labor market policy. There are some general principles which have to be adhered to (§ 7 SGB III), but within this framework the local agencies are free to decide.

In this process the new law itself offers some guidance to local decision makers. The first section of the SGB III contains these fundamental principles as well as the overall goals of the new law. § 1,(2) states that all funds have to be disbursed according to the general labor market policy aims of the federal government and that competitive jobs should not be endangered by these measures. However, there is no more detailed regulation on how this policy should be implemented. Instead, there is a preference ordering for the various measures of active labor market policy in § 4 and § 5 SGB III. *Guidance and placement assistance* (for more details on this and the other measures see below) should have priority over all other measures of active labor market policy which in turn should be preferred to the payment of benefits replacing work income. There are also several target groups mentioned in § 7 and § 8 SGB III which should be of prime interest in the allocation of active measures. These groups are: long-term unemployed, disabled persons, older workers, and individuals returning to the labor market, especially women.

Second, the catalog of instruments for employment promotion was enhanced by several new measures. Among these new measures introduced by the reform are the *preparatory training measures*, the *recruitment subsidies for newly self-employed*, the so-called *integration contracts* and the possibility to spend up to 10% of overall resources devoted to active labor market policy freely ("*Freie Förderung über Innovationstopf*", § 10 SGB III). The latter was designed in order to provide the local employment offices with the possibility to create measures of employment promotion for their specific needs. This could be measures not explicitly mentioned in the law and the funds could also be used, contrary to those allocated to other measures, for the support of *projects* as a whole. As before, the actual disbursement of the funds usually has to be decided upon on an *individual* case-by-case basis by the local staff.

Finally, local employment offices are required to implement an accounting system capturing the allocation of their funds to the various employment promotion measures at their discretion. The principal tools employed for this purpose are balance sheets documenting selected characteristics of recipients as well as the amounts granted to and the measures received by different recipient groups. Moreover, without further discussion of the fundamental methodological problems arising with such a request, the law demands that local labor offices also report the effectiveness of their policy

(§ 1 SGB III). Without further guidance this prescription must lead to the wrong answer, namely a pure accounting approach. Therefore, after characterizing individual measures in more detail in the next section, this paper will clarify in section 4 what the appropriate questions to ask would be, what alternative, conceptually misleading questions will be answered by any labor office's balance sheet, and how this accounting data could nevertheless be used to assess the economic effects of various measures of active labor market policy.

3. Measures of Employment Promotion

The SGB III³ is organised in 13 chapters (*Kapitel*) which in turn contain several sections (*Abschnitte*) and subsections (*Unterabschnitte*). Some of these chapters regulate special aspects of labor market policy, like e.g. work permissions for foreigners, which are not subject of this paper. Apart from chapter 1 which contains the general rules of the SGB III, the most interesting parts of this law are the chapters 4, 5, and 6. These chapters contain the regulations concerning payments to employees/unemployed (ch. 4), payments to employers (ch. 5), and the institutions carrying out some of the employment promotion measures (*Träger der Maßnahmen*; ch. 6).

All these measures will be described in more detail below. For this purpose and with regard to the further analysis we divide them into two categories: *non-discretionary* and *discretionary* measures. Non-discretionary measures are defined by the existence of a legal claim for workers who fulfill certain eligibility requirements. Their claim cannot be rejected by the employment agencies. By contrast, it is the discretionary measures which might be deliberately used as instruments aiming at particular target groups or labor market problems that employment offices would like to address. Most of the regulations in the SGB III could be amended by statutory orders of the *Federal Ministry of Labor and Social Affairs* or by decree of the *Federal Employment Services Agency*.

3.1 Non-Discretionary Measures

The most important non-discretionary measures are the several forms of benefit payments replacing work income (*Entgeltersatzleistungen*). § 116 SGB III gives an overview over these payments. There are six different entries:

³ All paragraphs are quoted with reference to Sozialgesetzbuch III – Arbeitsförderung, Beck-Texte, 4th edition.

1. Unemployment benefit for unemployed and partially unemployed (§§ 117–152);
2. Allowances for employees who attend training measures (§§ 153–159);
3. Temporary benefits for disabled persons who attend integration schemes (§§ 160–168);
4. Short-time allowance (§§ 169–182);
5. Insolvency allowance (§§ 183–189);
6. Unemployment assistance (§§ 190–206).

Additionally, there is a separate section regulating two special forms of benefit payments: the winter bad weather benefit and the winter allowance (§§ 209–216). For all of these payments a list of different eligibility prerequisites must be met and all of them vary in the duration and amount of entitlement (§§ 117 ff). *Table 1* documents the total amount of money spent by the employment offices and the amount for the various non-discretionary measures in 1998. By far the largest share of all non-discretionary expenditures was spent for unemployment benefit and assistance payments, i.e. approximately 96.3%. Compared to these entries the other benefits payments are quantitatively of minor importance.

Table 1
Selected Expenditures for Non-Discretionary Measures in 1998

Item	Amount in 1,000 DM	% of Total Expenditures
Total Expenditures	134,932,049	
of which: Total Non-Discret.	86,430,097	64.05
of which:		
Unemployment Benefits	52,826,984	39.15
Unemployment Assistance	30,437,600	22.56
Insolvency Allowance	2,054,663	1.52
Short Time Allowance	656,335	0.49
Winter Allowance	376,617	0.28
Winter Bad Weather Benefit	77,898	0.06

Source: Bundesanstalt für Arbeit (1999b); own calculations.

Unemployment benefit

There are several prerequisites which must be met for entitlement: (i) unemployment, (ii) a minimum period of work before unemployment, and (iii) registration as unemployed at local labor office. *Unemployment* means that an individual does not work more than 15 hours a month and is actively

looking for a new job. The duration of unemployment benefits depends on the number of years worked before unemployment and the age of the individual (§ 127 SGB III) and varies between 6 and 32 months. The amount of benefit paid depends on the level of earnings in the last job (§§ 129 ff SGB III) and varies between 60% and 67% of the last net earnings. The AFRG also introduces the possibility of a partial unemployment benefit for people who lose one out of several jobs and are therefore called partially unemployed.

Unemployment assistance

There are several forms of unemployment assistance. The most important one is the assistance paid after the claim to unemployment benefit is exhausted. (§ 191,(1), No. 1 SGB III). The duration of entitlement is 12 months and the amount paid varies between 53% and 57% of the last net earnings. The measure is means-tested since assets of the individual or the individual's spouse will be taken into account. Several details concerning unemployment assistance are regulated in an additional law called *Arbeitslosen-hilfereformgesetz* from June 1996.

Another benefit payment is the *short-time allowance*. Short-time work means that an individual is out of work for a temporary period, either due to a deterioration in the demand of her company's products or to other disturbances in the production process, but without being unemployed. A related benefit is the so called *structural short-time allowance* which could be granted in particular cases where the cancellation problem persists for a protracted period. This may be the case in serious deteriorations of the economic situation of a complete economic sector which is accompanied by the threat of mass dismissals of workers. The *insolvency allowance* is granted to employees of companies for which the bankruptcy proceedings are initiated or which have to be shut down permanently. The *winter bad weather benefit*, and the *winter allowance* are available to employees in the construction industry only. They aim at compensating workers for hours lost due to bad weather in the winter season (November 1 to March 31).

The next section describes in more detail the various measures of active labor market policy for which the local employment offices now have discretion in disbursement.

3.2 Discretionary Measures

The SGB III contains several measures of active labor market policy for which the Federal and State Employment Agencies and the local employment offices have a broad scope of action. For all these measures the local employment agencies decide on the amount of money to be spent. To con-

ceptualize affairs, we divide these measures into four broad categories: (i) monetary and non-monetary assistance for finding jobs, (ii) measures promoting the qualification of the unemployed (“human capital formation”), (iii) the provision of financial incentives, and (iv) the creation of jobs. Every category contains one or more instruments of active labor market policy, all of them with detailed regulations regarding entitlement, duration of measures and the amount of money granted. In general, all regulations of the SGB III concerning these measures could be amended by decree of the *Federal Employment Services Agency*.

(i) *Monetary and non-monetary assistance for finding jobs*

It is explicitly stated in § 4 SGB III that the placement of unemployed workers into employment has to be the main objective of the employment agencies. It enjoys priority over all other measures of active labor market policy. Therefore, the employment offices are obliged to give advice about job opportunities to people actively looking for jobs and to help in matching supply and demand on the labor market. Moreover, the local employment offices are prepared to reimburse some of the expenses resulting from the process of job application. This *guidance and placement assistance* (§§ 29–44 SGB III) covers cost for application material, traveling to job interviews and, if necessary, accommodation. In addition, unemployed individuals taking up a job offer can claim *mobility benefits* (§§ 53–56 SGB III) covering, among others, the cost for work clothes and equipment, daily cost of travel between home and work, and part of the cost for running a second household, if necessary.

(ii) *Measures promoting the qualification of the unemployed*

Three measures are available to improve the qualification of unemployed individuals with the objective of increasing their prospects in finding a job. The same measures are available to employed workers being at risk of dismissal. The principal motivation behind such measures is the insight that the probabilities of job loss are lower and those of hiring are typically higher, respectively, for skilled individuals (for empirical evidence see e.g. Schmidt (1999)). The first item are *training measures* (§§ 48–52 SGB III). These are measures aiming at (i) examining the ability of an unemployed individual for performing a specific job or another measure of active labor market policy (duration: four weeks), (ii) improving the ability to apply for jobs (e.g. application courses; duration: two weeks), and (iii) improving specific skills necessary for finding work (e.g. computer courses; duration: eight weeks). The employment offices cover cost like the course fees, as well as travel and child care expenditures. The payment of unemployment benefits or assistance continues for the duration of these courses.

The second measure is the *promotion of vocational training* (§§ 59–76 SGB III). The employment offices can grant an allowance for vocational training or courses trying to prepare young unemployed without any or with low qualifications for work. These allowances (partly) cover cost-of-living expenses, course fees, working clothes, travel expenditures, and child care, for workers unable to cover the cost themselves. This measure aims at facilitating the taking up of (preparatory) vocational training to improve the skills of young workers.

The last measure in this context is the *promotion of further training* (§§ 77–96 SGB III). Basically, the allowances granted for further training measures cover the same expenditures as those of the vocational training grant. Further training measures can be granted to individuals who are unemployed and have no formal job qualification or to employed individuals who can claim convincingly that their lacking job qualification puts them at a considerable risk of dismissal. In addition, it is usually a further prerequisite for entitlement that these persons must have had a job for at least 12 months during the last three years before taking up the measure. In special cases it is possible that this prerequisite is not insisted upon.

(iii) *Financial incentive Schemes*

The SGB III contains several measures directly aiming at increasing the willingness of employers to hire unemployed individuals. This is usually done by some form of wage subsidy and mainly differs in the amount and the duration of the grant. The detailed regulations are spread throughout several chapters of the SGB III and can be structured as follows.

a) Integration subsidies (§§ 217–224 SGB III)

Integration subsidies are wage subsidies which can be granted to employers hiring unemployed workers in order to compensate these employers for the presumably lower productivity of the unemployed. The most important features of the available subsidies are as follows:

- Subsidies for workers who *must be trained*. If the employer can claim convincingly that unemployed workers must be trained to be able to meet the requirements of their new jobs, a subsidy can be granted for six months which covers 30% of the subsidizable wage.
- Subsidies for *difficult-to-place workers*. Unemployed workers with characteristics that are unfavorable for their hiring prospects, e.g. long-term unemployed or persons with disabilities, are classified as difficult-to-place workers. For them a subsidy can be granted for 12 months which covers 50% of the subsidizable wage.

- Subsidies for *old workers*. For unemployed workers aged 55 and above a subsidy can be granted for 24 months, also covering 50% of the subsidizable wage.

The duration and/or the amount of the integration subsidies can be increased in cases where the productivity of a person is extremely low (§§ 221 and 222 SGB III). However, the law itself does not regulate how this should be measured or on which criteria the decision should be based.

b) Integration contracts (§§ 229–234 SGB III)

The integration contracts are a new measure introduced by the SGB III. These are temporary contracts aiming at improved hiring prospects for unemployed workers being classified as *difficult-to-place* workers. The principal idea is that employers abstain from taking the risk of hiring such workers, since they can not screen them sufficiently. Integration contracts constitute *no* formal work contract at present, but rather have the objective to pave the way for a permanent contract in the future. During the tenure of the integration contract, the unemployed workers should have had the opportunity to prove their ability and willingness to work and the employer should have had the chance to assess the quality of this person. The duration of such contracts is at most six months and can be terminated during this period by both sides without notice and without giving any reason. The employment offices reimburse any cost incurred through the absence of the employee, including the employers' share of social insurance contributions and, additionally, the employers can claim an integration subsidy.

c) Self-employment start-up and recruitment subsidies
(§§ 57–58 and 225–228 SGB III)

Compared with other countries the self-employment rates in Germany are quite low and some labor market observers see an increase in these rates as an avenue for reducing unemployment. The SGB III introduces two new employment promotion measures aiming at this aspect of the labor market. Firstly, people becoming self-employed in order to end or avoid unemployment can claim transitional benefit payments (*Überbrückungsgeld*) to secure a minimum standard of living during the first *six* months of self-employed work. The amount paid is the same as the unemployment benefit or assistance the individual has received or could have claimed, including social insurance contributions.

The prerequisites for entitlement are a minimum time of four weeks of receiving unemployment benefits or assistance and a statement by some designated expert (e.g. the industry and commerce chambers or banks) that the planned new business has a sound foundation. In addition to this start-up subsidy employers who have been running a new business for not more

than two years can claim a wage subsidy if they hire an unemployed person. As a prerequisite this person must be entitled to unemployment benefit or assistance payments and must be employed in a *newly* created position. Moreover, the employer must not have more than five employees. This subsidy is granted up to 12 months and covers 50% of the subsidizable wage.

(iv) *Active measures promoting the creation of jobs*

The employment offices have the possibility to support job creation schemes (*Arbeitsbeschaffungsmassnahmen*; §§ 260–271 SGB III), if the work done within these schemes is of public interest, e.g. projects aiming at the maintenance of public gardens. In addition, it is necessary that this work would either be delayed substantially into the future or not be performed at all without the support of the employment agencies. To be entitled for attending these schemes individuals usually have to be unemployed for more than one year and must also be entitled to some form of income-replacement benefit. Exceptions from this rule are possible under certain circumstances especially for the eastern part of Germany. Usually, the job creation schemes have to be carried out by private firms which can claim a subsidy of 50% to 75% of the subsidizable wage for each unemployed attending the scheme. For exceptional cases these grants can be increased up to 100%. The maximum duration is usually 12 months, but an extension up to 36 months is possible for special cases as well.

Moreover, it is possible to support *structural adjustment measures* (§§ 272–279 SGB III). Such measures are a special form of job creation schemes, originally introduced for the eastern part of Germany, which were extended to Germany as a whole in 1998 and are limited until the end of 2002. Such measures include e.g. environmental conservation and social services projects. The main difference to the usual job creation schemes is the fact that employers receive a *lump sum* subsidy equal to the average amount of unemployment benefit or assistance saved by the specific measure. The duration of the subsidy is usually up to 36 months but can be extended to 48 months if the employer is willing to hire the employee permanently after the subsidy expires.

In addition to the measures classified above the employment offices have the possibility to award grants for so called “social plan measures” (*Zuschüsse zu Sozialplanmassnahmen*; §§ 254–259 SGB III). This regulation aims at supporting workers at risk of being dismissed, in order to improve their chances to find a new job. For this purpose it is possible for the employment offices together with the employers to finance e.g. training measures to enhance the qualification of these workers. The decision whether to support such a “social plan” or not and, if so, the amount

granted, is under the responsibility of the *regional* employment offices. In 1998 the amount of money spent for such measures was negligible.

On the level of the regional employment offices the expenditure shares for the different groups of active labor market measures in 1998 are reported in *Table 2*. In West Germany most of the money spent for discretionary measures falls into category (ii) aiming at “human capital formation”. The shares for job creation schemes and the different incentives for employers are quite small. In East Germany approximately equal shares were spent for job creation schemes and measures aiming at “human capital formation”, whereas the money spent for direct incentives was even lower than in West Germany. In both parts of the country the money spent for guidance and placement assistance is very low. However, this does not accurately reflect the total amount of expenditures spent for the placement of unemployed into work since the cost for the staff of the employment offices and the various information services offered by them are not captured in these figures.

Table 2
Expenditures for Discretionary Measures in 1998

Measure	West Germany	East Germany
Guidance and Placement Assistance	0.4%	0.4%
Human Capital Formation	65.4%	45.8%
Incentives for Employers	8.2%	2.6%
Job Creation Schemes	16.4%	43.3%

Source: Bundesanstalt für Arbeit (1999a); own calculations.

4. Evaluation of Measures

Measures of active labor market policy (ALMP) compete with alternative programs for a substantial share of the tight public budget. Consequently, when engaging in ALMP policy makers as well as administrators are considered accountable by the general public for what happens with the taxpayers' money. For any informed judgement, though, the evaluation of effects and costs of an intervention is imperative, with the principal objectives of ranking alternative candidate interventions and assessing their cost-effectiveness. These insights are reflected, in principle, in the reform's requirement that local labor offices publish a balance sheet and offer a self-assessment of their success in alleviating unemployment.

Yet, contrary to what the wording of the reform law suggests, its direct implementation would neither enable labor offices to assess their own performance nor that of the various policy measures at their disposal. Far from

being a simple matter of accounting, the evaluation of policy interventions is a serious methodological challenge. The essential task for any evaluation analysis is the construction of a credible *counterfactual* situation – a precise statement of what would have happened in the absence of the policy intervention (see, for instance, Heckman et al. (1999), Manski (1995), or Schmidt (1999)). Only under very peculiar circumstances can this statement be based on an accounting exercise. In most cases, an elaborate scientific study is indispensable.

This section discusses the reform of employment promotion from the perspective of modern evaluation research. After stating the case for scientific evaluation, we clarify the nature of the major elements of program evaluation – the unit of analysis, the output measure, program cost, and the evaluation strategy – in this context. We then proceed to offer methodological guidance regarding the construction of the desired counterfactual situation, to discuss possible ways to construct this counterfactual, and to describe the natural limitations of any statistical inference. It will become transparent that while the legislator does not offer any real guidance as how to address the evaluation problem, the requirements of the reform law nevertheless generate the data basis needed for a credible evaluation of policy interventions in the labor market: the reform law and its implementation have to be understood as a beginning of this endeavor, not as its end.

4.1 Scientific Program Evaluation

Any serious evaluation effort has to follow well-accepted scientific standards – the strict reliance on evidence, a careful statement of data sources, consideration of sources of possible errors in inference, and the standard of publicness –, and cannot be done in-house as an addendum to the usual accounting procedures. This requirement does not at all question the honest planning, meticulous administration and careful delivery of policy measures by a well-trained and well-intentioned staff. Undeniably, though, the inception and design of the policy measure might rest on a false premise about the causes of the problem at hand. Second, one can hardly expect all participants in the design and delivery of policy measures to be completely impartial. Often the effects of policy measures are quantitatively small, and therefore even slight and inadvertent tendencies to emphasize positive aspects might invalidate their conclusions.

Finally and most importantly, attributing an effect to an underlying cause with considerable confidence is a task that is far more complex than is generally appreciated: in all instances, it requires the construction of a plausible *counterfactual* situation – identical to what is observed, apart from the

absence of the intervention – against which the actual situation has to be compared. Thus, at best the effect can only be estimated with confidence, but never measured with certainty. While in the US there is a long tradition of evaluating policy interventions scientifically on the basis of publicly accessible data and with publicized accounts of research methods and evidence (overviews are given by Friedlander et al. (1997), Heckman et al. (1999), Schmidt (2000), and Stanley et al. (1998)), this is underdeveloped in European economies (for an overview of recent European evidence see Kluge and Schmidt (2000)).

The role of the local labor offices and their advisory boards has been broadened by the reform. Not only do they have to implement the law, they now have to take an active role in the concrete design of policy. In this context, consider the stylized example of any given local labor office's choice of budget shares to be earmarked for either fostering the acquisition of human capital („training”) or for the creation of jobs in the public sector („job creation”). This example purposefully abstracts from many complexities, in particular the variety of measures subsumed under each heading – it is merely an instrument for organizing our thoughts. For ease of exposition, we also simply distinguish two possible expenditure levels, high and low, and presume that a high budget share for training implies a low budget share for job creation, and vice versa.

One way open to the advisory board for approaching the decision problem is to resort to economic theory. Unfortunately, as in most decision problems regarding economic policy, there are plausible arguments supporting both policy interventions. Training programs intend to enhance the human capital endowments of trainees and, since low human capital seems to be a major source of unemployment risk, more human capital might improve the labor market situation of the workers undergoing the program. This would speak in favor of allocating a high budget share to the first intervention. On the other hand, unemployed workers might foremost need the first step of getting back again into the labor market, thereby receiving the opportunity to retain their skills and to display their favorable characteristics – motivation and perseverance – to potential employers. This would argue for undertaking the second intervention.

A second approach would be to rely on the advice of experienced practitioners. These experts typically shaped their perspective on these matters while they were involved in the implementation of comparable policy interventions in the past, not in any controlled situation being designed for objectively evaluating effective output. Rather than the true program impact, this experience thus reflects the particular circumstances of these past interventions, most importantly the state of economic cycle, the characteris-

tics of the target population, and the specific design of the measures being implemented. Typically, those experts would equate previous success – whose accurate measurement had not been their priority – with efforts spent, both in terms of money and man-hours.

Third, one could consult the growing body of research on the evaluation of policy interventions. Unfortunately, none of the available studies would address exactly the same decision problem. In fact, much of the existing literature in policy-oriented journals is not aware of the conceptual subtleties of evaluation and, thus, not helpful at all. In essence, instead of a clear-cut answer to the decision problem, the best that a policy maker can hope for is a summary of the available evidence according to the well-respected standards of scientific research. One of these cornerstones of scientific research is the idea that only the weight of the evidence is able to answer any research question.

Finally, while this is no relief for the initial decision problem, as data become available over time, they can be used to perform an evaluation analysis whose results might then be able to guide future budget decisions. Yet, for this purpose one has to know (i) what questions to ask, (ii) how to extract the corresponding information from the available data, and (iii) how to assess the reliability of the results. The following sub-sections will make clear that even meticulous accounting for who participates in which intervention and experiences which post-intervention labor market success by itself does not answer any relevant evaluation question. By contrast, aspects (i) and (ii) necessitate a convincing empirical strategy tailored to the situation at hand.

Regarding the reliability of the results, even scholarly research is unable to eliminate all error. To derive an explicit assessment of remaining uncertainty, any study must invoke more or less restrictive assumptions – so-called *identification assumptions* – which are *assumed* to be true for the purposes of the analysis, and whose validity is not reflected in the usual measures of sampling variability. Indeed, more restrictive assumptions will generally lead to smaller sampling errors. But this raises the question whether the identification assumptions were correct to begin with.

In addition, well-respected scientific standards require that researchers publish their data and methods, thereby allowing any other researcher to assess in an independent replication of the evidence whether she would have arrived at similar qualitative conclusions. This requires also that the data pertaining to the local labor office's decision problem should be publicly accessible for research purposes. In the context of such a regionally aggregated analysis, questions of anonymity and data security – one of the central concerns when working with register data – are unlikely to pose the slightest problem whatsoever.

4.2 Constituent Elements of Program Evaluation

The first question to clarify in any evaluation study is what constitutes the relevant unit of analysis. This issue is determined by the nature of the intervention and the nature of the data that will become available for analysis. If the policy intervention under scrutiny targets individuals and individual-level data are collected, the formulation of the evaluation problem and its analysis will naturally be performed at the level of individuals. This would be the typical case, if the impact of a measure is investigated within a given region.

By contrast, if the intervention is delivered at the community level, but with varying intensity across regions, and the data will be collected at the regional level, the appropriate unit of analysis is the region. This is exactly the situation generated by the employment promotion reform. In this case we can expect the level and the particular mix of policy interventions to vary across regions, and we will be able to examine aggregate labor market indicators in a cross-regional comparison. Therefore, the subsequent discussion will treat regions as the unit of analysis⁴.

Second, we have to ask what should be considered as a success. In the case of employment promotion reform, any suitable outcome measure would arguably be a measure that captures whether the interventions are bringing the unemployed back into (stable) employment. But the same qualitative outcome may plausibly be measured in several ways, for instance by the regional job finding rate or by the regional employment or unemployment rates. In addition, interventions typically affect several outcome measures simultaneously. On the other hand, the interventions which are competing for shares of the budget might affect qualitatively different outcomes. Thus, the choice of the appropriate outcome measure can be quite complicated.

The third key aspect of the evaluation problem is the estimation of the costs of interventions. For any economic evaluation of policy interventions, valid estimates of the ensuing costs are as important as estimates of their impact. By contrast to the evaluation of program impact or *efficacy*, so far relatively little methodological work has been done on how to assess the *efficiency* of policy interventions, that is their impact per Euro spent. Even more detrimentally, in most economic evaluations of policy interventions so far only the direct outlays for the program have been determined, whereas the costs for the program participants and their families have been neglected.

⁴ A hybrid situation would be generated, if the intervention was at the community level, but for analysis a representative sample of individuals in every region was available. This situation is discussed in Augurzky and Schmidt (2000) and Schmidt et al. (1999).

The full costs of program participation include three components: the financial costs of treatment (the expenditures incurred by the participants, including fees, transport, and the costs of subsistence at a distant treatment site); the time cost of participation (the opportunity cost foregone by the participants); and the time cost of administrators and those involved in the delivery of the program (the opportunity costs of administrators). The latter cost component – including costs that arise from administrative overhead – is often neglected. However, it is quite naive to think that the efforts of existing administrative agencies are costless. Administrators being involved in this process could still perform another, potentially more valuable task.

Since the state of the labor market in any region is influenced by numerous factors, the fourth key issue is the attribution of labor market success to specific policy interventions. Since it will not be possible to measure the impact of any program with certainty, an appropriate empirical strategy has to be chosen for its estimation. Clearly not all estimation approaches are equally desirable. One would want to select an approach which would at least be able to yield the correct answer under ideal study conditions – conditions that could never hold in practical applications, such as the absence of any measurement error and an unlimited sample size. In technical terms, one would only select a strategy which *identifies* the entity of interest.

In the evaluation of interventions, the randomized controlled trial (RCT) is generally considered as the *gold standard*, particularly for medical interventions at the hospital level where treatments are randomized to different patients on an individual basis. The natural analog would be to randomize the budget share allocation across regions. The impact of the policy measures could then be evaluated by comparing average outcomes of regions with high budget shares allocated to training (the so-called „treatment group”) with those with low budget shares (the „control group”). Unfortunately, neither are controlled experiments easily implemented at the regional level (see Schmidt et al. (1999)), nor has scientific evaluation been a concern of the legislator at the time of employment promotion reform. The policy has been implemented long before a controlled experiment could be designed and executed. Thus, an appropriate non-experimental (also called *observational*) approach will have to be selected as the empirical strategy for evaluating measures of employment promotion.

4.3 The Formal Evaluation Problem

In recent years the evaluation literature in statistics and econometrics has developed a unified formal framework that facilitates the exploration of the potential and the limits of both experimental and non-experimental evalua-

tion strategies. Suppose that each region under study can be described by several key characteristics. Denote the state of affairs associated with allocating a high budget share to training by “1”, and that associated with a low budget share for training measures by “0”. The budget allocation is indicated by the region-specific indicator variable D_i . That is, if region i spends a large budget share on training, then $D_i = 1$. What we would like to compare is what would happen to region i in terms of the regional labor market, if i spent a large budget share on training ($D_i = 1$), as well as if i did not ($D_i = 0$).

Specifically, the relevant labor market outcomes in post-treatment period t are denoted by Y_{ti} if region i did not spend much on training, and by $Y_{ti} + \Delta_i$ if region i spent generously. This setup directly allows the formulation of the *causal impact* of allocating a high budget share to training (and correspondingly a low budget share to job creation) on the state of the labor market in region i as Δ_i . This concentration on a single region requires that the effect of the budget allocation on each region i not be affected by the budget decision of any other region. In the statistics literature this requirement is referred to as the *stable unit treatment value assumption* or SUTVA (Rubin (1986)).

Unfortunately, and this is the core of the evaluation problem, we can never observe Y_{ti} and $Y_{ti} + \Delta_i$ *simultaneously* for a given region – a region can either be a high-training region or not. Instead, only one of these two outcome variables can actually be observed for each region i . That is, the outcome Y_{ti} is the *counterfactual outcome* for those regions who do spend substantial resources on training ($D_i = 1$), whereas $Y_{ti} + \Delta_i$ is the *counterfactual outcome* for those regions which do not ($D_i = 0$). It is the budget decision, that is the value of D_i , that decides which of the two entries will be observed.

To give further structure to the discussion, presume that the underlying frequency distributions of the outcomes $Y_{ti} + \Delta_i$ and Y_{ti} across the population of regions (conceptually, think of actual regions being drawn from a large population of possible regions) are characterized by a set of region-specific characteristics Z_i and by pre-reform (period t') outcomes $Y_{t'i}$. That is, for each and every possible configuration of the characteristics $X_i = (Z_i, Y_{t'i})$, the respective conditional frequency distributions of Y_{ti} and $Y_{ti} + \Delta_i$ (which we do not know, but whose central aspects we want to estimate) describe the relative frequency with which every possible realization arises in the sub-population defined by X and $Y_{t'}$. Knowledge of these conditioning variables will allow to correct for “selection on observables!”. However, it is crucial not to condition on variables that themselves are outcomes of the policy measure under study, *exogeneity* of these conditioning characteristics is required: the budget decision must not alter the value of $(X_i, Y_{t'i})$ for any region i .

Since for each region i the respective counterfactual outcome is not observed, the available data comprise, in addition to observed outcomes Y_{ti} or $Y_{ti} + \Delta_i$, and characteristics X_i and Y_{ti} , the indicator for the budget decision D_i . In general, we would be very hesitant to impose that Δ_i is equal for all regions, not even for those regions sharing the same values of X_i and Y_{ti} . Some regions might be better off as a result of a specific budget allocation, some worse. There will thus be no opportunity to ever estimate region-specific gains with confidence. Interest in program evaluation is therefore on specific *evaluation parameters*, that is values that summarize the region-specific gains from being a high-training region appropriately.

In particular, one might still hope to be able to assess the population average (henceforth, population averages are denoted by the mathematical expectations operator $E(\cdot)$) of gains from awarding training a large budget share, since we know that the population averages of the frequency distributions of $Y_{ti} + \Delta_i$ and Y_{ti} can be estimated for high-training and low-training regions, respectively (since these are not the counterfactual, but the observed outcomes). One such evaluation parameter would be the so-called *mean effect of treatment on the treated*,

$$(1) \quad \begin{aligned} M_1(X) &= E(\Delta \mid X, D = 1) = E((Y_t + \Delta) - Y_t \mid X, D = 1) \\ &= E(Y_t + \Delta \mid X, D = 1) - E(Y_t \mid X, D = 1) \end{aligned}$$

conditional on the specific realization of the exogenous variables (note that individual indices are suppressed when expressing population averages). This parameter appropriately summarizes the region-specific gains in the population of high-training regions.

Alternatively one might consider the *so-called mean effect of treatment on a region randomly chosen*,

$$(2) \quad \begin{aligned} M_2(X) &= E(\Delta \mid X) = E((Y_t + \Delta) - Y_t \mid X) \\ &= E(Y_t + \Delta \mid X) - E(Y_t \mid X) \\ &= E(\Delta \mid X, D = 1) \cdot P(D = 1 \mid X) + \\ &\quad E(\Delta \mid X, D = 0) \cdot (1 - P(D = 1 \mid X)) \end{aligned}$$

where $P(D = 1 \mid X)$ denotes the share of high-training regions. This parameter summarizes the gains from budget re-allocation towards training to be expected for a region which is randomly chosen out of the population of regions.

After the population parameter of interest has been determined, one has to estimate population averages from the data in the sample. This estimate will unlikely be exactly the true population parameter itself. Instead, the es-

timate can only give an approximation to the true parameter, since it has been derived on the basis of only a subset of all members of the population. A successful estimation strategy requires that, as the sample taken from the population becomes larger and larger, the approximation become more and more exact. In the limit, the approximation should be indistinguishable from the true parameter. While in any finite-sample situation one has to accept some *noise*, one would certainly like to avoid *bias*. Noise are any *unsystematic* deviations of the estimate from the true value that would wash out in independent replications of the process of data collection and estimation if that process could indeed be repeated arbitrarily often. Bias is any *systematic* deviation of the estimate from the true value that would consistently arise in such repetitions.

A population parameter is *identified* from observable data, if it could be estimated correctly with infinite precision by collecting abundantly many observations from the underlying population. If the sample size could be made abundantly large, statistical inference would simply be based on relative frequencies, since the relative frequency distribution would converge (that is, come closer and closer until complete resemblance) to the probability distribution in the population. One of the two population averages featured in equation (1) is identified from observable data, while the other is not: in principle, one could estimate $E(Y_t + \Delta_i | X, D = 1)$ with considerable precision from the available data on high-training regions, but one could not even hypothetically estimate the population average $E(Y_t | X, D = 1)$, since no sample size would alleviate the fact that Y_{ti} is not observed for regions with a high training share. Similarly, $Y_{ti} + \Delta_i$ cannot be observed for low-training regions. For this reason, both right-hand-side entries of expression (2) are not identified.

These arguments characterize the fundamental problem facing program evaluation. This *evaluation problem* is the problem of finding appropriate identification assumptions that allow replacing the counterfactual population averages $E(Y_t | X, D = 1)$ in (1) and (2), and $E(Y_t + \Delta | X, D = 0)$ in (2), respectively, with entities that are identified from observable data. They are counterfactuals because they indicate what would have happened to high-training regions, on average, if they had only allocated a small share of the budget to training, and what to low-training regions, had they attributed a high budget share to training, respectively. Finding credible identification assumptions is a problem that cannot be solved by more or by refined measurement. It can only be resolved by finding a plausible comparison group of regions⁵.

⁵ Note that under some circumstances past observations of regions may serve as the comparison group.

In principle, three conceptually distinct and non-exclusive errors may plague any attempt at program evaluation. First, one might not find comparable low-training regions. For instance, it would be impossible to assess the impact of an intervention affecting regions with a large number of manufacturing firms (as opposed to say, high-tech regions), if every region with that characteristic spent a large share of its budget on training. In that case, the corresponding evaluation parameter is undefined. Second, while there might be comparable candidates among high-training and low-training regions for every configuration of observable characteristics, their relative shares might be disproportionate. For instance, if more manufacturing-based regions are among the high-training regions, but one were to take simple averages over high-training and low-training regions, then the average of the high-training regions' outcome might be relatively unfavorable. This is a problem that will be fended off by an appropriate conditioning on observable characteristics.

Third, there might be *selection bias*. Even when one compares comparable regions for all relevant configurations of observable characteristics *and* weighs the corresponding means appropriately, there might be unobservable factors that invalidate the comparison. In formal terms, we would have

$$E(Y_t | X, D = 1) \neq E(Y_t | X, D = 0),$$

For instance, some local labor offices might have a better connection to local employers than others, or might be more selective in accepting into their training measures only the most promising candidates among the unemployed workers. These might also be the regions where a high budget share for training is typical. Then, even if resources spent on training would have been small, the population average of the counterfactual outcomes Y_{ti} would have been higher among the high-training regions ($D_i = 1$) than the average observable outcome is among the low-training regions ($D_i = 0$).

Finally, it is important to realize what can and what cannot be identified by the information generated by employment reform. The technical apparatus developed above clarifies that variation over time of both intensity of measures and corresponding outcomes might serve to identify the effects of general policy shifts, while cross-sectional variation in the extent of training provides the opportunity to identify the effects of budget allocations. That is, while the creation of new data material is an important positive result of employment promotion reform – at last, data material will become available over a broad range of regions, enabling researchers to seriously address the evaluation of various policy measures–, the impact of the reform itself cannot be evaluated on the basis of this new, post-reform data.

If we were interested in how the general shift in employment promotion policy, most prominently the decentralization of budget decisions, affected the potential of the public sector to alleviate unemployment problems, we would require pre-reform information of the same breadth and quality. This temporal dimension would be necessary, since all regions were affected together by the reform, none was excluded. In its absence, we will therefore never be able to construct any credible counterfactual situation in which the only difference to the observable situation is the absence of decentralized decision making by local labor offices.

4.4 Experimental and Observational Studies

Empirical approaches typically follow a common principle of *analogy*. In order to formulate an estimate of population parameters, one searches for the corresponding concept in the sample at hand. Actual estimation in the sample is then performed by taking the appropriate sample averages. Since so much of the evaluation process depends on the data, improving the quality of the data on which evaluations are conducted should be a priority for future research. In this respect, in its request for the preparation and publication of data, the German employment promotion reform took an important step. Yet, it is instructive at this stage of the current paper to consider thoroughly what alternative study designs could be used to perform evaluation research, and to embed the potential offered by German employment reform into this general framework.

Under the fundamental requirement that an experiment completely replicate the intervention that will be implemented in the field, experimental studies generally are a convincing approach to the *evaluation problem*. The key concept of any experiment is the *randomized assignment* of regions into „treatment” and „control” groups. For regions who voluntarily would spend a large share of their budgets on training ($D_i = 1$) the random mechanism decides whether they are in fact allowed to allocate their budget in this fashion or not. This assignment mechanism is a process that is completely beyond the regions’ control and that also does not discriminate as to which region will spend substantial resources on training. Beyond the initial assignment phase, this approach requires considerable control by the researcher about the delivery of the intervention and about the compliance of regions with the experiment. In effect, if sample sizes are sufficiently large, randomization will generate a complete balancing of all relevant observable and unobservable characteristics across treatment and control groups, thus facilitating comparability between experimental treatment and control groups.

As long as the randomization is uncompromised (and samples are not outrageously small), there is no need for any sophisticated statistical analysis. However, a randomized controlled trial might not be a feasible approach at all, for political, ethical, logistic, or financial reasons, or a randomized trial might be contaminated by influences beyond the control of the researcher designing the study. This holds *a fortiori* for community-based interventions (see Schmidt et al. (1999)). Communities are not simply large-sized individuals, their decisions regarding budget allocation and regarding their compliance with the experiment are rather the consequence of the complex aggregation of their members' preferences.

By contrast to experimental analyses, in non-experimental or *observational* studies (a seminal source is Rosenbaum (1995)) the data are not derived in a process that is completely under the control of the researcher. Instead, the possibility to allocate a large budget share to training might have been particularly attractive for some regions. These might be regions, for instance, whose administrators entertain close connections to the local business community. What is collected instead of the desired experimental data, is an account of how regions actually performed after budget allocation. For high-training regions this means observation of $Y_{ti} + \Delta_i$, for low-training regions observation of Y_{ti} . The objective of any observational study is to use this information in an appropriate way such as to replace the comparability of treatment and control groups *by design* – the objective of experimental analyses – by a plausible alternative identification condition.

Irrespective of the particular estimation approach chosen, all observational studies adhere to this common principle: in experiments, random assignment of treatment ensure a balancing between treatment and control groups of all aspects relevant to the process, observable and unobservable. The desire in any observational study is to use the observable information (on Z_i and on Y_{ti}) such that in sub-populations defined by these observables, for instance high-manufacturing regions with a low unemployment rate in a pre-reform period t' , any remaining differences between high-training and low-training regions – apart from the budget allocation – can be attributed to chance. Then, using a random sample from this sub-population, the impact of the program can be estimated by forming the difference between means of actual outcomes for high- and low-training regions. One of the considerations in choosing an appropriate identification strategy is sample size. One would certainly not place high confidence in averages taken only over a handful of regions. This is the final methodological aspect which we will address here.

4.5 Sampling Distributions

Whenever a sample is used to estimate a population average, the answer given by the estimate will unlikely be exactly the population parameter itself. Instead, the estimate can only give an approximation to the true parameter, since it has been derived on the basis of only a subset of all members of the population. Moreover, although a successful estimation strategy requires that, as the sample taken from the population becomes larger and larger, the approximation become more and more exact, it does not mean that one will ever receive the correct answer in any given estimation attempt.

Instead, what this strategy would suggest in a finite sample situation, is that if one were to perform many repetitions of the sequence *drawing a random sample – estimating the population parameter – storing the estimated parameter value*, then the central tendency of the resulting frequency distribution would be on the correct value. Yet irrespective of the sample size in each of these replications, there would be some dispersion around the true population parameter. Generally, this dispersion would decrease with growing sample size, but never vanish completely. This remaining uncertainty or *noise* about the true value should be reported in any decent empirical study in order to indicate, whether large confidence should be placed in the conclusions of the study or not. A reported impact estimate that is not accompanied by an indication of the sampling variability around it is absolutely worthless.

Typically, researchers report *standard errors* or *confidence intervals* to this effect. Yet, while this principle is conceptually clear to practioners, what is generally less appreciated is that the remaining uncertainty that will be reported will always reflect the researcher's conviction that all systematic deviations between the answer given by the estimation strategy and the true population parameter have been successfully eliminated by invoking the correct identification assumption. Stricter identification conditions typically lead to lower assessments of remaining uncertainty. Thus, by contrast to the perception still widely held among practioners small noise is not the only important aspect of an empirical study. Instead, any evaluation effort that wants to be taken seriously should aim at a convincing identification strategy that eliminates all systematic tendencies to deviate from the correct population parameter.

5. Conclusions

This paper places German employment promotion policy and its reform of 1998 into the perspective of contemporaneous economic research. It introduces the main characteristics of the regime change in 1998, especially the shift of budget authority towards the local employment offices. Furthermore, the different measures of employment promotion, subdivided into *non-discretionary* and *discretionary measures*, are explained and – in deviation from the wording of the law classified into broad, economically meaningful categories.

Moreover, the paper explains the problem of evaluating the impact of this reform, clarifying that the semi-aggregate nature of the regional data, in contrast to the usual individual level data, needs to be addressed carefully in any serious evaluation attempt⁶. Most importantly, it is demonstrated within the formal framework of recent evaluation research in economics and statistics, that the problem of evaluating the implementation of this reform cannot be solved by the input accounting approach demanded by the new law. Instead, we discuss the methodological issues involved in the construction of the desired counterfactual situation, outline possible ways to construct this counterfactual, and describe the natural limitations of any statistical inference.

In doing so it becomes transparent that, while the legislator does not offer any real guidance as how to address the evaluation problem, the requirements of the reform law nevertheless generate a step towards the data basis needed for a credible evaluation of policy interventions in the labor market. Moreover, it has to be understood that in the absence of extensive information on past history, employment promotion reform and the accompanying requirement for data collection might allow different measures of active labor market policy to be evaluated, but do not allow an evaluation of the reform itself since all regions experienced the same policy change at the same time.

These insights provide a starting point for further research. Any credible evaluation attempt aiming at policy implications has to be a mixture of two complementary levels of research. Firstly, on a semi-aggregate (regional) level one should assess the effect of a specific set of measures and its regional heterogeneity. In a second step, a thorough *process analysis* of the local spending decisions has to be conducted to uncover the underlying reasons for any such differences. Therefore, it is inevitable to collect additional data

⁶ For an application in an international context see e.g. Hamermesh and Trejo (2000); for a survey, Heckman et al. (2000).

supplementing the available figures regularly published in the balance sheets (*Eingliederungsbilanz*) of the local labor offices.

As already pointed out above it would be extremely helpful to know more details about the decision process in the local labor offices. Specifically, we would like to ask how the local employment offices assess the local labor market situation, which information is regularly collected in this endeavor and how it is processed. A serious evaluation of the effectiveness of ALMP at the local level and an understanding of the underlying reasons for success or failure require to know, for instance, if the expertise of academic researchers was involved, and if there is a system of formal or informal discussion with other offices and local employers. This exchange might touch upon the decision making process, experiences with implemented measures, the principal objectives of employment policy, and the actual budget shares disbursed to the different measures. Moreover, it would be interesting to know how local labor offices decided to measure the success of the chosen policy mix, and if a local monitoring system has been set up.

In this endeavor the role of scientific evaluation and the analysts performing this task should be seen as that of partners for administrators and policy makers. Only a joint effort of both sides of this partnership will promise to improve the design and implementation of employment promotion, thereby moving closer to the ultimate aim of finding the best mix of policy measures.

References

- Augurzky, Boris / Schmidt, Christoph M.* (2000), The Evaluation of Community-Based Interventions. A Monte Carlo Study. IZA · Discussion Paper No. 270.
- Bundesanstalt für Arbeit* (1999a), Daten zu den Eingliederungsbilanzen 1998. Sondernummer der Amtlichen Nachrichten der Bundesanstalt für Arbeit. Nürnberg.
- Geschäftsbericht 1998. Nürnberg.
- Clever, Peter* (1998);, Sozialgesetzbuch III – Das neue Recht der Arbeitsförderung. ZFSH/SGB 37, 3-21.
- Deutscher Bundestag* (1996), Drucksache 13/4941 vom 18. 06. 1996.
- Fitzenberger, Bernd/ Prey, H.* (2000), Evaluating Public Sector Sponsored Training in Eastern Germany. Oxford Economic Papers, forthcoming.
- Friedlander, Daniel/ Greenberg, David H. / Robins, Philip K.* (1997), Evaluating Government Training Programs for the Economically Disadvantaged. Journal of Economic Literature 35, 1809-1855.
- Gagel, Alexander* (1999), Einführung in das SGB III – Arbeitsförderung. Beck-Texte, 4. Auflage, IX-XXXII.
- Hamermesh, Daniel S. / Trejo, Stephen J.* (2000), The Demand for Hours of Labor: Direct Evidence From California. The Review of Economics and Statistics, 82, 38-47.

- Heckman, James J. / Hohmann, Neil / Smith, Jeffrey (2000), Substitution and Dropout Bias In Social Experiments: A Study of an Influential Social Experiment. *The Quarterly Journal of Economics*, 115, 651-694.
- Heckman, James J. / LaLonde, Robert J. / Smith, Jeffrey A. (1999), The Economics and Econometrics of Active Labor Market Programs. In: Ashenfelter, Orley and David Card (eds.) *Handbook of Labor Economics*, vol. III, Amsterdam et al.: North-Holland, 1865-2097.
- Hübler, Olaf (1997), Evaluation beschäftigungspolitischer Maßnahmen in Ostdeutschland. *Jahrbücher für Nationalökonomie und Statistik* 126, 21-44.
- Hujer, R. / Maurer, K.O. / Wellner, M. (1997), Estimating the Effect of Training on Unemployment Duration in West Germany – A Discrete Hazard-Rate Model with Instrumental Variables. *Frankfurter Volkswirtschaftliche Diskussionsbeiträge*, University of Frankfurt.
- Kluve, Jochen / Schmidt, Christoph M. (2000), Can Training and Financial Incentives Combat European Unemployment? A Survey of Recent Evaluation Studies. mimeo., University of Heidelberg.
- Lechner, Michael (1998), Training the East German Labor Force – Microeconomic Evaluations of Continuous Vocational Training After Unification. *Physica-Verlag: Heidelberg*.
- (1999), Earnings and Employment Effects of Continuous Off-the-Job Training in East-Germany After Unification. *Journal of Business and Economics Statistics* 17, 74-90.
- Manski, Charles F. (1995), *Identification Problems in the Social Sciences*. Cambridge, Mass. et. al.: Harvard University Press.
- Rosenbaum, Paul R. (1995), *Observational Studies*. New York: Springer Series in Statistics.
- Rubin, Donald B. (1974), Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology* 66, 688-701.
- (1986), Which Ifs Have Causal Answers? *Journal of the American Statistical Association* 81, 961-962.
- Schmidt, Christoph M. (1999), Knowing What Works. The Case for Rigorous Program Evaluation. IZA Discussion Paper No. 77.
- (2000), Training, Incentives, and Public Service Jobs. The North American Experience With Labor Market Programs. mimeo., University of Heidelberg.
- Schmidt, Christoph M. / Baltussen, Rob / Sauerborn, Rainer (1999), Evaluation of Community-Based Interventions. Group-Randomization, Limits and Alternatives. Dept. of Economics Discussion Paper No. 281, University of Heidelberg.
- Sozialgesetzbuch III – Arbeitsförderung vom 24. 03. 1997, Beck-Texte, 4. edition, 1999.
- Stanley, Marcus / Katz, Larry / Krueger, Alan (1998), Impacts of Employment and Training Programs: The American Experience, Background paper prepared for the British Chancellor of the Exchequer for the 1998 G-8 Meeting.