RWI : Discussion Papers

Michael Fertig and Jochen Kluve

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany

No. 13

RWI
ESSEN

RWI : Discussion Papers
No. 13

Published by Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstrasse 1/3, D-45128 Essen, Phone +49 (0) 201/81 49-0

Editor: Prof. Dr. Christoph M. Schmidt, Ph.D.
ISSN 1612-3565 – ISBN 3-936454-22-1

RWI : Discussion Papers
No. 13

Michael Fertig and Jochen Kluve

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany

RWI
ESSEN

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISSN 1612-3565
ISBN 3-936454-22-1

Michael Fertig and Jochen Kluve*

A Conceptual Framework for the Evaluation of Comprehensive Labor Market Policy Reforms in Germany

Abstract

Over the last year the German government has introduced a comprehensive set of labor market policy reforms, the so-called *Hartz reforms,* which aim at a significant reduction of unemployment. To this end, (a) many of the existing instruments of active labor market policy are modified considerably, (b) a set of new instruments is introduced, and (c) the administrative framework in which these measures operate is changed substantially. In order to be able to judge the success of these measures by the end of the current legislative period in mid-2006, the government has asked academic experts to set up an evaluation concept capable of generating reliable empirical evidence by that date. The task is therefore to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own, facing substantive constraints regarding data availability and a short time horizon. This paper presents such a concept. We discuss essential guidelines for an ideal evaluation design, conceptual and practical difficulties that arise in the context of evaluating the Hartz reforms, and ways to overcome these obstacles. After detailing the three main analytical steps – analyses of effectiveness, efficiency, and implementation and process analysis – we present the concrete evaluation design, specific methods applicable to particular instruments, and a sampling scheme for collecting the required data. In addition to the fact that our concept is directly implementable, it also has the advantage of being extensible for future evaluations.

JEL-Classification: J0

Keywords: Program Evaluation, Identification, Active Labor Market Policy, Policy Reform, Evaluation design

* Both: Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstraße 1/3, D-45128 Essen, and IZA, Bonn. The authors are grateful to Helmut Apel, Werner Friedrich, Helmut Hägele, Christoph M. Schmidt and participants of the BMWA workshop "A Concept for the Evaluation of the Suggestions of the Hartz-Commission", December 04-05, 2003, in Berlin for helpful comments. Corresponding author: Jochen Kluve, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, Fax: +49-201-8149-236, e-mail: kluve@rwi-essen.de.

1. Introduction

Every year Germany spends several billions of Euro on active measures of employment promotion, with the explicit intention to contribute to the reduction of unemployment. Yet, unemployment has turned out to be a persistent problem throughout the last two decades, raising doubts regarding the effectiveness of such employment promotion measures. In line with a general tendency to redefine and modernize various kinds of administrative institutions, several major reforms of employment promotion policy have been launched since 1998. The latest set of reforms in this row are the so-called *Hartz reforms*: These were first introduced at the end of 2002, and the legislation process continued through to the end of 2003.

The Hartz reforms constitute a comprehensive reform of labor market policy. Indeed, it seems safe to argue that this is the most far-reaching reform endeavor in Germany in the last decades, and the government pursues an ambitious set of objectives with these reforms. They are supposed, in particular, to contribute to a significant reduction of unemployment by a quicker and more sustainable job placement, as well as to the creation of new jobs and employment opportunities. To this end, (a) many of the already existing instruments of active labor market policy are modified considerably, (b) a set of new instruments is introduced, and (c) the framework in which these measures operate is changed substantially. The latter does not only apply to the new administration and management of labor market policy within the *Federal Employment Agency* ("Bundesagentur für Arbeit, BA"), but also, for instance, to tighter sanctioning measures that aim at imposing more pressure on unemployed workers to take up jobs.

The success of these reforms, however, is anything but guaranteed *a priori*. To learn about policy success, i.e. policy effectiveness, clearly requires systematic evaluation of any implemented measure. This task is especially challenging if policy changes are multifaceted, detailed, and comprehensive, as is the case for the Hartz reforms.

Over the last decades, European policy makers, and German policy makers alike, have not shown much interest in initiating or considering systematic evaluation of policy measures – unlike the US, where scientifically sound evaluations have commonly accompanied policy interventions at least since the 1960s (cf. Kluve, Schmidt 2002 for details on the US-European comparison in this regard). Fortunately, and only very recently, also a European "evaluation culture" has started to emerge, and Germany is no exception. In the particular case of the Hartz reforms, the German parliament ("Bundestag") explicitly connected the evaluation task to the implementation of the reforms – with the objective to be informed, on a scientifically solid basis, about reform effectiveness before the end of the current legislative period in

mid-2006. This decision led to the situation that, for the first time, academic evaluation experts are involved from the very beginning in the implementation and evaluation of policy reforms in Germany, which in turn will result in the first systematic evaluation study for such measures.

As the Hartz reforms are currently being set into practice, the task with which the government approached academics can be summarized as follows: to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own. The main restriction regarding this task is that the concept needs to allow for generating of robust evidence until mid-2005 (1st report) and mid-2006 (final report), respectively.

This paper presents the main features of the concept that fulfills this task, i.e. we suggest a conceptual framework for a comprehensive evaluation of the effectiveness and efficiency of the Hartz reforms.[1] Specifically, we discuss the *conceptual* and *practical* problems of evaluating current labor market reforms in Germany. We will devote particular attention to the following research questions on the individual, i.e. micro, as well as the macro level:

a) How can we assess the *effectiveness* of labor market reforms?

b) How can we assess the *efficiency* of labor market reforms?

c) How can we determine the *reasons* for the estimated effectiveness and efficiency of the reforms?

In this endeavor, it will be of prominent relevance to take into account the intimate connection between changes in the *general framework* of labor market policy and the *modification and introduction of particular instruments.* The interaction between these two reform components induces the difficulty to isolate the impact of a specific policy change. The latter, however, is a prerequisite to identify the *causal effect* of any intervention. Furthermore, many of the reforms induce a *universal treatment* situation, i.e. every unemployed individual is affected by these changes, rendering the construction of a valid comparison group difficult. We will discuss these difficulties in detail and provide suggestions for ways to overcome thesm.

The paper is organized as follows. Section 2 establishes the context of the evaluation. It discusses objectives and elements of the Hartz reforms, presents guidelines for a suitable evaluation concept, and details the conceptual and practical problems that arise for the evaluation in this specific context.

[1] The paper is based on a pilot study by RWI-Essen and ISG Köln (cf. Fertig et al. 2004) that develops the evaluation concept for the Hartz reforms *in extenso.* The pilot study was prepared on behalf of the German Federal Ministry for Economic Affairs and Labor ("Bundesministerium für Wirtschaft und Arbeit, BMWA"). A second pilot study by ZEW Mannheim also intends to formulate such a concept.

Section 3 focuses on the general methodological framework for the evaluation, i.e. how program effectiveness and program efficiency can be identified. In Section 4 we discuss the concrete evaluation design and delineate specific methods applicable to the elements of the Hartz reforms. This includes details on the appropriate sampling procedure. Section 5 concludes.

2. The Hartz-Reforms: Establishing the context of evaluation

2.1 Objectives of the Hartz reforms

The new measures and modifications of labor market policy in the framework of the Hartz reforms aim at quick and sustainable job placement, the creation of new employment opportunities, and the simplification of moving into employment. This serves the overall objective of reducing unemployment in Germany. Hence, success or failure of the Hartz reforms must be judged in relation to the claim that the reforms contribute in a sustainable way to combating unemployment. The evaluation concept must therefore focus on the employment performance of unemployed individuals as the central outcome of interest.

The universe of labor market policy reforms within the Hartz framework can be divided into two major groups.

I. Modification and Implementation, respectively, of particular – i.e. aimed at unemployed individuals or individual enterprises – instruments of labor market policy.

- Direct wage subsidies ("Eingliederungszuschüsse", EGZ)
- Self-employment start-up subsidies ("Ich-AG", "Überbrückungsgeld, ÜG")
- Qualification measures ("Förderung der beruflichen Weiterbildung", FbW)
- Regulations regarding low-wage jobs ("Mini-Jobs")
- Regulations regarding low-wage jobs subject to social insurance contribution ("Midi-Jobs")
- Personnel Service Agencies ("Personal-Service-Agenturen", PSA)
- Integration measures provided by a third party ("Integrationsverträge")
- Reform of social plan measures and structural short-term allowance ("Umgestaltung präventiver Instrumente, Transferleistungen")
- Reform of regulations regarding temporary help Service workers ("Arbeitnehmerüberlassungsgesetz")

- Reform of regulations regarding temporary contracts with old workers and wage subsidies for old workers
- Placement vouchers for private agencies ("Vermittlungsgutscheine")

II. Modifications in the general framework of labor market policy.

(a) Modifications regarding objectives, administration, and steering system.

- Merging of measures of direct job provision in the public sector ("Arbeitsbeschaffungsmaßnahmen", ABM, "Strukturanpassungsmaßnahmen", SAM)
- Simplification of calculation of unemployment benefits
- Output-oriented steering system ("Ergebnis-orientiertes Steuerungskonzept")
- "Job-Center" as uniform contact point for all job seekers

(b) Modifications regarding benefit regulations.

- Reducing of duration of unemployment benefit entitlement
- Tightening of benefit sanctions ("Sperrzeitregelung", "Zumutbarkeitsvorschriften")
- Merging of unemployment assistance and social assistance ("ALG II")
- Uniform regulation regarding employment duration necessary for benefit entitlement
- Social security coverage of persons in compulsory military and community service
- Obligation for early unemployment registration

Table 1 presents a detailed overview of these measures in terms of their respective core features and objectives. Clearly, it is a challenging task to evaluate individual policy measures separate from each other given interactions and overlap in their contents and objectives. The evaluation aims at two things: First, an assessment of the overall effect of the Hartz reforms in their entirety; second, to isolate and quantify the effect of particular elements of the reforms.

2.2 Guidelines for a suitable evaluation concept

As previously outlined, the task is to develop a ready-to-implement framework for the evaluation of the full set of reforms in their entirety as well as each instrument on its own. This task faces the restriction that the evaluation design needs to allow for generating of robust evidence until mid-2005 and mid-2006, respectively. Hence, we believe that the concept

appropriate to fulfill this task must follow a set of guidelines: Practicability, focus, consistency, multidisciplinarity, robustness, and extensibility. Specifically:

1) *Practicability*: Given the constraints of a short time horizon and limited data quality and availability, the evaluation concept must be *practicable*. The actual implementation must allow for generating of robust evidence until mid-2005 and mid-2006.

2) *Focus*: Given the multitude of objectives and sub-objectives of labor market policy, as well as the multitude of specific policy changes constituting the Hartz reforms, the evaluation concept must be *focused*. Since reduction of unemployment is the main goal of both the Hartz reforms and labor market policy in general, the evaluation concept must necessarily focus on the effect of the Hartz reforms on the employment situation of unemployed individuals. For some policy measures, the secondary objective "employability" rather than direct integration into the labor market has been formulated by policy makers as the outcome measure of interest – clearly, this outcome would be somewhat more difficult to measure and assess in the evaluation.

3) *Consistency*: The evaluation concept must offer a *consistent* framework for the analysis of well posed evaluation questions; a framework that allows for comprehensive answering of these evaluation questions and for generating robust and convincing evidence. At the same time, given the heterogeneity of policy measures within Hartz, this consistent framework must be sufficiently flexible to accommodate tailor-made evaluations of specific policies. The concept must be based on recurrent steps of analysis for each policy measure, following three central questions:

a) What are the effects of Hartz interventions, i.e. are these policy measures and policy changes effective?

b) What costs have been generated through these interventions, i.e. are these policy measures efficient?

c) What reasons can be identified for the observed effects?

4) *Multidisciplinarity*: For a comprehensive assessment of reforms within a consistent analytical framework, the evaluation concept must follow a *multidisciplinary* approach. This requires a balanced combination of sound theoretical analysis, econometric competency, and extensive knowledge of implementation and process analysis.

5) *Robustness*: The evaluation concept must encompass an analytical approach that – given time and data constraints – allows for generating of *robust* evidence. This requires an intimate connection of available data and empirical method. Application of highly complex methods that do not fit the

data, and that may generate results that are highly sensitive to specification changes, is certainly counterproductive.

6) *Extensibility*: The evaluation concept must allow for an *extension* of the evaluation horizon beyond 2006. Given the late coming into effect of some policy measures within Hartz, and the important question regarding long-term effectiveness, a continuation of the evaluation beyond 2006 is both sensible and necessary. Hence, the evaluation concept must be transferable in a straightforward manner to later evaluation dates.

In the light of these background guidelines, the main elements of the appropriate concept for evaluating the comprehensive Hartz reforms result in the following procedure in practice.

- Compiling and reporting of main statistical data (monitoring function).
- Analysis of the improvements of services provided by the employment administration (employment offices) as the main instrument for increasing the effectiveness of labor market interventions.
- Microeconometric analysis of the instruments of active labor market policy that are implemented or modified within the Hartz framework. This analysis focuses on establishing the effectiveness and efficiency of policy measures regarding the employment performance of unemployed individuals.
- Macroeconomic analysis of labor market policy reforms, taking into account regional, cyclical, and other factors.

A major objective of the Hartz reforms concerns the provision of "modern services" by the employment administration. The fundamental hypothesis is that the effectiveness of a particular instrument of labor market policy does not only result from the instrument itself, but also from the counseling, informing, program allocation, and monitoring provided by the employment offices. Quite clearly, the effectiveness of a training program depends on the extent to which program contents are congruent with market demands, on the type of quality assurance used by employment offices, and on the type of participants selected for a specific program. Therefore the evaluation concept encompasses as central elements the consideration of the quality of service provisions, i.e. essentially the quality of labor offices, and the implementation of the new control system within the employment offices.

2.3 Conceptual and Practical Problems of the Evaluation

Any evaluation study that aims at identifying the causal effect of a policy measure must isolate the intervention to the best possible extent, in order to avoid measuring the effect of various interventions overlaying each other. If

there are indeed various measures implemented at the same time – as is obviously the case of the Hartz reforms – then it is the task of the evaluation design to guarantee the identification of the isolated effect. This requires the construction of a genuinely comparable comparison scenario. The comparison scenario aims at answering the counterfactual question, what would have happened (in terms of the outcome of interest) if the policy measure had not been implemented? The difference in the outcome between the factual scenario – the policy measure being implemented – and the counterfactual scenario then measures the causal effect of the intervention. Quite intuitively the construction of the counterfactual scenario has to follow the principle "compare the comparable", since ideally the factual and counterfactual states should differ only in the policy intervention being implemented or not (and associated outcomes), and nothing else.

The evaluation of the Hartz reforms therefore faces *two conceptual problems.*

The intertwining of changes in the general framework of labor market policy and individual instruments of labor market policy

As discussed above, some measures of the Hartz reforms regard the general objectives, administration and steering system of labor market policy. Specifically, this concerns the new outcome-oriented steering system, which aims at simplification of administrative procedures as a means to let employment office staff focus more intensely on their main task, job placement of unemployed individuals. If this new system were indeed to work, and, for instance, employment office staff could allocate more time to support unemployed workers and select measures of active labor market policy suitable for them, then this would clearly also affect the effectiveness of such programs. Hence, the modification in the general framework overlays any comparison of the reintegration performance of an active labor market program before and after the reforms. This methodological problem would be even more severe, if actual support intensities under the new framework were to vary across employment offices.

It is possible, and necessary, to appropriately reflect the role of the concrete implementation of the new steering system in the employment offices on the reintegration performance of individual programs: Each evaluation study must include as control variables indicators for the quality of service provision for individual employment offices, in order to avoid potentially severe bias in the impact estimates ("omitted variable bias"). Such quality indicators comprise, for each employment office, the number of job search assistants relative to unemployed individuals, the amount of online job search facilities relative to unemployed individuals, the average duration of counseling interviews, etc.

Several specific policy interventions constitute "universal treatments"

Some policy measures – such as the Mini- and Midi-Jobs, respectively, and the start-up subsidies – either affect every person in the labor force in the same way, or access to the program is open to every job seeker. That is, these measures constitute "universal treatments", a fact that complicates the construction of a genuinely comparable counterfactual scenario, since the most straightforward scenario – "not affected by the policy measure" – does not exist. The identification of an appropriate counterfactual then requires meticulous construction of the comparison group. This is detailed for some measures in section 3.

In addition to the two conceptual problems, *two practical problems* also arise in creating the appropriate evaluation design for the Hartz reforms.

Implementation of the new steering system

Up to this moment – i.e. the point in time when the evaluation concept has to be set up – it remains unclear in what way exactly the steering system will be designed in detail, and to what extent it has been implemented in the employment administration. Moreover, it is still an open question which information on the interaction between employment office staff and unemployed individuals will be collected and eventually made available to the evaluators.

With regard to the design of the evaluation this shortcoming implies that the concept can only suggest an ideal procedure, and the actual evaluation will – where applicable – have to adjust or extend the set of indicators for the quality of labor offices correspondingly. As mentioned above, the importance of these indicators for a meaningful and comprehensive evaluation cannot be overemphasized: Without indicators for the quality of labor offices it is not possible to identify the causal effect of individual labor market policy measures – only the composite effect of single instruments together with the change in the general policy framework would be identifiable.

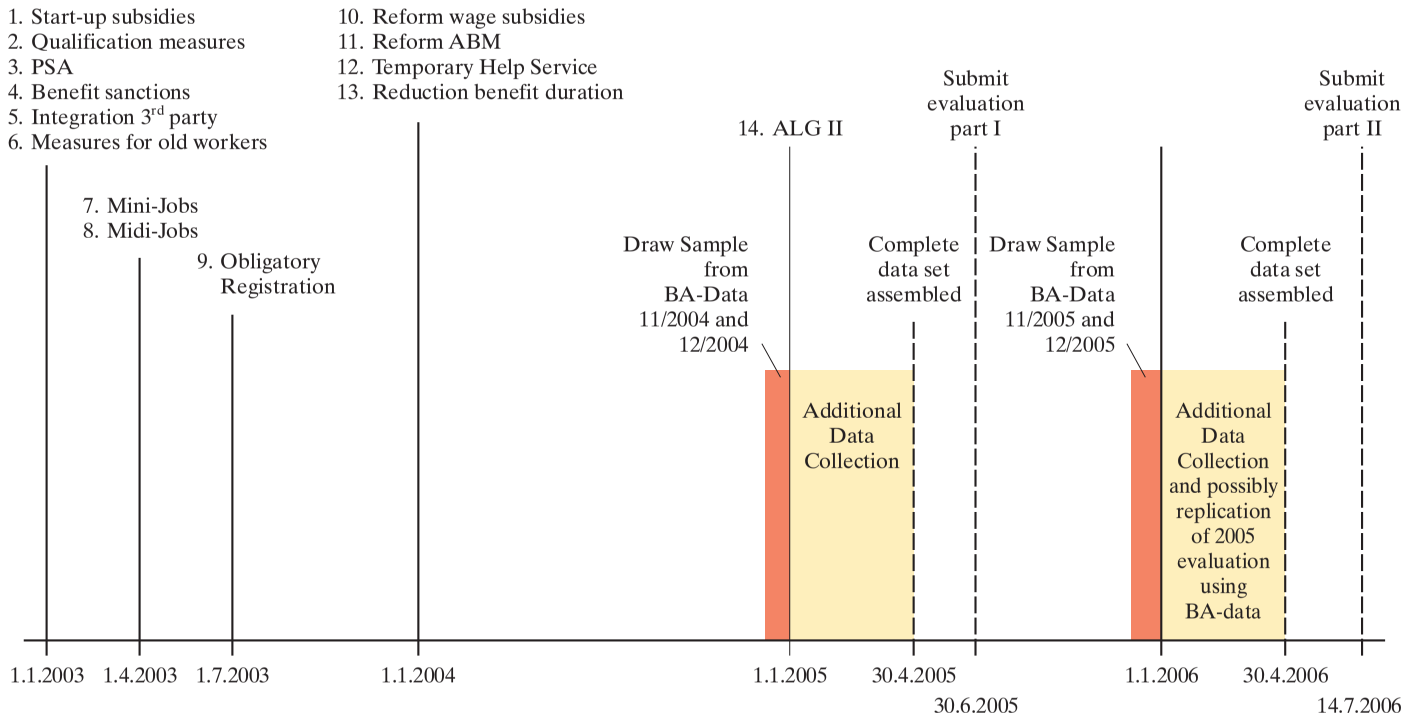
Data availability

Data availability is a crucial issue in setting up an evaluation concept. Currently, the data suitable for evaluation purposes provided by the employment administration are about 12 months behind real time. This lag, as pointed out by the employment administration, can at best be reduced to 8 months. Since the first evaluation report will be due mid-2005, the data gap must be overcome by additional data collection.

In principle, all necessary data on socioeconomic characteristics and, most importantly, employment histories of unemployed individuals are available,

Figure 1

Time Frame of the Evaluation



Note: Placement vouchers since 27.03.2002

and must then be complemented by data collection on *current* employment status and living situation. Moreover, depending on the specific program being analyzed, additional data on program participation must be collected. Moreover, data on the indicators for the quality of labor offices (cf. above) must be added. Gathering these data will take several months, and subsequently the evaluator will need at least two months for analyzing the data. Figure 1 depicts the resulting time frame for evaluating the Hartz reforms. The figure shows the starting date of several core measures of Hartz, along with the timing of sampling from the available data, additional data collection, and the evaluation.

Given the late starting date of some measures, the fact that program duration may be around 12 months, and the months needed subsequently for data collection and evaluation, it is clear that not all instruments of the Hartz reforms can be evaluated comprehensively until mid-2005 or mid-2006. Hence, the evaluation concept allows for a straightforward extension of the evaluation procedure beyond 2006, along the lines suggested in Figure 1, in order to assess long-term effects. The proposed time frame also has the advantage that, starting with 2006, each year the evaluation results of the previous year, which were then based on additional data collection, can be checked for consistency utilizing the now available register data from the employment administration (due to the 12 month time lag).

3. Methodological framework

The objective of any policy evaluation is the determination of the effects of the policy, its efficiency, and the reasons for success or failure. Specifically, these three steps are:

a) Analysis of *effectiveness*: "What is the impact of the policy measure?"

b) Analysis of *efficiency*: "What is the relation of benefits vs. costs of the measure?", or: "At what cost has the effect been achieved?"

c) *Implementation and process analysis*: "What are the reasons for the estimated effectiveness and efficiency?"

In principle this procedure holds for the evaluation on both micro and macro levels. Clearly, all three steps of the analysis are closely connected, and the omission of one of the steps would result in loss of important information for the decision maker. Specifically, simply focusing on step c) and presuming that a measure must be effective if it is merely well-implemented (a practice certainly not uncommon in the past in Germany) would be quite naïve. Instead, the crucial step for generating robust empirical evidence is a cogent analysis of effectiveness. Steps b) and c), while equally important, are logically subsequent, and aim at relating the established effect to its cost and possible reasons, respectively.

3.1 Descriptive analysis

The first part of the empirical analysis of effectiveness concerns a descriptive account of the policy measure under scrutiny. This account provides the basis for the subsequent analysis. In the context of measures of labor market policy, the following elements are of particular interest:

- Description of objectives, core features, state of the implementation, and possible regional differences of the policy measure.
- Assessment of incentive structures.
- Description of characteristics and composition of program participants.
- Gathering of information for identifying an appropriate comparison group, such as similarity in observable socioeconomic characteristics and employment histories.

The descriptive analysis mainly aims at delineating and summarizing core statistical data that relate to program implementation, thus putting the evaluation into the appropriate context. This mainly serves a monitoring function.

3.2 Analysis of Effectiveness

The first, and most important, step in evaluating a policy measure regards the identification of its effect. Estimating the causal effect of a program in general involves answering a so-called *counterfactual* question, i.e. making a statement about the hypothetical state of the world in the absence of the program. For instance, if we are interested in the effect of a qualification measure, and we observe an outcome of interest – such as employment probability – for program participants, then we would need to answer the counterfactual question: What would have happened to program participants in terms of their employment probability, if they had not participated in the program?[2] Whereas, obviously, program participants cannot be observed in two states of the world – with and without participation – at the same time, answering the counterfactual is *the* methodological challenge of any program evaluation.[3] Without answering the counterfactual it is not possible to assess the impact of a policy measure, since it is exactly the difference between factual and counterfactual outcomes that measures the causal effect of the program.

[2] The evaluation parameter corresponding to this counterfactual is the so-called "Average Treatment Effect on the Treated".

[3] Hence, this problem is frequently referred to as the "evaluation problem", cf., for instance, Heckman et al. (1999) or Kluve, Schmidt (2002). The causal model underlying this counterfactual notion of causality is commonly labeled "potential outcome model", since only one of the two outcomes required for causal inference is actually observable, and the second one is not, and thus a potential quantity. For further methodological details cf. Holland (19869 and Kluve (2004).

The fact that the counterfactual is not observable is called, in technical terms, *not identified*. It is therefore necessary to consider one or more *identification assumptions* that allow replacing the unobservable counterfactual with an observable counterpart. Ideally, the counterfactual would differ from the factual state in merely the fact that the policy measure was not implemented.

Besides these methodological challenges, several practical steps are important in analyzing program effectiveness. First, the unit of observation needs to be well-defined. In most cases concerning the Hartz reforms, unemployed individuals will be the units of interest. In addition, for some measures, individual enterprises will constitute the appropriate units of observation. Second, the outcome of interest has to be singled out. Usually the outcome of interest corresponds to the objectives that the policy measure intends to achieve. It may be the case, however, that a program pursues several objectives at the same time, and that some of them may be competing objectives. In the case of the Hartz reforms, the most interesting outcome by far is employment, both since almost all particular measures aim at increasing individual participants' employment probability, and since the overall objective of the Hartz reforms is a reduction of the unemployment rate.

Constructing the appropriate counterfactual, however, remains the most challenging task of the evaluation analysis. Given a modern economy with continuously growing production possibilities, hard-to-identify cyclical swings around a presumed growth path, and substantial heterogeneity in almost all economic aspects across individuals, sectors, and regions, it is clearly a difficult exercise to specify if some outcome of interest has taken on a "relatively high" value or not. Hence, finding a suitable comparison scenario determines success or failure of any impact analysis.

One important aspect in deciding on an identification strategy is individual heterogeneity of program participants. Quite clearly, program participants will differ from non-participants in a whole set of attributes, some of which will be observable, and others will be unobservable. Observable heterogeneity includes variables such as standard socioeconomic characteristics (age, gender, marital status, etc.) or individual employment histories. These characteristics are contained in the available data, and can therefore be utilized in the analysis, i.e. they can be controlled for when comparing the outcome of the group of participants (usually called "treatment group") with the outcome of a comparison group constructed from the pool of non-participants.

Unobserved heterogeneity, however, regards characteristics that are not contained in the data. We cannot observe whether an individual is characterized by e.g. motivation, stamina, or diligence. Unfortunately, these could be characteristics that partly determine the outcome of interest. If the

treatment group systematically differs from the comparison group in one or more of these characteristics, then estimates of the program effect will be biased. Hence, when choosing an identification strategy it is a core challenge to take into account unobserved heterogeneity to the best possible extent. It then depends on the study design whether the strategy will be successful.

In general, an experimental study design is considered the "gold standard" and conceptually most convincing strategy for identifying causal effects. Since in an experimental study a population of individuals eligible for program participation is randomized into a treatment and control group, it can be shown that there will be no systematic differences – neither observable nor unobservable – between the two groups. Therefore the control group provides the counterfactual in a straightforward and conceptually conclusive way. Most evaluations of measures of active labor market policy, however, cannot rely on experimental data. Experimental evaluations in the labor market context have almost exclusively been put into practice in the US, with a very small number in Europe, and none of them in Germany. Likewise, the evaluation of the Hartz reforms relies on non-experimental data only.[4]

The statistical and econometrics literature discusses a large number of possible identification strategies based on different assumptions. All identification assumptions have in common that they are not statistically testable, i.e. they cannot be right or wrong *a priori*, or proven correct or false *a posteriori*; they can only be more or less plausible, or more or less easily violated. The literature contains many examples showing that, in analyzing the same program, different identification assumptions can lead to different impact estimates.

Choosing the appropriate evaluation strategy, i.e. identification strategy, for a specific program therefore involves the collection of relevant information that justify the identification assumption. This information could come from details on e.g. program features, program participants, and program implementation, etc., but it generally requires detailed knowledge of program specifics. In the context of the Hartz reforms, one core variable in constructing comparison groups is the employment history. It has been frequently argued in the literature that such employment histories may capture possible differences between individuals due to unobserved heterogeneity to a large extent (Card, Sullivan 1988; Heckman, Smith 1999; Kluve et al. 1999). Moreover, the appropriate evaluation strategy for the Hartz reforms relies on controlling for that type of observable heterogeneity originating in differences in the quality of services provided by the labor offices. Also the labor market

[4] For an overview of evaluation studies of German labor market policy prior to the Hartz reforms cf. Fitzenberger, Speckesser (2000); Fitzenberger, Hujer (2002). Kluve, Schmidt (2002) provide a review of the experience with active labor market policy in a European context.

context, specifically local or regional labor market conditions, constitute important control variables.

The following identification strategies are commonly used in evaluation research:

A) Nonparametric methods

- Before-After
- Difference-in-Differences (DiD)
- Matching

B) Parametric methods

- Linear regression
- Fixed-Effects Panel model
- Duration analysis

C) Instrumental Variables methods

- "Classic" IV method
- Arellano-Bond dynamic panel method

D) Discrete Choice methods

In the context of the Hartz reforms, in the majority of cases non-parametric methods, in particular matching and DiD, will be the appropriate identification strategy. Some examples are given in section 4, where we discuss such suitable evaluation designs for specific instruments.

The basic idea of matching methods is to mimic a randomized experiment *ex post*. Utilizing information on a set of observable characteristics X, matching constructs – from a pool of potential comparison units – a retrospective comparison group as similar or comparable as possible to the treatment group in terms of X. The comparison group thus substitutes for the experimental control group. The main difference is that, whereas randomized assignment in an experiment balances both observable and unobservable attributes across treatment and control groups, matching can only control for observable covariates. The identification assumption, which matching is based on, is commonly referred to as "conditional independence assumption" (CIA) or "unconfoundedness". Essentially it says that selection into treatment and comparison group is based on observables, and that, conditional on X, the non-participation outcome of the participant population (i.e. the desired counterfactual) can be replaced by the non-participation outcome of the non-participant population.

A related approach, the DiD method compares the difference in outcomes of participants and non-participants before and after the intervention. The difference of the two differences then measures the treatment effect. The approach is based on the identification assumption that unobserved heterogeneity between the two groups is intertemporally invariant; i.e. any differences in the outcomes due to unobserved characteristics do not change over time and will therefore be "differenced out". Further details of these and other identification strategies can be found in Heckman et al. (1999), and Blundell/Costas-Dias (2000); a discussion of identification strategies in the context of the Hartz reforms is given in Fertig et al. (2004).

3.3 Efficiency analysis

Building on the impact analysis, the next step in evaluating a policy measure is to ask with which expenses the established effect has been achieved. This assessment of the benefits of the measure versus the costs of implementation is called *efficiency* analysis. If the preceding analysis of effectiveness has found a (qualitatively) positive effect of the intervention, then this is merely a necessary condition for a successful policy measure. In order to judge the success of the program, this effect needs to be confronted with the costs of the program. Clearly, this judgment is a relative measure, but most likely a program that causes only moderate improvement of the initial situation would not be allowed to be very cost intensive.

While consideration of the full costs of the program is desirable, the measurable *direct* costs – such as program fees or subsistence of participants – constitute only part of the total costs. Depending on the specific measure, also *indirect* costs have to be taken into account. For instance, opportunity costs may arise for program participants or firms: A program participant cannot exercise a job and effectuate earnings during program participation, and a firm may face additional administrative costs when managing a wage subsidies program. Moreover, it could be the case that a program has unintended negative side effects. For instance, a training measure could result in training participants displacing workers who did not participate in the program, or the training of participants would have been implemented even in the absence of government funding (cf. section 3.5).

Whereas, in principle, all direct costs can be taken into account in an efficiency analysis, this is much more difficult for indirect costs. Frequently, for instance, only coarse estimates for opportunity costs for workers and firms can be included. Unintended negative side effects can usually be assessed on an aggregate level, though methodological difficulties remain. However, consideration of program costs to the best possible extent is crucial for

determining the success of a policy measure, as otherwise evaluation results might seriously misguide policy makers.

3.4 Implementation and process analysis

The logically last step of a full evaluation is an analysis of the implementation and the processes of a program. This step aims at identifying the *reasons* behind the effectiveness and efficiency established in the preceding stages. Necessarily, the implementation and process analysis proceeds much less formally, and focuses on qualitative rather than quantitative aspects of the program.

An important part of this analytical step is a discussion of the policy framework and context within which the individual employment offices act, along with the type and design of their cooperation with third parties. These two aspects constitute the frame for the *input* of resources and the provision of services as concrete activities for the clients. On the other hand, the actual services provided constitute the *output* for the clients.

In the empirical implementation of the evaluation, much of the data required for this part of the analysis will come from surveys of (a) local employment office employees, and (b) program participants or, more generally, registered unemployed individuals, i.e. all those eligible to receive services from the employment offices. Typical questions for this analysis are (for a full catalogue of questions for each policy instrument, cf. Fertig et al. 2004):

- From the perspective of local employment office employees: Do possibilities for facilitating administrative procedures exist?
- Opinions on new/reformed instruments: Is the instrument (more easily) applicable?
- Are there any specific program features, specific efforts towards selecting or looking after particular workers, etc.?
- Reactions of program participants or of those affected by the policy measure.
- Analysis of data regarding indicators of labor office quality (cf. section 2.3): Are there systematic differences in program effectiveness between employment office districts? If so, how do they relate to differences in the quality of service provision?

3.5 Distinct features of the macro evaluation

In principle, the same methodological problems as on the micro level also arise for the evaluation on the macro level. The aim here is to identify the effect of the *entire* set of measures of labor market policy on an aggregate level. Again

this requires answering the counterfactual question: "What would have happened to the outcome of interest (such as the unemployment rate) if the Hartz reforms had not been implemented?"

Similar to the micro level, appropriate identification assumptions are needed for construction of the counterfactual. Attention must be paid, though, to the potential endogeneity of policy measures: A labor market policy influenced by the current state of the labor market would imply that certain outcome variables (such as the unemployment rate or the share of long-term unemployed) determine the design of labor market policy, and not vice versa. Moreover, possible interactions between specific measures, and the interplay between labor market policy on a nationwide level and the levels of the federal states, need to be taken into account.

With respect to efficiency, the interest lies in identifying detrimental side effects that may potentially arise on the macro level (cf. Calmfors 1994):

- *Displacement effects*: Program participants take on jobs that, in the absence of the program, would have been taken on by non-participants.
- *Substitution effects*: Labor demand for other types of non-participants decreases due to changes in relative wages.
- *Deadweight loss*: The state funds programs that would have been implemented also in the absence of state funding.
- *Tax effects*: Financing of the active labor market policy through taxes has effects on non-participants.

In addition, the reduction of costs that may be brought about by a reduction in the number of unemployed needs to be estimated.

With respect to the implementation and process analysis, the following criteria should be taken into account (For a full catalogue see Fertig et al. 2004):

- How do local employment offices determine local labor market conditions, i.e. which information constitutes the basis for decision processes on spending of available funds?
- What is the influence of employment offices on the federal state level in this regard?
- Is there any cooperation or exchange of information between employment offices in the same region?
- Is there a local monitoring system, if so, does it entail essential features, and do monitoring results influence future implementations of measures?

Of particular importance are those criteria that influence the quality of service provision.

4. Evaluation design

In this section, we delineate the salient practical components of an evaluation concept which adheres to the above mentioned guidelines and which is able to achieve the primary objective of generating *robust* and *comparable* evidence within the short time frame available for the concrete implementation of the evaluation. Clearly, the most challenging task in this endeavor is the generation of comparable evidence for the *effectiveness* of the different components of the Hartz reforms. Both the analyses of efficiency of the measures as well as the process and implementation analyses decisively depend on the results of the investigations regarding the measures' effectiveness. Therefore, we focus on the salient elements of the evaluation design aiming at this step of the overall evaluation strategy.

Table 2 provides an overview of the most important measures of the Hartz reforms, and details suggestions for the evaluation of their effectiveness[5]. Specifically, the table presents relevant counterfactual questions and outlines the main ingredients of an identification strategy able to provide an answer to these questions. We would like to emphasize that direct comparisons between some of the instruments mentioned in Table 2 are recommendable, in order assess which of the respective instruments is relatively more successful in promoting the job prospects of unemployed individuals.

To render such an approach feasible, a unified data collection and estimation strategy is indispensable. The cornerstone of our evaluation concept is therefore a unified scheme for the collection of individual data on participants in a particular program and their respective comparison group. Regarding the implementation of our concept, this crucial element implies that only *one* comparison group needs to be constructed that can be used simultaneously for the evaluation of the majority of instruments. Not only does this setup reduce the effort as well as the costs for the collection of data, but it also ensures comparability and unified interpretability of evaluation results across instruments.

In order to estimate the mean effect of treatment on the treated a variety of individual-level data for participants and non-participants as well as additional information is necessary. Specifically, we need *individual employee data*, especially the current labor market status and the employment history, as well as socio-economic characteristics and information on family background (labor market and earnings situation of partner and children). The labor market history as well as the family background information of the individuals

5 As pointed out various times throughout the paper, the Hartz reforms comprise a multitude of particular policy changes and policy interventions. For the sake of brevity, in this section we focus on the main elements in the whole set of reforms. For further details on secondary measures see Fertig et al. (2004).

are of particular importance, since this information is decisive for alleviating the potential problem of unobserved heterogeneity between participant and non-participant groups (section 3).

Furthermore, for some measures (especially the evaluation of Mini- and Midi-Jobs) it is necessary to collect *individual employer data*. That is, we need information on the number of employees (disaggregated by their social insurance coverage) as well as on company characteristics like the year of foundation, main sector of business, investment activities, organizational change etc.

Finally, data on the level of the local labor offices needs to be collected. Specifically, it is indispensable to collect indicators for the quality of service provision in the 180 local labor offices. Without such indicators the impact of the various components of the Hartz reforms cannot be identified, and cannot be isolated from the effect of the changes in the administration and steering of labor market policy.

The primary data source for providing this information is the *Federal Employment Agency*. However, due to the time lag in data processing or partially incomplete data for evaluation purposes (section 2), it is necessary to complement this information by additional data collection. Figure 2 and Figure 3 illustrate the unified data collection scheme for both evaluation reports due 2005 and 2006, respectively.

The idea behind this scheme is to draw random samples of participants for the various measures, who enrolled in these measures between January and June 2003 and 2004, respectively, from the database of the *Federal Employment Agency*. For these participant groups only one comparison group of untreated unemployed has to be drawn. [6] The participant groups as well as the common comparison group then have to be interviewed regarding their labor market status during 2004 and 2005, respectively, as well as with respect to their family background.[7] The collected information has to be matched to the existing data at the BA to construct the samples necessary for the empirical investigation of program effectiveness. Clearly, since all participants entered the different measures within the same period, it is possible to compare the labor market outcomes of different participant groups directly, in order to investigate which measure is relatively more successful in increasing the labor market prospects of the unemployed. Finally, for a small set of components of the Hartz reforms, especially those regarding benefit sanctions, it is necessary to construct additional comparison groups. However, this can be accomplished with comparably little effort.

---

6 "Untreated unemployed" denotes registered unemployed individuals who did not participate in any measure of ALMP.

7 For cost reasons our concept recommends doing this by sending out questionnaires, since the number of individuals to be interviewed is rather large.

Figure 2

Uniform Sampling Concept, short-term

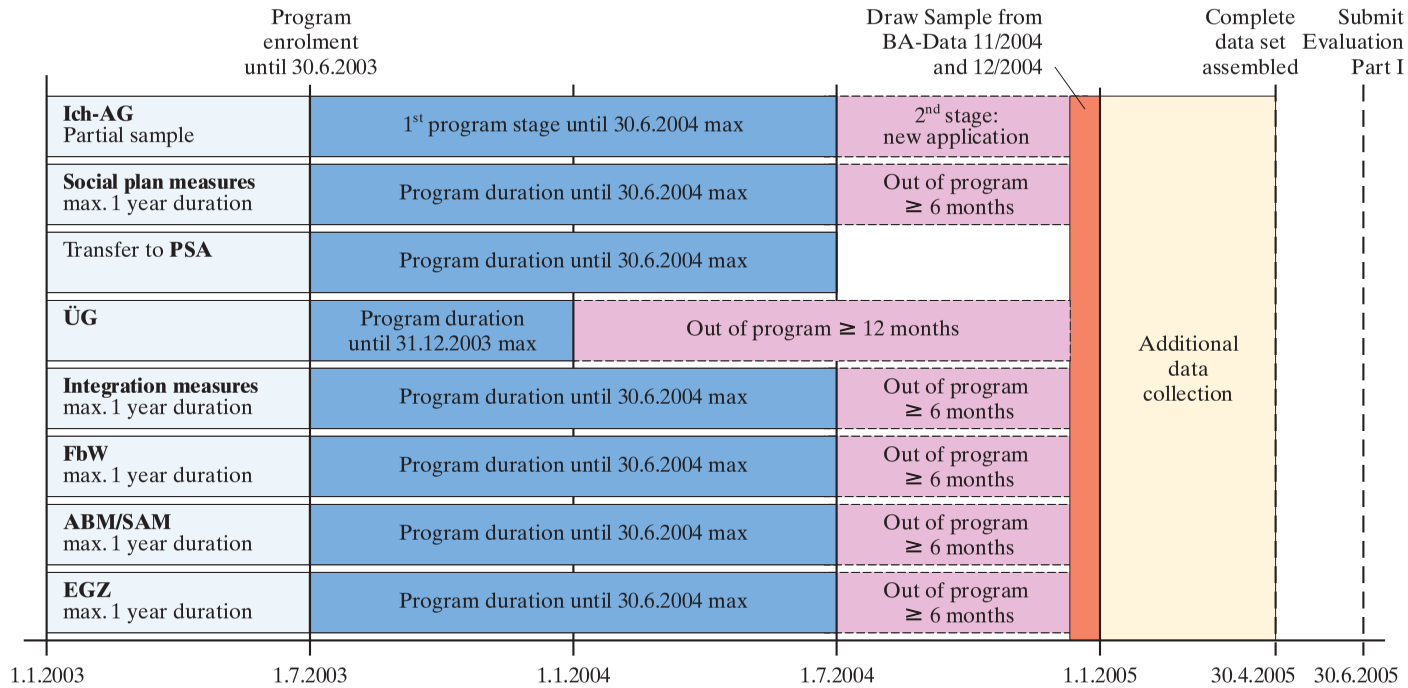
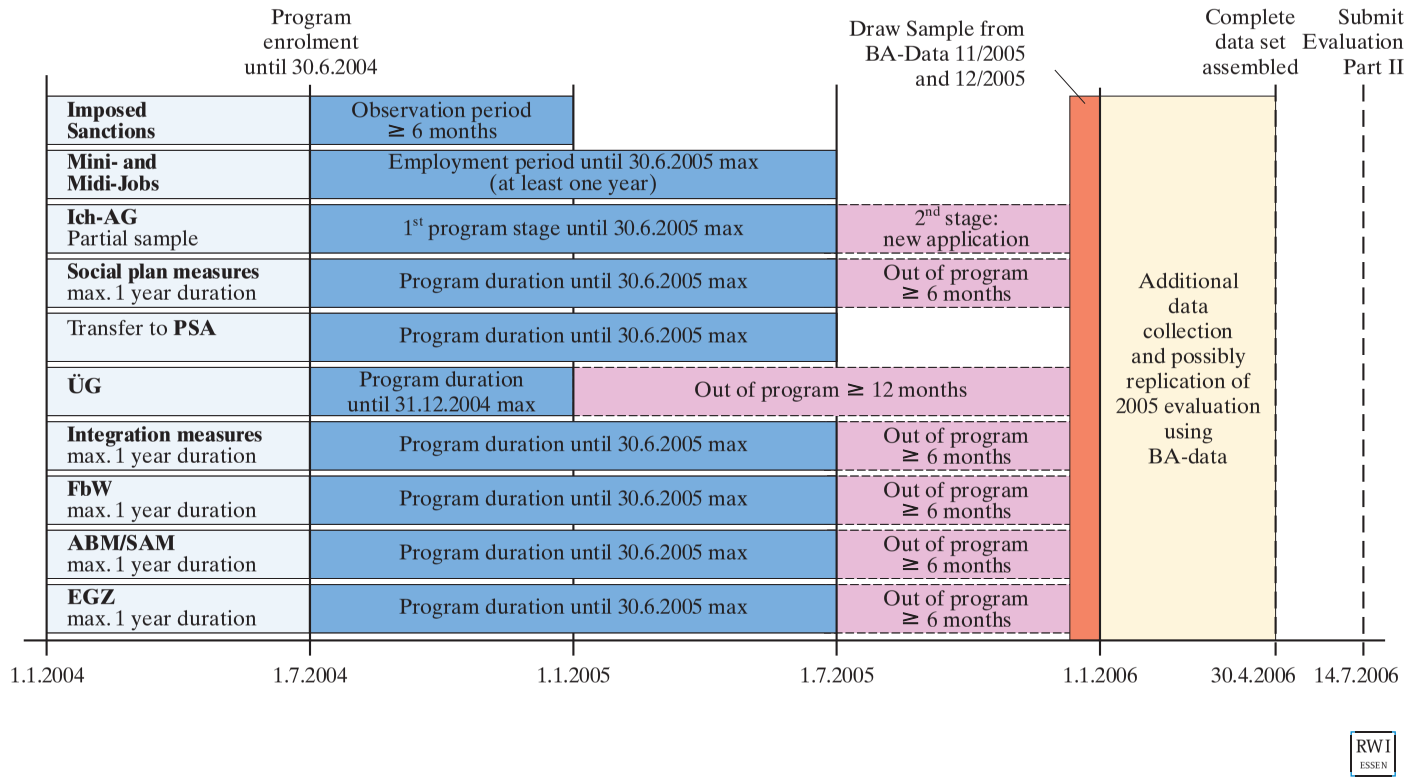


Figure 3

Uniform Sampling Concept, medium-term



5. Conclusion

It seems likely that the Hartz reforms are among the most ambitious, comprehensive, and far-reaching policy reforms in Germany over the last decades. Aiming at a sustainable contribution to reducing unemployment, the Hartz reforms entail a multitude of specific labor market policy changes, introduction of new instruments, and modifications of the general administrative framework in which German labor market policy operates. It is, however, anything but clear *a priori* that this bold endeavor will attain its objectives.

Learning about the success, predominantly the effectiveness, of any policy measure requires systematic evaluation. Given the manifoldness and scope of the Hartz reforms, this is a particularly challenging task. Fortunately, German policy makers have decided to rely on academic expertise in this regard, and to include academic experts from the very outset of evaluating the reforms, i.e. setting up an appropriate evaluation concept. This creates a novel situation for German evaluation practice in two regards: For the first time academic experts were consulted prior to the implementation of a policy, which in turn will result in the first systematic and comprehensive policy evaluation in Germany. The significance of this development cannot be overemphasized.

As the Hartz reforms are currently being set into practice, the task with which the government approached academics can be summarized as follows: to develop a ready-to-implement concept for the evaluation of the full set of reforms in their entirety, as well as each instrument on its own. The main restriction regarding this task is the time constraint: The concept needs to allow for generating of robust evidence until mid-2005 (1st report) and mid-2006 (final report), respectively.

In this paper, we have outlined such a concept. At the outset, the paper has discussed elements and objectives of the Hartz reforms, and has detailed essential guidelines for an appropriate evaluation framework. While, in principle, outlining an ideal evaluation concept, we have discussed limitations that arise for this concept both in practice and from a conceptual point of view. More importantly, facing these limitations, we have discussed ways to overcome practical and conceptual obstacles.

The paper has proceeded to discuss central methodological issues regarding the three main steps of an evaluation study, i.e. analyses of effectiveness, efficiency, and implementation and process analysis. Finally, we have presented the core features of a suitable evaluation design for the Hartz reforms, mainly focusing on tailor made evaluation approaches for specific policy measures, and a sampling scheme for the data collection that is both conceptually appropriate and practicable.

In summary, the evaluation design we suggest takes into account the contextual specifics of the Hartz reforms, provides feasible solutions to conceptual and practical problems, and will enable the evaluator to generate the desired empirical evidence given substantive time and data restrictions. In addition to the fact that our concept is directly implementable, it also has the advantage of being extensible in a straightforward way, once the data basis has been set up. Hence, we are confident that this concept for evaluating the Hartz reforms can contribute substantially to a cogent, comprehensive, and lasting evaluation of labor market policy in Germany in the long run.

References

Blundell, R. and M. Costas-Dias (2000), Evaluation Methods for Non-experimental Data. *Fiscal Studies* 21: 427–468.

Calmfors, L. (1994), Active Labour Market Policy and Unemployment – A Framework for the Analysis of Crucial Design Features. OECD Economic Studies 22. OECD, Paris.

Card, D. and D. Sullivan (1988), Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment. *Econometrica* 56: 497–530.

Fertig, M., J. Kluve, C.M. Schmidt, H. Apel, W. Friedrich, H. Hägele (2004), Die Hartz-Gesetze zur Arbeitsmarktpolitik – Ein Umfassendes Evaluierungskonzept. RWI : Schriften 74. Berlin: Duncker & Humblot. In preparation.

Fitzenberger, B. and R. Hujer (2002), Stand und Perspektiven der Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland. *Perspektiven der Wirtschaftspolitik* 3: 139–158.

Fitzenberger, B. and S. Speckesser (2000), Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland: Ein Überblick. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 33: 357–370.

Heckman, J.J., R.J. LaLonde and J.A. Smith (1999), The Economics and Econometrics of Active Labor Market Programs. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*. Vol. 3. Amsterdam: Elsevier, 1865–2097.

Heckman, J.J. and J.A. Smith (1999), The Pre-program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies. *Economic Journal* 109: 313–348.

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

Kluve, J. (2004), On the Role of Counterfactuals in Inferring Causal Effects. *Foundations of Science* 9: 65–101.

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

Kluve, J. and C.M. Schmidt (2002), Can training and employment subsidies combat European unemployment?" *Economic Policy* 35: 409–448.

Table 1

Elements and Objectives of Hartz Reforms

Measure	Key Elements	Objectives
	I. Introduction of New Instruments, and Reform of Existing Instruments of Active Labor Market Policy	
Direct Wage Subsidies ("Eingliederungszuschüsse", EGZ)	Reduction of the number of wage subsidies to two types: (i) EGZ for unemployed with integration problems and (ii) EGZ for unemployed with disabilities; reduction and unification of benefit duration and amount	(i) Integration of unemployed by temporarily subsidizing their wage (ii) simplifying the implementation of the instrument and increasing its attractiveness; (iii) increasing the efficiency of the instrument
Self-Employment Start-Up Subsidies ("Ich-AG"; "Überbrückungsgeld", ÜG)	Ich-AG: Temporary (max. 3 years) subsidy for newly self-employed if expected yearly income does not exceed 25,000 €; amount of subsidy declines every year; application must be renewed each year ÜG: Subsidy for newly self-employed for six months amounting to unemployment benefits or assistance	(i) Integration of unemployed by supporting self-employment; (ii) creation of additional jobs by newly established companies
Qualification Measures ("Förderung der beruflichen Weiterbildung", FbW)	Introduction of vouchers ("Bildungsgutschein") for choice of provider of qualification measure; certification of providers by external agencies	(i) Integration of unemployed by improving their qualification (ii) more competition among providers of qualification measures (iii) higher quality of measures; (iv) reduction of administrative effort
Integration Measures Provided by Third Party ("Integrationsverträge")	Possibility to award integration measures to providers by call for tenders; honorarium for especially successful measures; output-oriented steering of measures	(i) Integration of unemployed by customized and innovative measures (ii) increasing competition among providers of integration measures
Personnel Service Agencies ("Personal-Service-Agenturen", PSA)	Implementation of at least one PSA in every local labor office district; PSA employs unemployed workers with integration problems for a maximum period of 12 months and engages as a temporary help service company; during periods in which the former unemployed does not work, the PSA is supposed to provide adequate qualification measures	(i) Integration of unemployed by temporary help service jobs (ii) increasing the acceptance and quality of temporary help service work
Special Wage Subsidies for Old Workers ("Entgeltsicherung für ältere Arbeitnehmer") and Reform of Regulation Regarding Temporary Contracts With Old Workers	Special Wage Subsidy: Temporary allowance for workers aged 50+ if they take up employment with a lower wage compared to their last job before unemploy- ment; eligibility depends on a residual claim for unemployment benefits of at least 180 days; application by worker required Temporary Contracts: Reduction of age limit for temporary work contracts from 58 to 52	(i) Provision of incentive to take up jobs with a lower wage for old workers by partial compensation of loss in income (and pension claims) (ii) integration of old workers by facilitating temporary employment
Reform of Regulations Regarding Temporary Help Service Workers ("Änderungen im Arbeitnehmerüberlassungsgesetz")	Prohibition of synchronization, repeated dismissal and recruitment, and maximum duration of temporary help service jobs abolished; easing of exemptions for construction sector	(i) Integration of unemployed by temporary help service jobs (ii) increasing the acceptance and quality of temporary help service work (see also PSA)

Table 1 continued

Elements and Objectives of Hartz Reforms

Measure	Key Elements	Objectives
Transfer Measures and Transfer Short-Time Allowance (replacing former social plan measures and structural short-time allowance; "Umgestaltung präventiver Instrumente in Tranferleistungen")	Transfer measures (former social plan measures): All measures aiming at the integration of workers in danger of becoming unemployed because their company will be closed; employer has to take over 50 % of costs; transfer measures must be implemented by external provider who is obliged to have a quality assurance system; parallel promotion by other instruments no longer possible Transfer Short-Time Allowance (former structural short-time allowance): Profiling measure necessary for eligibility; reduction of maximum duration of entitlement to 12 months (instead of 24); employer is committed to provide placement assistance or adequate qualification measures	(i) Facilitate transition from work to work by a better co-ordination of instruments (ii) avoidance of transitory unemployment periods; (iii) abolish former practice of early retirement via structural short-time allowance
Placement Vouchers for Private Agencies ("Vermittlungsgutscheine")	Workers being three months or more unemployed can apply for a placement vouchers for a private agency; this voucher is valid for three months; private agencies receive honorarium depending on integration success	(i) Integration of unemployed by utilizing private placement services (ii) increasing competition for labor offices
Mini- and Midi-Jobs ("Geringfügige Beschäftigungsverhältnisse und Einführung einer Gleitzone in der Sozialversicherung")	Mini-Jobs: Jobs with wages up to 400 € per month are exempted from income tax and social security contributions for the employee; employer has to pay a 25 % flat-rate (taxes and social security contributions); this regulation also applies for secondary jobs; voluntary additional contributions to pension system possible; Mini-Jobs for private households receive privileged status; unemployed workers holding a Mini-Job can earn up to approx. 200 € without losing part of their benefit payments Midi-Jobs: Partial reduction of social security contributions for employees for jobs between 401–800 €; social security contributions increase linearly from 4.25 % to the full rate of contribution (approx. 21 %)	(i) Integration of unemployed by providing incentives to take up a Mini- or Midi-Job which is supposed to serve as a steppingstone into regular employment (ii) increasing the flexibility of employment in low-paid jobs (iii) reduction of non-wage labor costs; (iv) subsidizing low-income sector (v) reduction of illegal and shadow work (esp. in private households) (vi) reducing bureaucratic effort for employers (vii) creation of additional opportunities to earn money for workers in low-paid jobs
	II. Modifications Regarding Objectives, Administration, and Steering System	
Merging of Measures of Direct Job Provision in the Public Sector ("Zusammenlegung von ABM und SAM")	Merging of ABM and SAM into ABM-new; reduction of administrative complexity; participants are no longer covered by social security and therefore do not generate new benefit eligibility; objective of ABM-new no longer integration of unemployed, but employability	(i) Increasing employability of unemployed (ii) avoidance of crowding-out effects and benefit churning
Output-oriented Steering System ("Ergebnis-orientiertes Steuerungskonzept")	Implementation of target agreements and contract management system for the steering of labor market policy between the different levels within the BA	(i) Faster and more sustainable integration of unemployed by a more flexible and individually targeted labor market policy (ii) conversion of labor offices into customer-oriented service centers

Table 1 continued

Elements and Objectives of Hartz Reforms

Measure	Key Elements	Objectives
Job-Center ("Job-Center als einheitliche Anlaufstelle für Arbeitsuchende")	Implementation of a joint center for all recipients of ALG II (see below) providing them with all kind of services necessary for labor market integration	(i) Activation of unemployed (ii) counseling and monitoring of job seekers (iii) co-ordination of placement and advisory activities
	III. Modifications Regarding Benefit Regulations	
Reduction of Duration of Unemployment Benefit Entitlement	Reduction of duration to 12 months (for workers aged 55+: 18 months); transitory period until February 2006	(i) Reduction of unemployment duration (ii) avoidance of practice of early retirement
Tightening of Benefit Sanctions ("Sperrzeitregelung, Zumutbarkeitsvorschriften"")	Differentiated and unified system of temporary refusal to pay benefits in response to misconduct by unemployed; reversal of the burden of proof from labor office to unemployed Unemployed without family are obliged to move to another region if they receive a job offer	(i) More flexibility in sanctions system (ii) provision of incentives to take up jobs by reducing benefits (iii) increasing geographical mobility of unemployed
Merging of Unemployment Assistance and Social Assistance (ALG II)	ALG II as new benefit system for former unemployment assistance or social assistance recipients on the level of former social assistance; stepwise transition form unemployment benefit (ALG I) to ALG II by temporary extra-allowance; possibility to promote all recipients of ALG II by measures of ALMP; transitory period of 6 months for former recipients of unemployment assistance	(i) Integration of long-term unemployed by providing incentives to take up employment (lower level of benefits for former recipients of unemployment assistance and lower earnings threshold for former recipients of social assistance) (ii) reduction of expenditures for long-term unemployed
Obligation for Early Unemployment Registration	Obligation to register as unemployed immediately after notice of dismissal or 3 months prior to the end of a temporary work contract; upon violation: reduction of benefits possible	(i) Speeding-up of placement process (ii) avoidance of begin of unemployment spell

RWI ESSEN

Table 2

Main Elements of Hartz Reforms - Suggestions for Analyses of Effectiveness

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/ Comparison Group	Control Variables
			a) Public Job Provision Programs (ABM/SAM)[1]		
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6/12 months after end of program	Matching on observables (requires CIA to hold)	Participation in ABM/ SAM; comparison group: untreated unemployed[2]	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
			b) Qualification Measures (FbW)[1]		
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6/12 months after end of program	Matching on observables (requires CIA to hold)	Participation in FbW; comparison group: untreated unemployed[2]	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if the reform of FbW had not happened?	Individual unemployed	Employed (yes/no): 6/12 months after end of program	Difference-in-differences (ass.: inter-temporal invariance of unobserved heterogeneity)	Participation in FbW before and after the reform; comparison group: untreated unemployed[2] before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
			c) Direct Wage Subsidies (EGZ)[1]		
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Matching on observables (requires CIA to hold)	Supported by EGZ; comparison group: untreated unemployed[2]	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if the reform of FbW had not happened?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Difference-in-differences (ass.: inter-temporal invariance of unobserved heterogeneity)	Supported by EGZ before and after the reform; comparison group: untreated unemployed[2] before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
			d) Personnel Service Agency (PSA)[1]		
What would have happened to the employment situation of employees in PSAs, if they had not been employed there?	Individual unemployed	Regularly employed (yes/no) during or directly after PSA-period	Matching on observables (requires CIA to hold)	Employment in PSA; comparison group: untreated unemployed[2]	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation

Table 2 continued

Main Elements of Hartz Reforms - Suggestions for Analyses of Effectiveness

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/ Comparison Group	Control Variables
			e) Self-Employment Start-Up Subsidies (Ich-AG and ÜG)[1]		
What would have happened to the employment situation of participants, if they had not participated?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Matching on observables (requires CIA to hold)	Supported by Ich-AG or ÜG; comparison group: untreated unemployed[2]	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of participants, if they had received alternative funding?	Individual unemployed	Employed (yes/no): 6/12 months after end of subsidy	Matching on observables (requires CIA to hold)	Supported by Ich-AG or ÜG; comparison group: self-employed receiving Start-Geld or Micro-Darlehen from KfW	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation
			f) Mini- and Midi-Jobs		
What would have happened to the employment situation of unemployed taking up a Mini- or Midi-Job, if they had not done this?	Individual unemployed	Regularly employed (yes/no): 6/12 months after start of Mini- or Midi-Job	Matching on observables (requires CIA to hold)	Unemployed taking up a Mini- or Midi-Job; comparison group: untreated unemployed[2] without Mini- or Midi-Job	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation
What would have happened to the employment situation of unemployed taking up a Mini- or Midi-Job, if the reform of these jobs had not happened?	Individual unemployed	Regularly employed (yes/no): 6/12 months after start of Mini- or Midi-Job	Difference-in-differences (ass.: inter-temporal invariance of unobserved heterogeneity)	Unemployed taking up a Mini- or Midi-Job before and after the reform; comparison group: untreated unemployed[2] without Mini- or Midi-Job before and after the reform	Socio-economic characteristics; (un-) employment history; indicators for regional labor market situation
What would have happened to the number of Mini- or Midi-Jobs, if the reform of these jobs had not happened?	Individual companies	Number of employees in Mini- or Midi-Job	Regression model for period before reform; (ass. structural stability of estimates and evaluating post-reform covariates with them yields predicted jobs after reform); difference between actual and predicted number of jobs yields effect of reform	Comparison group: all other companies in the sample	Company-specific characteristics; indicators for regional labor market situation

Table 2 continued

Main Elements of Hartz Reforms - Suggestions for Analyses of Effectiveness

Counterfactual Question	Unit of Observation	Outcome Measure	Identification Strategy	Treatment Indicator/ Comparison Group	Control Variables
g) Merging of Unemployment Assistance and Social Assistance (ALG II)					
What would have happened to the employment situation of ALG II-recipients, if they had received unemployment assistance instead?	Individual unemployed	Employed (yes/no): 5 months after start of benefit receipt	Matching on observables (requires CIA to hold); utilizing transitional regulation	Receipt of ALG II and entitled for unemployment assistance in old regulation context; comparison group: receipt of unemployment assistance directly before the reform came into effect	Socio-economic characteristics; (un-) employment history; indicators for quality of labor offices; indicators for regional labor market situation
What would have happened to the employment situation of ALG II-recipients, if they had received social assistance instead?	Individual unemployed	Employed (yes/no) on a monthly basis	Before-after comparison in duration analysis framework	Receipt of social assistance or ALG II	Key socio-economic characteristics (family situation, age, education); duration of spell

Notes: [1]Direct comparisons (using matching techniques) between these instruments of active labor market policy recommended. – [2]Untreated unemployed denotes individuals receiving no special treatment, i.e. individuals not participating in any measure of active labor market policy.