

Huber, Martin; Lechner, Michael; Wunsch, Conny; Walter, Thomas

Working Paper

Do German welfare-to-work programmes reduce welfare and increase work?

IZA Discussion Papers, No. 4090

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Huber, Martin; Lechner, Michael; Wunsch, Conny; Walter, Thomas (2009) : Do German welfare-to-work programmes reduce welfare and increase work?, IZA Discussion Papers, No. 4090, Institute for the Study of Labor (IZA), Bonn,
<https://nbn-resolving.de/urn:nbn:de:101:1-20090330299>

This Version is available at:

<https://hdl.handle.net/10419/35506>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 4090

Do German Welfare-to-Work Programmes Reduce Welfare and Increase Work?

Martin Huber
Michael Lechner
Conny Wunsch
Thomas Walter

March 2009

Do German Welfare-to-Work Programmes Reduce Welfare and Increase Work?

Martin Huber

SEW, University of St. Gallen

Michael Lechner

*SEW, University of St. Gallen
and IZA*

Conny Wunsch

SEW, University of St. Gallen

Thomas Walter

ZEW

Discussion Paper No. 4090
March 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Do German Welfare-to-Work Programmes Reduce Welfare and Increase Work?

Many Western economies have reformed their welfare systems with the aim of activating welfare recipients by increasing welfare-to-work programmes and job search enforcement. We evaluate the three most important German welfare-to-work programmes implemented after a major reform in January 2005 ("Hartz IV"). Our analysis is based on a unique combination of large scale survey and administrative data that is unusually rich with respect to individual, household, agency level, and regional information. We use this richness to allow for a selection-on-observables approach when doing the econometric evaluation. We find that short-term training programmes on average increase their participants' employment perspectives and that all programmes induce further programme participation. We also show that there is considerable effect heterogeneity across different subgroups of participants that could be exploited to improve the allocation of welfare recipients to the specific programmes and thus increase overall programme effectiveness.

JEL Classification: J68

Keywords: welfare-to-work policies, propensity score matching, programme evaluation, panel data, targeting

Corresponding author:

Michael Lechner
University of St. Gallen
Swiss Institute for Empirical Economic Research (SEW)
Varnbühlstr. 14
CH-9000 St. Gallen
Switzerland
E-mail: michael.lechner@unisg.ch

1 Introduction^{*}

Over the last decade many OECD countries faced increasing numbers of welfare recipients. In Germany, for example, the number of recipients of welfare payments had risen to about 4.5 million people by the end of 2004. Many countries reacted by conducting welfare reforms that resulted in a shift of labour market policies from passive benefit receipt towards increased job search and work requirements among welfare recipients. Post-reform programmes typically focus on activation of welfare recipients to encourage employment and to reduce welfare receipt and incentives to stay welfare-dependent. Needy but employable welfare recipients are obliged to participate in activation programmes, and they can be sanctioned by benefit cuts when not complying.

Welfare research has traditionally focused on the USA, see for instance Grogger (2003), who investigates the effect of time limits for benefits and the Earned Income Tax Credit (EITC) on welfare receipt and work. Welfare-to-work efforts were considerably increased across US states over the 1990s. In the course of the reforms, an extensive literature evaluating the various welfare programmes has evolved, see Blank (2002) and Moffitt (2002) for a review of the US welfare reforms and of the related empirical literature.

In Europe, where unemployment insurance (UI) is usually more generous and where there are larger numbers of UI claimants than in the US, the literature has almost exclusively focused on the evaluation of programmes targeted at UI rather than welfare recipients (see e.g. the surveys by Martin and Grubb, 2001; Kluve and Schmidt, 2002; Kluve, 2006; Wunsch, 2006). However, the results are not easily extendable to welfare recipients because they differ

^{*} The second author has further affiliations with ZEW, Mannheim, CEPR and PSI, London, IZA, Bonn and IAB, Nuremberg. This paper is part of the project *Evaluation of the Experimentation Clause in §6c SGB II*. Financial support from Germany's Federal Ministry of Labour and Social Affairs (BMAS) is gratefully acknowledged. The data originated from a joint effort of ZEW/SEW with IAB, Nuremberg, TNS Emnid Bielefeld and IAQ, Gelsenkirchen to use administrative and survey data for welfare evaluation. The usual disclaimer applies.

considerably from regular UI recipients with respect to their labour market relevant characteristics and employment perspectives.¹ These differences may be particularly relevant as the programmes have been found to exhibit considerable effect heterogeneity with respect to participant characteristics (for Germany see for instance Caliendo, Hujer, and Thomsen, 2005, Lechner, Miquel and Wunsch, 2006, 2007, and Wunsch and Lechner, 2008). Recently, European countries conducted welfare reforms that also increase work incentives and job search enforcement for welfare recipients and that introduce substantial welfare-to-work programmes, thus raising considerable interest in the effects of these costly measures.² One of the most substantial welfare reforms, the German so-called Hartz IV reform, has taken place in the beginning of 2005.

In this paper we use regression adjusted caliper propensity score matching on unique data that combines exceptionally rich survey, administrative and regional data to evaluate the three most important welfare-to-work programmes used in Germany since the Hartz IV reform. The programmes considered comprise (i) very short training programmes that include basic job search assistance, work tests and minor adjustment of skills, (ii) further training that aims at improving job-related skills, and (iii) so-called 1-Euro-jobs, which are effectively similar to a work requirement with some small extra remuneration (workfare).³ The programmes started between October 2006 and March 2007, so that we naturally must focus on short-run outcomes 6 to 12 months after programme start.

¹ Eligibility for either welfare benefit receipt or UI hinges on an individual's unemployment history. Thus, welfare recipients and unemployed are by definition distinct with respect to characteristics relevant for labour market success.

² Surveys on welfare reforms in European countries are provided by Torfing (1999), Kildal (2001), and Halvorsen and Jensen (2004) for the Nordic countries, Finn (2000) and Beaudry (2002) and Dostal (2008) for the UK, Finn (2000) and Knijn (2001) for the Netherlands, and Wunsch (2008), Jacobi and Kluve (2007) and Konle-Seidl et al. (2007) for Germany.

³ The following recent papers look at other policies targeted specifically at German welfare recipients: Bernhard et al. (2008) study wage subsidies, Wolff and Nivorozhkin (2008) investigate start-up programmes and Schneider (2008) analyses benefit sanctions.

We find no significant effects of the programmes on the likelihood of future welfare receipt and that programme participation induces further subsequent programme participation. With respect to the employment effects of the programmes, we find positive and significant effects for some programmes and groups of participants, in particular for short training and for welfare recipients without a migration background. Our results are in line with Wolff and Jozwiak (2007) who investigate the effect of participation of welfare recipients in short-term training, as well as with Hohmeyer and Wolff (2007) who evaluate the effectiveness of 1-Euro-jobs. Both studies use propensity score matching. However, they solely rely on administrative data and focus on programmes that started at the beginning of 2005, i.e. directly after the reform. This period was characterised by strong implementation and data collection problems, which may have affected their results. By considering more recent programmes, our findings cannot be attributed to those temporary phenomena of the introductory phase of the new regime. We also provide more robust evidence because we use much more informative data than the earlier studies.

Moreover, we add to the literature in further dimensions: (i) We also evaluate more substantial further training that provides job-related skills. (ii) We investigate effect heterogeneity in a detailed way and investigate a variety of outcome variables, thus providing considerably more comprehensive results than earlier studies. (iii) We assess the optimality of the allocation process of welfare recipients to the different programmes and find considerable scope for improvement with respect to both taking up employment and leaving welfare.

The remainder of the paper is organized as follows: Section 2 provides background on the economic conditions and relevant institutions in Germany since 2005. In Section 3, we introduce the data and our evaluation sample followed by a discussion of the definition of programmes and participation. Section 4 displays descriptive statistics for the evaluation sample. Identification and estimation of the effects of interest as well as the simulation of

alternative allocations into treatments are discussed in Section 5. In Section 6, we present the effect estimates and simulation results. Section 7 concludes.

2 Economic conditions and institutions in Germany since 2005

2.1 German unemployment insurance and welfare

Recent reforms of German welfare and labour market policies focused on the activation of welfare recipients based on improved employment services to enhance individual employment prospects ('Fördern') and on making greater demands on individuals to actively participate in and speed up the reintegration process ('Fordern'). The so-called Hartz reforms⁴ were gradually implemented in the beginning of 2003 (Hartz I and II), 2004 (Hartz III), and 2005 (Hartz IV).⁵ Jacobi and Kluve (2007) provide an excellent survey of the reform package.

Before 2005, unemployed with no or expired unemployment benefit entitlements (henceforth UB) were either eligible for unemployment assistance (UA), which was conditional on previous employment, or for social assistance (SA), or a combination of both (if UA was 'too low'). Both UA and SA were means-tested. When Hartz IV and the Social Code II came into force in 2005, unemployment benefit II (UBII) replaced both UA and SA. In contrast to UA, which replaced up to 57% of the previous net earnings, UBII, like the former SA, does not depend on former earnings. Furthermore, it is means-tested and the test is based on the wealth and income of all individuals in the household.

At the beginning of 2005, the standard UBII amounted to 345 EUR in West Germany and 331 EUR in East Germany. Meanwhile, the level of UBII in East Germany was adjusted to the western level and UBII was slightly increased in both parts to compensate for inflation (351

⁴ The reforms were named after the chairperson of the commission proposing the reforms, Peter Hartz, who was a Human Resources executive and a member of the board of executives of the German car producer VW until July 2005. Ironically, Hartz was convicted of embezzlement in January 2007.

⁵ Hartz I-III focused on labour market institutions and unemployment benefit recipients, whereas Hartz IV is targeted at welfare recipients.

EUR in January 2009). Besides the standard UBII, welfare payments also include compulsory social insurance contributions, rents and housing costs. Further costs for special needs might be covered as well. According to Ochel (2005), standard UBII is less generous than former UA (on average EUR 550 in 2003 in West Germany).

UBII access is conditional on claimants being 'employable', i.e. on being capable of working for at least 15 hours per week. Employable claimants have to register with the local employment office. One important innovation is that welfare recipients are obliged to participate in welfare-to-work programmes. The job seekers' rights and duties in the activation process are usually set out in writing in a so-called integration contract. This binding agreement between the employment office and the welfare recipient contains obligations concerning programme participation and job search activities as well as services provided by the employment office. Non-compliance and/or the rejection of 'acceptable'⁶ job offers can be sanctioned by temporary benefit cuts.⁷

The administration of activation programmes and welfare receipt is in most cases executed by local agencies that are formed as joint ventures between the local employment office of the public employment service (PES) and the local community. However, in 69 out of 429 offices the agencies are run by the local community alone, entirely outside of the responsibility and competency of the PES.⁸

The Hartz IV reform constitutes a remarkable change in German welfare policy. For the first time welfare recipients are a target group of labour market activation. Before 2005, almost no effort was made to reintegrate those persons into the labour market. Thus, there is neither

⁶ According to the legislation, almost any job is acceptable, even if it does not correspond to the individual's former profession or education.

⁷ Though job seekers might be threatened with temporary benefit cuts, Jacobi and Kluve (2007) argue that they are rarely enforced in practice, since they frequently entailed costly lawsuits with benefit claimants.

⁸ In this study, we only consider the regular joint ventures.

experience nor any evidence on the effectiveness and efficiency of welfare-to-work programmes prior to the reform.

2.2 *Economic conditions in Germany*

The Hartz-IV reform came into effect in a period of mild recovery of the German economy. After stagnation and a decline in GDP in 2002 and 2003, GDP grew moderately in 2004 (1.1%) and 2005 (0.8%). In 2006, GDP growth was up to 2.9% while 2007 saw a moderate slow down (2.5%).⁹ The number of persons receiving welfare amounted to roughly 4.5 million in January 2005. It increased steadily during 2005 – partly due to the adjustment to the new welfare system – and reached a peak of 5.5 million in April 2006. Since then it has declined to just fewer than 5 million in August 2008. In January 2005 there were 2.3 million unemployed persons receiving welfare. This number increased during the following months to peak at roughly 3 million at the beginning of 2006. Since then unemployment among welfare recipients declined to 2.2 million in August 2008.¹⁰

2.3 *German welfare-to-work programmes since 2005*

This section gives a brief overview over the most important activation programmes targeted at welfare recipients. The post 2005 programmes are more in favour of a 'work first' approach, i.e. their primary goal is to integrate benefit claimants into the labour market quickly. Table 1 displays entries into the most important activation programmes for the period 2005-2007 along with the corresponding annual expenditures in millions of EURO. The so-called 1-Euro-jobs are by far most frequently assigned, followed by short trainings and further training. Expenditures on short trainings are comparably low due to their short duration. Other important categories are job placement by third parties, wage subsidies, and start up grants. Let us now discuss the different programmes in more detail.

⁹ Figures according to the Federal Statistical Office; see www.destatis.de.

¹⁰ Figures according to the Federal Employment Agency; see www.arbeitsagentur.de.

Table 1: Entries in and expenditures for selected activation programmes*

Category	Entries			Expenditures in million €		
	2005	2006**	2007**	2005	2006	2007
Short training	410,884	480,675	545,960	157.5	164.1	163.3
Further training	69,906	124,169	167,200	196.3	377.6	503.7
1-Euro-jobs/other public employment progr.	633,938	815,380	798,774	1104.5	1381.2	1321.5
Job placement services by third parties	272,627	153,381	119,390	62.9	63.7	47.5
Wage subsidies	60,675	111,372	135,806	145.7	316.7	408.2
Start up grants	20,097	48,751	52,718	21.9	63.7	71
Others	291,475	536,408	666,319	1435.9	1473.8	1706.2
Total	1,707,410	2,270,136	2,486,167	3124.7	3840.8	4221.4

* If not stated differently, figures are for joint ventures alone.

** Includes both joint ventures and community controlled agencies.

Source: Statistics of the Federal Employment Agency at <http://www.pub.arbeitsamt.de/hst/services/statistik/detail/e.html>.

Short training courses are comparably cheap programmes with durations of usually a few days to two weeks, but in any case not more than 12 weeks. They are rather heterogeneous with respect to their content and target group and pursue two main objectives. Firstly, they are used to check the welfare recipients' occupational aptitude and availability for the job market, as well as to provide basic job search assistance. Typical examples are job application and job interview trainings. Secondly, the focus of short training is on minor adjustments of job relevant skills. This includes language courses, computer classes, and occupation specific trainings. Quantitatively, short training is rather important as a programme type, as in 2007, roughly 546,000 welfare recipients received some form of short training.

Further training comprises a more substantial human capital investment and focuses on the adaption of the professional skills and qualifications of participants to recent labour market requirements, e.g. to mitigate mismatches in times of structural change. The programmes typically aim at adjusting skills in the profession held, the completion of an additional qualification, receiving a first professional degree, retraining, and the participation in practice firms simulating a job in a specific field of profession. Their planned duration reaches from a few months to up to three years.

The so-called *1-Euro-jobs* are public-sector-related workfare programmes that were introduced in 2005. In contrast to short and further trainings, which are also open to unemployed not receiving welfare, 1-Euro-jobs exclusively apply to the activation of welfare recipients. According to the legislator, these programmes should be of public interest and 'additional' in the sense that the work assigned as 1-Euro-jobs would otherwise not be accomplished by existing public and private sector companies.¹¹ Exiting to regular employment, if possible, has clear priority over carrying out 1-Euro-jobs. The work load typically consists of 20-30 hours per week over a period of 3-9 months. Legally, 1-Euro-jobs do not constitute any form of employment. Thus, participants do not receive a (subsidized) wage, but merely a 'compensation for job related extra costs' that is paid additionally to standard benefits. The name of the programme originates from this compensation being set to 1 EUR per hour in many cases. Since their introduction in 2005, 1-Euro-jobs have largely substituted other forms of public employment programmes. In 2007, almost 800,000 individuals were assigned to 1-Euro-jobs (or other employment programmes). Expenditures amounted to more than 1.3 billion Euros representing more than 30 % of total spending for activation programmes.

In order to introduce more competition with respect to the placement of welfare recipients, the Hartz reforms opened the market to *job placement services* of private companies ('third parties') who compete with public institutions. The market is, however, only quasi-liberal, as services are bought through vouchers or by means of public biddings organized by the local agencies, instead of direct intervention by welfare recipients (see Bernhard et al., 2008). Third parties might either be partly or exclusively involved in the job placement activities. The remuneration of private providers by the agency is partly dependent on the placement success.

¹¹ Critics who doubt the usefulness of workfare programmes therefore argue that they merely create 'symbolic', non-productive employment without providing marketable skills to the participants, see for instance Dostal (2008).

Job placement by third parties has been decreasing in numbers, but amounted still to almost 120,000 in 2007.

Wage subsidies are paid to firms which employ individuals facing competitive disadvantages on the job market during the first months of employment. They should generate an incentive to hire such individuals by compensating employers for initial productivity gaps. Roughly 136,000 jobs for welfare recipients were subsidised in 2007.

Similarly, *start up grants* are bridging allowances for taking up self-employment. They are either granted to young firms hiring welfare recipients in the form of wage subsidies or directly to benefit claimants as promotion of self-employment and private start ups. More than 20,000 welfare recipients benefitted from start up grants in 2007.

3 Data and definition of sample and participation status

3.1 Data

Our analysis is based on a unique and exceptionally informative data set that combines various data sources characterizing welfare recipients. The core of these data is a survey of welfare recipients who have been interviewed in two waves at the beginning (January - April 2007) and around the end of 2007 (November 2007 - March 2008). The survey consists of about 25,000 realised interviews in each wave including both a stock sample (roughly 21,000) of welfare recipients in October 2006 and a small inflow sample (roughly 4,000) from August to December 2006. Despite 93% of interviewees agreeing in the first wave to participate in the follow-up interview, attrition was non-negligible, mainly due to 'relocation problems' and 'refusal to participate'. To make up for these losses, the second wave contains a refreshment sample (7,086) that was drawn from the same population as the panel sample (13,914). The new participants of the refreshment sample had to answer retrospective questions to make up for the information collected from the panel cases in the first wave.

It is important to note that our sample is not drawn randomly from the population of welfare recipients, but is instead stratified.¹² Stratification is based on the following individual characteristics: age (15-24 / 25-49 / 50-64), children under 3 in the household, and being a lone parent. This is done to ensure that the number of observations is sufficiently high for these groups, each of which resembles one target group of welfare-to-work programmes. The data contain sample weights, denoted by η_j in the following, for each individual j in the sample that take into account both stratification and attrition.

One problem with the survey data is that information is not symmetric in the panel and refreshment sample, because the retrospective questions for the latter do not fully match the questions of the first wave. In particular, there is an information asymmetry concerning programme starts in the two subsamples.

Despite this problem, the survey is unique with respect to the information available for German welfare recipients and household characteristics as well as sample sizes. The survey includes individual characteristics such as gender, age, marital status, education, nationality and migration background. It also contains details concerning labour market status, welfare receipt, participation in activation programmes, past performance on the labour market and job search behaviour. Finally, it includes information on the household such as the number, age and employment status of other household members as well as the interviewees' relation to them.

These survey data have been merged with administrative data on welfare recipients provided by Germany's Federal Employment Agency (FEA) for the period 1998-2007. They combine spell information from social insurance records, programme participation records and the benefit payment and jobseeker registers of the PES. The database comprises very detailed

¹² In addition, our sample is restricted to a subgroup of agencies. However, the sampled agencies and the composition of welfare recipients within this subgroup is very similar to all other agencies in Germany.

information in several dimensions. Personal characteristics include education, age, gender, marital status, number of children, profession, nationality, disabilities and health. The benefit payment register provides information on type and amount of benefits received as well as remaining benefit claims. The jobseeker register includes additional information on the desired form of employment and compliance with benefit rules. Moreover, the data comprise information on previous employments including the form of employment, industry, occupational status and wages. With respect to programme participation, the data covers the type of the programme as well as its actual duration, and the planned duration (for training only).

The combined administrative and survey data were linked to further data at the agency and regional level. They include a wide range of regional information reflecting labour market conditions (e.g. share of unemployed and long term unemployed, share of the elderly among unemployed, share of welfare recipients, GDP per worker, share of migrants, population density, industry structure) and variables that characterize the agencies' organisational structure (e.g. generalised or specialised case management, number and qualification of caseworkers, welfare recipients per caseworker, placement strategies, counselling concept).

3.2 Sample and treatment definition

The inflow sample of about 4,000 individuals is too small for this application. We therefore evaluate programme effects for the stock sample, which is drawn from the population receiving welfare in October 2006 (sampling date). Note however, that this sample is endogenous with respect to previous programme participation, as it contains in particular those persons who did not succeed in exiting welfare receipt, whereas successful exits are not observed in the sample. Therefore, treatment effects prior to October 2006 cannot be identified. The sample is, however, exogenous with respect to programme participation after the sampling date conditional on prior programme participation, benefit receipt, employment,

and further characteristics. We therefore confine our analysis to the effects of programme participation after October 2006.

To obtain sufficiently large samples, we use both the panel and the refreshment samples. However, information asymmetries related to differences in the survey design of the two samples imply that start dates coming from the survey data can only be consistently identified for all individuals from January 2007 onwards. For November and December 2006, only the start dates in the administrative records are considered. As survey data are only available up to the second interview, which took place between November 2007 and February 2008, and administrative records end in December 2007, our evaluation window is rather short. Thus, we restrict attention to the first programme (after the first sampling period) that starts before April 2007 in order to have a follow-up period for measuring the outcomes which is not too short.

To obtain sufficiently large samples the programmes are aggregated to broader treatment categories, see Table 2. Nonparticipants are defined as those individuals not receiving any treatment between November 2006 and March 2007. Among the treated categories, the programme groups 1-Euro-jobs, short trainings, and further training with a planned duration of up to 3 months have sufficient numbers of observations to estimate programme effects semiparametrically.¹³ These are also the most important programmes in terms of participants and/or expenditures (cf. Table 1). Hence, in the following we focus on these three programmes only.¹⁴

¹³ Further training activities with a planned duration longer than 3 months are not considered as the follow-up period would be too short. Note, however, that 1-Euro-jobs also might have longer durations than 3 months. But for them we only observe actual duration which is potentially endogenous, so that we do not want to group based on the realized durations.

¹⁴ Note that short and further training are also used for UI recipients. Evidence for UI recipients suggests that there are initial lock-in effects and at most minor positive longer-run employment effects of these programmes (see e.g. Wunsch and Lechner, 2008). 1-Euro-jobs are exclusively targeted at welfare recipients. Therefore, no pre-reform evidence for welfare recipients exists.

Table 2: Programme categories

Category	Description of the programmes	Sample sizes
Nonparticipants (NP)	No treatment from November 2006 to March 2007	8,091
Job placement by third parties	Job placement services by third parties	154
Job creation schemes	Job creation schemes	72
'integration grants'	Subsidized employment	103
Promotion of self-employment	Promotion of self-employment	68
1-Euro-jobs (OE)	Public workfare programmes	656
Short training (ST)	Check of the occupational aptitude and availability, job application training, job training, job trial, internship	479
Further training up to a planned duration of 3 months (FT)	training, skill formation, language course, job preparation, job orientation, completion of school leaving qualifications	394
Other programmes	Residual group of several small and very heterogeneous programmes	471
Total stock sample		10,675

Starting with this sample, we make three further adjustments. Firstly, since we measure conditioning variables and outcomes relative to programme start, which is only available for participants, we simulate hypothetical start dates for all nonparticipants. We (i) regress the time between sampling and programme start on individual characteristics¹⁵ in the pool of participants and (ii) use the coefficient estimates along with randomly drawn residuals to predict the nonparticipants' start dates.¹⁶ We drop all nonparticipants whose simulated start date is not between November 2006 and March 2007. Secondly, we ensure that only individuals that are in welfare at the sampling date and just prior to the programme start remain in the sample. Thirdly, all individuals not available to the labour market due to pregnancy, retirement, eased welfare receipt and (contemporaneous) long-term health problems and severe disability are deleted from the final evaluation sample that includes 5,210 nonparticipants and roughly 350 to 610 participants in the tree treatment categories we analyse, see Table 3.

¹⁵ Stratification characteristics, gender, education, marital status, variables reflecting the employment state history and benefit receipt, and regional variables are used as predictors.

¹⁶ This procedure has been suggested by Lechner (1999). The implemented version is analogous to Wunsch and Lechner (2008).

Table 3: Gross stock sample and evaluation sample

	NP	OE	ST	FT
Stock sample	8,091	656	479	394
Simulated programme start for nonparticipants is not within November 2006 and March 2007	1,466	-	-	-
Nonparticipants not receiving welfare or in (old) programme at the simulated start date	1,164	-	-	-
Not receiving welfare at sampling date (October 2006)	40	32	44	32
Not receiving welfare just prior to programme start	4	6	18	11
Reduced job search requirements: Pregnant, retired, 'eased' welfare receipt, long term health problems and severely disabled	207	7	2	4
Final evaluation sample	5,210	611	415	347

4. Descriptive statistics

Table 4 displays the composition of participants in 1-Euro-jobs, short training and further training, respectively, and of nonparticipants with respect to individual characteristics, regional attributes, and employment histories. This allows investigating the selectivity of various subgroups into a specific treatment. Women constitute 59 percent of the nonparticipants but account for less than half of the participants in any programme. Lone parents and individuals living with children younger than three are over-represented in the group of nonparticipants, too. It is also worth noting that the average age is considerably higher among nonparticipants and individuals in 1-Euro-jobs compared to other participants. Not surprisingly, the share of individuals aged 15 to 24 is higher in short and further trainings, whereas the converse is true for those aged 50 to 64 (not in table). Participants in 1-Euro-jobs are somewhat less frequently married and face more often health problems.

The share of individuals without school-leaving qualifications is relatively constant across groups, whereas the fraction of persons without vocational degree is somewhat lower among nonparticipants. German citizens are over-represented among nonparticipants and 1-Euro-jobs, individuals with a migration background (non-German nationality, foreign born or family language not German) are under-represented in these categories. 1-Euro-jobs are more extensively used in East Germany, whereas short and further trainings are less important. By

looking at the employment histories, one can see that on average, individuals in 1-Euro-jobs have spent fewer months in regular employment than other groups and more time in programmes and received welfare for more months (since 2005). However, a smaller share of them has never been employed since 1998. The fraction of unemployed since 1998 is lower in short and further trainings than in other groups.

Table 4: Selected descriptive statistics (shares in % in subsample)

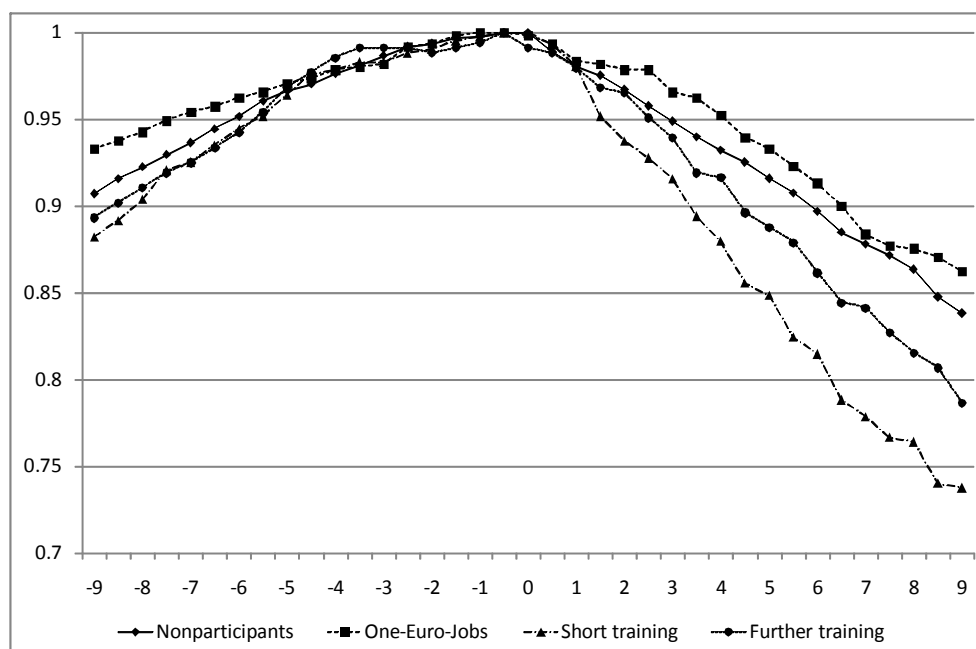
Subsample	Nonparticipation	One-Euro Jobs	Short training	Further training
Observations	5,210	611	415	347
Woman	59	46	47	49
Lone parent	22	15	15	15
Child below age 3 in household	24	11	15	17
Age in years	39	40	34	34
Married	38	31	35	35
Health limitations	15	17	13	13
No school degree	17	16	16	17
Lower secondary school degree	41	50	49	45
Upper secondary school degree	25	25	22	23
Polytechnical college or university entrance degree	9	8	8	9
No vocational degree	41	45	45	48
Completed apprenticeship training	44	50	45	38
Polytechnical college or university degree	4	3	2	5
German citizenship	85	90	80	78
Migration background*	30	23	34	33
East Germany	19	23	16	18
Months between sampling date and programme start	2,9	2,8	3,5	3,1
Months of welfare receipt since 2005	16,7	17,8	16,0	16,2
Months of minor employment since 2005	2,7	1,5	2,3	1,7
Months of regular employment since 2005	1,8	1,1	1,4	1,8
Months of unemployment since 2005	14,3	13,5	12,8	13,1
Months of programme participation since 2005	1,7	4,8	3,1	2,8
No employment since 1998	35	30	35	36
No programme participation since 1998	46	19	37	35
Fraction of time unemployed since 1998	31	31	26	27

Note: Entries are means and, if not stated otherwise, in percent. All variables are calculated from administrative records and are measured at the time when the sample was defined (October 2006). * Partly from survey data.

Thus, nonparticipants, participants in 1-Euro-jobs, and individuals in trainings differ considerably in observed characteristics. There are, however, only minor differences between the treated in short and further trainings. Participants in 1-Euro-jobs seem to have the worst

labour market relevant preconditions, as is indicated by their frequent welfare receipt, repeated programme participation, and fewer periods of regular employment.

Figure 1: Welfare receipt before and after programme start before matching



Note: Horizontal axis: months relative to programme start (month 0).

Figure 1 displays the proportion of welfare recipients across programmes and nonparticipation at different points in time relative to the (simulated) programme start. Month 0 indicates the programme start, -1 and 1 represent one month before and one month after start, respectively. The pre-start fractions give insights into programme selectivity. Post-start fractions are first indications of potential programme effects. They are however not yet corrected for any selectivity into the programmes. Nine months before programme start, the share of welfare recipients is higher than 85% in all categories and highest among participants in 1-Euro-jobs. Nonparticipants show the second highest fraction, followed by further and short training. In general, differences are not large, which points to only limited selectivity with respect to pre-programme welfare receipt. As expected, all individuals receive welfare at the start of the programme, as this is a precondition for being selected into the sample. After programme start, the fraction of welfare recipients declines least among participants in 1-Euro-jobs and 9

months later it is still higher than 85%. The share is somewhat smaller among nonparticipants and just below 80% for further training. Participants in short training improve their position relative to other groups with a fraction of welfare recipients of 75%. Note the lines of the different states show the same order before and after the programme which is indicative of some potential selection bias related to this variable. This issue will be taken up in the next section.

5. Econometrics

5.1 Programme effects of interest and identification

We are interested in identifying the average effects of participation in one of the three programmes versus nonparticipation for the respective population of participants. Let D denote the 'treatment', i.e. participation in some programme. Then, $D = 0$ denotes the case of nonparticipation and $D \in \{1, 2, 3\}$ denotes the case when an individual participated in one of the three programmes. Let Y^0, Y^1, Y^2, Y^3 denote potential labour market outcomes, e.g. the employment states that would have been realised for each treatment. For an early discussion of the potential outcome framework, see Rubin (1974). In reality, for each individual only the state related to the actually received treatment is observed. The observed outcome is thus

$$Y = \sum_{d \in \{0,1,2,3\}} Y^d \cdot \mathbf{1}\{D = d\}.$$

We want to learn about the mean effect

$$\theta^d = E[Y^d - Y^0 \mid D = d], d \in \{1, 2, 3\}.$$

θ^d denotes the expected effect for an individual randomly drawn from the population of participants in treatment d , i.e. the average treatment effect on the treated (ATET).

ATETs are generally not identified so that additional assumptions are needed for identification. A restriction often encountered in the programme evaluation literature is the

conditional independence assumption (CIA), which states that all factors that jointly affect selection into treatment and outcomes are observed in the data and, therefore, they should be controlled for. Under this condition, the potential outcomes are independent of the treatment conditional on the observed covariates. As this assumption is not testable, it needs to be plausibly justified. Such arguments become more convincing when the analysis is based on data that are rich in information with respect to the required covariates. The following expression formalizes the CIA on the relevant subspace χ of the covariate space:

$$Y^0, \dots, Y^3 \perp\!\!\!\perp D \mid X = x, \forall x \in \chi$$

where $\perp\!\!\!\perp$ denotes conditional independence and X is the vector of observed covariates (see Imbens, 2000, and Lechner, 2001, for the exact conditions and identification results in this multiple treatment framework). We obtain expressions for the mean potential outcomes conditional on covariates that are functions of the participation status, observable outcomes, and covariates only:

$$E[Y^d \mid D = d', X = x] = E[Y \mid D = d, X = x], \forall d, d' \in \{0, 1, 2, 3\}, \forall x \in \chi$$

Effect identification of the effects of the programmes compared to nonparticipation (ATET) also requires that there is common support in X among the respective treated and non-treated population:

$$p^d(x) \equiv \Pr(D = d \mid X = x, D \in \{0, d\}) < 1, \quad \forall x \in \chi, d \in \{1, 2, 3\},$$

where $p^d(x)$ denotes the so-called propensity score. Then, the ATET of participation in treatment $D = d$ versus nonparticipation is identified as

$$\begin{aligned} \theta^d &= \int E[Y \mid D = d, X = x] f_{X \mid D=d}(x \mid D = d) dx - \int E[Y \mid D = 0, X = x] f_{X \mid D=d}(x \mid D = d) dx \\ &= E(Y \mid D = d) - \int E[Y \mid D = 0, X = x] f_{X \mid D=d}(x \mid D = d) dx, \quad d \in \{1, 2, 3\}, \end{aligned}$$

where $f_{X \mid D=d}(x \mid D = d)$ denotes the conditional density of X given the respective treatment $D = d$. Instead of directly conditioning on X , Rosenbaum and Rubin (1983) for binary

treatments and Imbens (2000) and Lechner (2001) for multiple treatments have shown that identification is equivalently obtained by conditioning on a so-called balancing score, such as the one-dimensional propensity score $p^d(x)$ (this is useful to circumvent the curse of dimensionality related to a nonparametric regression of Y on a high-dimensional X).

5.2 Plausibility of conditional independence assumption in this setting

We now discuss the plausibility of the CIA in our research framework given our data. The selection process lies formally in the hands of the caseworkers who assign welfare recipients to activation programmes that are in principle compulsory, even though there is a limited possibility for bargaining between the caseworker and the welfare recipient. Jacobi and Kluve (2007) point out that recent welfare reforms have further increased the caseworkers' power over the selection process in order to improve the targeting of activation measures.

Post-reform programme allocation is related to a profiling process based on an interview in which the caseworker screens the welfare recipient's skills, deficiencies and labour market perspectives. The welfare recipient is subsequently classified according to her employment chances. This classification also influences the types of programmes she might potentially be assigned to. As noted by Jacobi and Kluve (2007), short training is targeted very broadly at those with reasonable employment prospects. Further training should be provided to those who benefit most from the newly provided skills and is mainly targeted at individuals with good labour market prospects. 1-Euro-jobs are targeted at those welfare recipients with otherwise very limited employment chances. They are frequently used in regions with particularly bad labour market conditions.

Given the wealth of individual and household information outlined in Section 3.1 our data are very well suited to capture the factors that determine individual employment prospects. In particular, we not only observe the standard characteristics like age, gender, marital status, household composition, nationality, migration, education and profession, but also health and

disability information. Moreover, we reconstruct the frequency, duration, and quality of employment, unemployment, benefit receipt, and programme participation of each individual from January 1998 to December 2007. What is lacking in our data are direct measures of individual motivation, attitudes and aptitude. It is, however, likely that these characteristics are relatively persistent over time such that they have impacted on the labour market success prior to the programme start. For this reason it is crucial that we are able to condition on individual employment histories in a detailed way. This is also emphasized by Card and Sullivan (1988) and Heckman, Ichimura, Smith and Todd (1998).

Furthermore, even though the profiling process is standardised, the organisational structure of the agencies might play a role in the judgment of which programme is considered to be most appropriate. We control for such differences by using agency level information about the form of case management, the number and qualification of caseworkers, and the number of welfare recipients per caseworker, among other factors.

Moreover, local labour market conditions are also crucial for employment prospects. Our data contain a large variety of measures of local labour market conditions including - among many others - unemployment, vacancies, GDP per worker, industry structure, migration, remoteness and distance from the next big city. Thus, we are confident that we capture all major factors that affect both selection into the programmes and our labour market outcomes of interest (see Section 6.1 for details on the latter).

5.3 *Estimation of the programme effects*

Having established identification of the effects, the question of the appropriate estimator arises. All possible parametric, semi- and nonparametric estimators are (implicitly or explicitly) built on the principle that for every comparison of two programmes and for every participant in one of those programmes we need a comparison observation from the other programme with the same characteristics regarding all factors that jointly influence selection

and outcomes.¹⁷ Here, we use propensity score matching estimators to produce such comparisons.¹⁸ An advantage of these estimators is that they are semi-parametric and that they allow arbitrary individual effect heterogeneity (see Heckman, LaLonde, and Smith, 1999; Imbens, 2004, provides an excellent survey of the recent advances in this field).

We use a matching procedure that incorporates the improvements suggested by Lechner, Miquel, and Wunsch (2006). These improvements aim at two issues: (i) To allow for higher precision when many 'good' comparison observations are available, they incorporate the idea of caliper or radius matching (e.g. Dehejia and Wahba, 2002) into the standard (nearest-neighbour) algorithm. (ii) Furthermore, matching quality is increased by exploiting the fact that appropriate weighted regressions that use the sampling weights from matching have the so-called double robustness property. This property implies that the estimator remains consistent if either the matching step is based on a correctly specified selection model, or the regression model is correctly specified (e.g. Rubin, 1979, Joffe, Ten, Have, Feldman, and Kimmel, 2004). Moreover, this procedure should reduce small sample bias as well as asymptotic bias of matching estimators (see Abadie and Imbens, 2006) and thus increase robustness of the estimator. The actual matching protocol is shown in Table B.1 in the appendix and contains more technical information about the estimator.

¹⁷ Of course, parametric models may construct such a group artificially outside the support of the data.

¹⁸ We estimate $p^d(x)$ by probit specifications. Among individual characteristics, gender, age, marital status, children younger than 3, nationality, and education appeared to be good predictors for selection into treatment. Individuals aged 50 to 64 are less likely to participate in any programme, and children under 3 decrease the probability of being assigned to further training. Furthermore, variables related to the employment history have considerable explanatory power. They include last occupation, duration of the last minor or regular employment, time in employment since 2005, time in programmes since 1998, average programme duration and number of programmes since 2005, time spent out of labour force since 1998, number of months in welfare receipt between sampling date and start date. Also regional variables characterize the treatment assignment. E.g. a large proportion of long term unemployed increases the likelihood to be assigned to 1-Euro-jobs. The exact specifications and results are available upon request.

As discussed in Section 3.2, our sample is not randomly drawn due to stratification. Since we are interested in ATETs and since participation is not random, we cannot simply use the sample weights η_j that account for stratification with respect to the total population of interest. Rather, we have to compute the probability of being part of a particular subpopulation conditional on treatment status. Using Bayes' law this can be done for each individual j using

$$\pi_j^0 = P(X = x_j | D = d) = \frac{P(D = d | X = x_j)P(X = x_j)}{P(D = d)} = \frac{p^d(x_j) \eta_j}{P(D = d)}.$$

When calculating the mean potential outcomes in each state d , the factor π_j has to be multiplied with the weight of the individual obtained by matching (1 for treated). Note that because stratification and attrition are independent of the participation status it suffices for the consistency of the first-step estimation of the propensity scores $p^d(x)$ to include all characteristics used to compute the sample weights η_j as explanatory variables, see Manski and Lerman (1977).

We use the fixed-weight standard error estimator proposed by Lechner, Miquel, and Wunsch (2006). It is the same as the one suggested by Lechner (2001) and applied in Gerfin and Lechner (2002) and Lechner (2002) except that heteroscedasticity is allowed for. See Lechner and Wunsch (2008) for the motivation and all details of this variance estimator that shows some resemblance to the estimator suggested by Abadie and Imbens (2006).

5.4 Simulating alternative allocations of welfare recipients to programmes

To answer the question whether programmes are targeted efficiently, we investigate the optimality of the allocation process. In contrast to the identification of ATETs, which is based on mean potential outcomes, the determination of the optimal allocation of welfare recipients into various programmes requires the knowledge of the potential outcomes for each individual in the sample. We therefore would have to know all potential outcomes Y^0, \dots, Y^3 for all

persons, even though only one out of four, i.e. the realised outcome Y , is observed. Our approach to predict the unobserved counterfactuals is similar to the one in Lechner and Smith (2007). Four aspects have to be taken into account. First, selection has to be controlled for, again by conditioning on the propensity score. Second, the potential outcomes have to be predicted as accurately as possible, including characteristics observed by the caseworkers suspected to influence their decision to allocate the welfare recipients. We therefore include vocational degree, regional characteristics, and variables reflecting the employment history as predictors. Third, due to the high dimensionality of the characteristics to be accounted for, nonparametric estimation of the potential outcomes is infeasible. Therefore, we use probit specifications for the potential outcome predictions, as the outcome variables are binary (see Section 6.1). Fourth, all characteristics used to compute the sample weights η_j have to be included in the probit specifications, too, for the estimation to be consistent for the stratified sample. To obtain representative average potential outcomes, the individual potential outcomes are multiplied with the respective sample weight.

Estimation of the coefficients required to predict Y^d is based on the subsample having the respective treatment status $D = d$. In each group, the binary outcome is estimated as a function of the propensity scores for all relevant comparisons, the variables used in the computation of weights, and characteristics observed by the caseworkers who decide upon programme allocation. The coefficient estimates are then used to predict the potential outcome Y^d for all individuals in the sample and this is done for all $D \in \{0, \dots, 3\}$. Based on the predicted potential outcomes, the results for different allocation rules regarding the assignment of welfare recipients into the programmes are simulated. However, it has to be remarked that the probit coefficients are estimated rather imprecisely due to small sample sizes in 1-Euro-jobs, short training and further training. This is not accounted for in the optimal allocation, which is determined by comparing the potential outcomes for each

individual and choosing the best one. In particular, we do not test whether differences in potential outcomes are statistically significant. In the interpretation of the results we therefore have to bear in mind that the potential outcomes are estimated with higher uncertainty for programme participants than for nonparticipants.

6 Results

6.1 Outcomes of interest and their measurement

From a policy perspective, the main interest is whether the programmes are able to reduce welfare dependency of their participants and whether they help them to find some form of employment. Moreover, since we focus on the first programme after the sampling date, it is interesting for the interpretation of the results to what extent individuals participate again after the first programme.

We measure outcomes in two ways. On the one hand, we use the administrative records to construct half-monthly outcome measurements starting with the first period after programme start. Focusing on the beginning rather than the end of the programme accounts for endogeneity of actual programme durations and rules out that programmes appear to be successful just because people are busy in the programme. We observe outcomes for all individuals in the sample up to 9 months after programme start. This period is relatively short but this is the price to pay when looking at very recent programmes. However, the half-monthly measurements allow looking at the short-run dynamics of the effects, thus potentially providing first indications of trends of the evolution of the effects in later periods. Moreover, they allow picking up potential lock-in effects of the programmes (cf. van Ours, 2004; Lechner, Miquel and Wunsch, 2006, 2007; Wunsch and Lechner, 2008). One drawback of using administrative records is that information on employment is missing after 2006.

The second set of outcomes is constructed from the second wave of the survey from self-reported employment status at the time of the second interview. Here, we are able to observe all outcomes of interest, but there are drawbacks as well. On the one hand, when individuals report to be in a programme we do not know whether this is the programme we evaluate or some other programme. Therefore, we do not report the survey results for this outcome measure. On the other hand, for each individual the second interview took place at different distances to programme start. Thus, when measuring outcomes we pick up a mixture of short (in particular of potential lock-in effects) and longer-run effects.

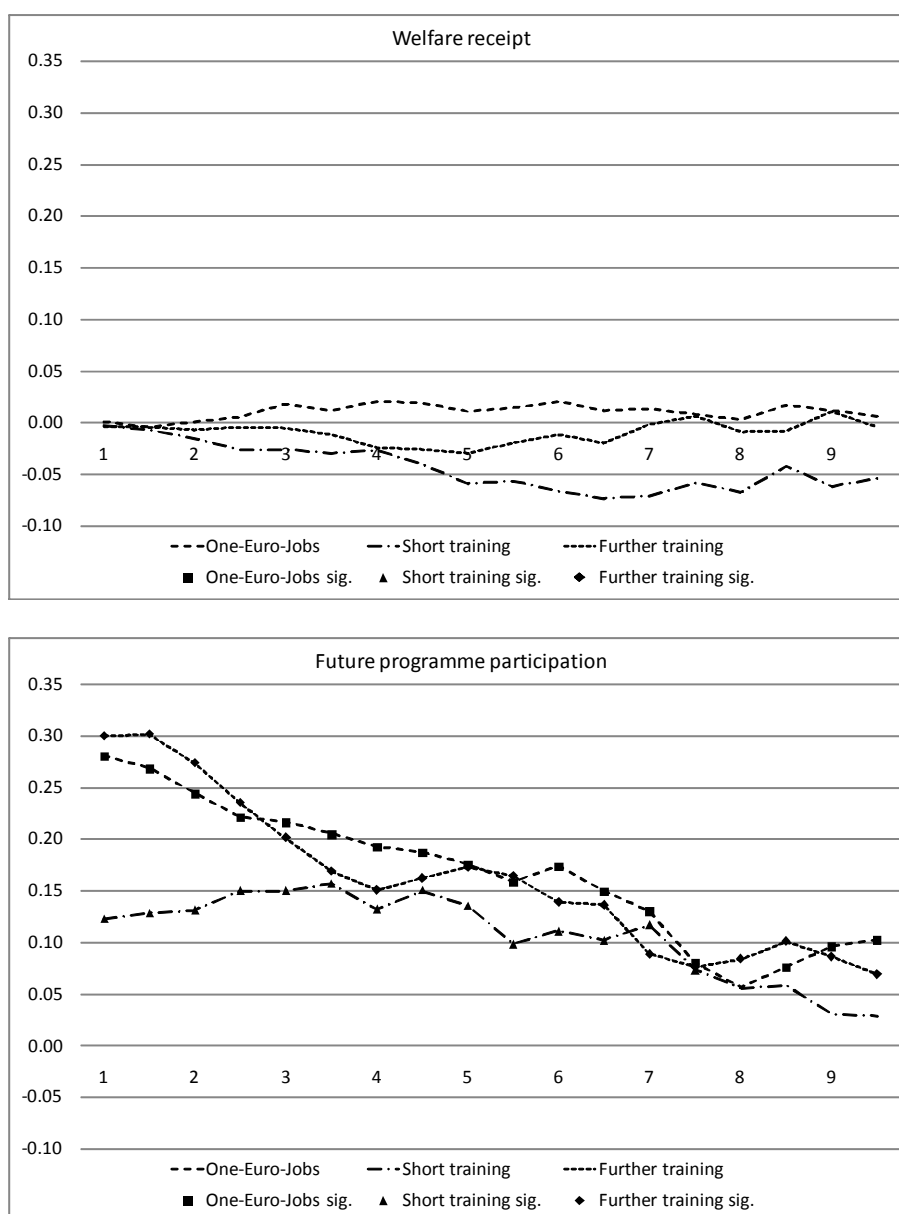
6.2 *The effects of the programmes*

Figure 2 shows the evolution of the effects of the programmes on welfare receipt (upper panel) and future programme participation (lower panel) for the comparison with nonparticipation for the first 9 months after programme start based on administrative records. It turns out that within this period none of the programmes significantly reduces welfare dependency. Only for short training the effect stabilizes at a reduction of about 5 %-points but the effect is still not significant. Programme participation seems to induce considerable future participation compared with nonparticipation (see lower panel of Figure 2).¹⁹

It is important to note, though, that sample sizes are too small to detect small effects of the programmes (standard errors vary between 0.06-0.07 for welfare receipt and between 0.03-0.04 for future programme participation). Thus, concluding from the results that the programmes are ineffective would not be appropriate.

¹⁹ Unfortunately, we cannot investigate whether there are positive long-run effects of participation in a sequence of programmes. Besides looking at a very short outcome window, our sample is too small to account for dynamic treatment effects as suggested by e.g. by Lechner (2009). However, it is not very likely that there are positive effects in the long-run because the estimated (insignificant) effects of programmes on welfare receipt are quite stable in the last three month of our observation period and do not indicate any future change.

Figure 2: Dynamics of the effects compared to nonparticipation (in %-points/100)



Note: Horizontal axis: months after programme start. Sig.: effect is significant on the 5% level (point-wise). Outcomes are calculated from administrative records. Standard errors vary between 0.06-0.07 for welfare receipt and between 0.03-0.04 for future programme participation.

The above results are confirmed with respect to welfare receipt when looking at the self-reported employment status at the time of the second interview in Table 5. With respect to employment, we find positive and significant average effects for participants in short training and it seems that the gain is in terms of 'self-sufficient' employment (individuals are employed and not welfare dependent). For the other programmes, there are some indications of positive employment effects as well but they are not significant.

Table 5: Effects of the programmes compared to nonparticipation in %-points/100

	One-Euro Job	Short training	Further training
Observations	611	415	347
Welfare receipt	0.030 (0.07)	-0.046 (0.06)	-0.034 (0.07)
Insured employment	0.056 (0.04)	0.091 (0.04) **	0.035 (0.04)
Minor employment	-0.037 (0.04)	-0.014 (0.03)	-0.005 (0.04)
Employed or self-employed	0.023 (0.05)	0.063 (0.05)	-0.012 (0.05)
Employed or self-employed, no welfare receipt	0.007 (0.03)	0.080 (0.03) **	0.041 (0.04)

Note: Standard errors in brackets. ***/**/* Effect is significant at the 1/5/10% level. Outcomes are calculated from self-reported employment status from the second wave of the survey.

6.3 Effect heterogeneity

In this section, we investigate whether there are some groups of participants that particularly benefit from the programmes. For example, we are interested in whether the programmes help those groups of welfare recipients that face particularly severe problems in reducing welfare dependency by returning to the first labour market. For this purpose, we estimate the programme effects in strata defined by gender, age, presence of small children in the household, lone-parent status, region, and migration background. The results are displayed in Table 6. Note however, that again the sample sizes are too small to draw definite conclusions from insignificant effects.

In contrast to the average effects, we find positive and weakly significant employment effects for participants in 1-Euro-jobs who are men, who are not lone parents and who do not have a migration background. However, these employments do not seem to be self-sufficient, i.e. pay enough to eliminate welfare benefit receipt. Moreover, the differences to the respective groups with opposite characteristics are small so that it cannot be concluded from the results that one group really benefits more than the other.

Table 6: Effects of the programmes compared to nonparticipation in %-points/100 for various subgroups

	Men	Women	Age 15-24	Age 25-49	Age 50-64	Child below age 3	No child below age 3
One-Euro Job versus nonparticipation							
Welfare receipt	0,040 (0.09)	0,010 (0.11)	0,056 (0.10)	0,098 (0.10)	-0,011 (0.18)	0,120 (0.16)	0,024 (0.07)
Insured employment	0,088 (0.05) *	0,064 (0.06)	-0,052 (0.07)	0,039 (0.07)	0,061 (0.06)	0,029 (0.12)	0,039 (0.04)
Minor employment	-0,027 (0.04)	-0,040 (0.06)	0,005 (0.04)	0,010 (0.06)	-0,003 (0.08)	-0,017 (0.08)	-0,031 (0.04)
Employed or self-employed	0,026 (0.07)	0,068 (0.08)	0,010 (0.08)	0,018 (0.09)	0,004 (0.11)	-0,055 (0.15)	0,030 (0.06)
Employed or self-employed without welfare receipt	0,014 (0.05)	0,029 (0.05)	-0,002 (0.07)	-0,077 (0.06)	0,046 (0.05)	-0,077 (0.06)	0,011 (0.04)
Observations	328	283	150	265	196	66	545
Short training versus nonparticipation							
Welfare receipt	-0,050 (0.08)	-0,035 (0.10)	-0,117 (0.10)	0,026 (0.09)	-0,187 (0.18)	-0,098 (0.16)	-0,008 (0.06)
Insured employment	0,089 (0.06)	0,106 (0.06)	-0,018 (0.06)	0,072 (0.07)	0,083 (0.06)	0,102 (0.11)	0,048 (0.05)
Minor employment	-0,020 (0.04)	-0,025 (0.05)	-0,020 (0.05)	-0,054 (0.05)	0,041 (0.10)	0,160 (0.08) **	-0,006 (0.03)
Employed or self-employed	0,075 (0.07)	0,038 (0.08)	-0,022 (0.08)	0,001 (0.08)	0,111 (0.12)	0,217 (0.12) *	0,038 (0.05)
Employed or self-employed without welfare receipt	0,096 (0.05) *	0,062 (0.05)	0,135 (0.05) ***	0,018 (0.06)	0,158 (0.07) **	0,175 (0.09) **	0,040 (0.04)
Observations	219	196	158	186	196	62	353
Further training versus nonparticipation							
Welfare receipt	-0.032 (0.09)	-0,023 (0.11)	-0.133 (0.11)	-0.008 (0.09)	0.062 (0.16)	-0.048 (0.18)	0.001 (0.07)
Insured employment	0.056 (0.06)	0,020 (0.07)	0.143 (0.06) **	0.010 (0.07)	0.009 (0.08)	0.156 (0.16)	0.001 (0.05)
Minor employment	0.023 (0.04)	-0,054 (0.06)	-0.002 (0.05)	0.031 (0.05)	0.000 (0.08)	0.072 (0.11)	-0.020 (0.04)
Employed or self-employed	0.016 (0.07)	-0.071 (0.09)	0.150 (0.07) **	-0.054 (0.08)	0.055 (0.12)	0.144 (0.18)	-0.057 (0.06)
Employed or self-employed without welfare receipt	0.059 (0.05)	0.022 (0.05)	0.158 (0.06) ***	-0.025 (0.05)	-0.015 (0.08)	0.099 (0.15)	0.015 (0.04)
Observations	328	283	150	265	196	66	545

Note: Standard errors in brackets. ***/**/* Effect is significant at the 1/5/10% level. Outcomes are calculated from self-reported employment status from the second wave of the survey.

- To be continued -

Table 6: Effects of the programmes compared to nonparticipation in %-points/100 for various subgroups (continued)

	Lone parent	No lone parent	East German	West German	Migration background	No migration background
One-Euro Job versus nonparticipation						
Welfare receipt	0,032 (0.15)	0,030 (0.07)	-0,038 (0.17)	0,038 (0.08)	-0,079	0,008 (0.08)
Insured employment	0,009 (0.07)	0,074 (0.04) *	0,118 (0.08)	0,057 (0.05)	0,117	0,071 (0.04) *
Minor employment	0,126 (0.09)	-0,032 (0.03)	-0,002 (0.06)	-0,042 (0.04)	-0,082	0,003 (0.03)
Employed or self-employed	0,146 (0.12)	0,040 (0.06)	0,139 (0.12)	0,029 (0.06)	0,001 (0.11)	0,072 (0.06)
Employed or self-employed without welfare receipt	0,001 (0.06)	0,012 (0.04)	0,038 (0.08)	0,005 (0.04)	0,071 (0.07)	0,015 (0.04)
Observations	92	519	138	472	140	471
Short training versus nonparticipation						
Welfare receipt	-0,008 (0.16)	-0,064 (0.07)	-0,012 (0.16)	-0,044 (0.07)	0,022	-0,113 (0.08)
Insured employment	0,084 (0.13)	0,117 (0.04) ***	0,185 (0.11)	0,067 (0.05)	0,030	0,113 (0.05) **
Minor employment	-0,109 (0.07)	-0,010 (0.03)	-0,116 (0.05) **	0,005 (0.04)	0,051	-0,081 (0.04) **
Employed or self-employed	0,008 (0.16)	0,101 (0.05) *	0,102 (0.13)	0,055 (0.06)	0,040 (0.09)	0,079 (0.06)
Employed or self-employed without welfare receipt	0,061 (0.10)	0,100 (0.03) ***	0,038 (0.08)	0,074 (0.04) **	0,060 (0.06)	0,121 (0.04) ***
Observations	64	351	67	348	142	273
Further training versus nonparticipation						
Welfare receipt	-0.101 (0.19)	-0.035 (0.07)	-0.019 (0.20)	-0.013 (0.07)	0.029	-0.024 (0.09)
Insured employment	0.043 (0.12)	0.060 (0.05)	0.150 (0.13)	0.005 (0.05)	-0.023	0.072 (0.05)
Minor employment	-0.014 (0.09)	-0.024 (0.04)	0.108 (0.11)	-0.011 (0.04)	0.049	0.004 (0.04)
Employed or self-employed	-0.097 (0.15)	0.008 (0.06)	0.174 (0.15)	-0.035 (0.06)	-0.008 (0.10)	0.023 (0.06)
Employed or self-employed without welfare receipt	0.036 (0.09)	0.052 (0.04)	0.002 (0.08)	0.033 (0.04)	-0.015 (0.08)	0.075 (0.04) *
Observations	92	519	138	472	140	471

Note: Standard errors in brackets. ***/**/* Effect is significant at the 1/5/10% level. Outcomes are calculated from self-reported employment status from the second wave of the survey.

The positive average effects of short training on self-sufficient employment seem to stem predominantly from participants who are either young or elderly, who have small children, or who have no migration background. For the latter as well as for East Germans it also seems that minor employments have been reduced in favour of regular insured employment. In contrast, the employment effect for participants with small children seems to stem from a substantial increase in minor employments.

For further training we now find evidence for positive employment effects for young participants and individuals without a migration background.

6.4 Optimal allocation of welfare recipients to programmes

Given that the programmes exhibit some effect heterogeneity with respect to participant characteristics it is interesting to investigate whether caseworkers send those welfare recipients to the programmes who benefit most from them. Focusing on the two most important outcome variables, i.e. welfare dependency and self-sufficient employment or self-employment, we compare the average outcomes of different allocations of welfare recipients to programmes using predictions of the respective variable as a function of characteristics for each individual in our evaluation sample. Table 7 presents the mean outcomes of the actual allocation and three alternatives for cost-neutral reallocations that keep the share of participants in each programme type constant.

The first interesting result is that the caseworker allocation and a random allocation yield very similar results for both outcomes of interest. However, caseworkers still do considerably better than in the worst-case scenarios, which would yield a 5 percentage points higher rate of welfare dependency or an about 4 percentage points lower employment rate. The overall scope for improvement by switching to the optimal allocation is for both outcomes about 9 percentage points which indicates a substantial inefficiency of the allocation process.

Table 7: Mean outcomes for different allocations

	Welfare receipt	Employment or self-employment without welfare receipt
Actual allocation	78.65	14.37
Random assignment	77.98	15.13
Outcome maximization	83.79	23.28
Outcome minimization	69.50	10.06
Difference between optimal and actual policy	-9.15	8.91

Note: Entries are in percent. Shaded cells indicate the optimal policy.

7 Conclusions

We use a unique data set that combines rich survey, administrative and regional data to provide early evidence on the effects of the three most important welfare-to-work programmes used in Germany since the last major welfare reform in 2005. This so-called Hartz IV reform constitutes the starting point for labour market activation of welfare recipients in Germany. Precisely, we look at short and further training as well as a workfare programme called 1-Euro-jobs that were conducted between October 2006 and March 2007, and consider short-run outcomes up to 12 months after programme start.

On average, we do not find significant effects of the three types of programmes on future welfare receipt. But all programmes induce further programme participation. Only short training, which is a combination of job-search assistance, work-tests and minor adjustment of skills, has on average a significant positive effect on self-sufficient employment. Moreover, all programmes exhibit considerable effect heterogeneity meaning that there are several subgroups of participants that do benefit from the programmes. We find positive and weakly significant employment effects for participants in 1-Euro-jobs who are men, who are not lone parents and who do not have a migration background. Short and further training is effective for young participants and non-migrants. In addition, short training also shows positive employment effects on the elderly and people with small children.

Given this effect heterogeneity we investigate whether caseworkers send those welfare recipients to the programmes who benefit most from them. We find that the observed

allocation is not optimal in terms of welfare receipt and employment. An optimal targeting of programmes that keeps the share of participants in each programme type and hence, programme costs constant would reduce welfare dependency by about 9 percentage points and would increase employment by a similar amount.

The results of this paper shed light on the short-term effects of the three quantitatively most important activation measures used since the Hartz IV legislation. However, sample sizes are currently too small to draw definite conclusions about the short-run effectiveness of the programmes. Further research is also required to evaluate long-term effects of a broader range of programmes and activation measures. This will eventually allow judging on the overall effectiveness of an important component of the recent welfare reforms in Germany.

Literature

- Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, 74, 235-267.
- Bernhard, S., K. Hohmeyer, E. Jozwiak, S. Koch, T. Kruppe, G. Stephan, and J. Wolff (2008): "Aktuelle Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung. Aktive Arbeitsmarktpolitik in Deutschland und ihre Wirkungen", IAB Forschungsbericht 2/2008.
- Beaudry, R. (2002): "Workfare and Welfare: Britain's New Deal", working paper, Department of Economics, University of York.
- Blank, R. M. (2002): "Evaluating Welfare Reform in the United States", *Journal of Economic Literature*, 60, 1105-1166.
- Caliendo, M., R. Hujer, and S. Thomsen (2005): "Identifying Effect Heterogeneity to Improve the Efficiency of Job Creation Schemes in Germany?", *Applied Economics*, 40 (9), 1101-1122.
- Card, D., and D. Sullivan (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment", *Econometrica*, 56, 497-530.
- Dehejia, R. H., and S. Wahba (2002): "Propensity-Score-Matching Methods for Nonexperimental Causal Studies", *Review of Economics and Statistics*, 84, 151-161.

- Dostal, J. M. (2008): "The Workfare Illusion: Re-examining the Concept and the British Case", *Social Policy & Administration*, 42, 19-42.
- Finn, D. (2000) : "Welfare to Work: the local dimension", *Journal of European Social Policy*, 10, 43-57.
- Gerfin, M., and M. Lechner (2002): "Microeconomic Evaluation of the Active Labour Market Policy in Switzerland", *Economic Journal*, 112, 854-893.
- Grogger, J. (2003): "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families", *The Review of Economics and Statistics*, 85, 394-408.
- Halvorsen, R., and P. H. Jensen (2004): "Activation in Scandinavian welfare policy", *European Societies*, 6, 461-483.
- Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66(5), 1017-1098.
- Heckman, J. J., R. LaLonde, and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Vol. 3, 1865-2097, Amsterdam: North-Holland.
- Hohmeyer, K., and J. Wolff (2007): "A fistful of Euros. Does One-Euro-Job participation lead means-tested benefit recipients into regular jobs and out of unemployment benefit II receipt", IZA Discussion Paper 32/2007.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *The Review of Economics and Statistics*, 86, 4-29.
- Jacobi, L., and J. Kluve (2007): "Before and After the Hartz Reforms. The Performance of Active Labour Market Policy in Germany," RWI Discussion Paper 41.
- Joffe, M. M., T. R. Have, H. I. Feldman, and S. Kimmel (2004): "Model Selection, Confounder Control, and Marginal Structural Models", *The American Statistician*, 58-4, 272-279.
- Knijin, T. (2001): "Careful or lenient: welfare reform for lone mothers in the Netherlands", *Journal of European Social Policy*, 11, 235-251.
- Kildal, N. (2001): "Workfare Tendencies in Scandinavian Welfare Policies", *SES paper*, 5, Geneva: International Labour Office.

- Kluve, J. (2006): "The Effectiveness of European Active Labour Market Policy", IZA Discussion Paper 2018.
- Kluve, J., and C. Schmidt (2002): "Can training and employment subsidies combat European unemployment?", *Economic Policy*, 35, 411-448.
- Konle-Seidl, R., W. Eichhorst, and M. Grienberger-Zingerle (2007): "Activation policies in Germany : from status protection to basic income support," IAB Discussion Paper 200706.
- Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany after Unification", *Journal of Business Economics and Statistics*, 17, 74-90.
- Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.
- Lechner, M. (2002): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", *Review of Economics and Statistics*, 84, 205-220.
- Lechner, M. (2009): "Sequential Causal Models for the Evaluation of Labor Market Programs", *Journal of Business & Economic Statistics*, 27(1), 71-83.
- Lechner, M., R. Miquel, and C. Wunsch (2006): "Long-Run Effects of Public Sector Sponsored Training in West Germany", revised version of the ZEW Discussion Paper No. 05-02.
- Lechner, M., R. Miquel, and C. Wunsch (2007): "The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification", *German Economic Review*, 8(4), 468-507.
- Lechner, M., and J. Smith (2007): "What is the value added by caseworkers?", *Labour Economics*, 14(2), 135-151.
- Lechner, M., and C. Wunsch (2006): "Are Training Programs More Effective When Unemployment is High?", revised version of Discussion Paper 200623, Department of Economics, University of St. Gallen.
- Manski, C.F., and S. R. Lerman (1977): "The Estimation of Choice Probabilities from Choice Based Samples", *Econometrica*, 45, 1977-1988.
- Martin, J. P., and D. Grubb (2001): "What Works and for Whom: A Review of OECD Countries' experiences with active labour market policies", *Swedish Economic Policy Review*, 8, 9-56.
- Moffitt, R. (2002): "Welfare Programs and Labor Supply", in: A. Auerbach and M. Feldstein (eds.): *Handbook of Public Economics*, 4(4), Chapter 34, 2393-2430.

- Ochel, W. (2005): "Hartz IV - Welfare to Work in Germany," CESifo DICE Report 2/2005.
- Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, 41-50.
- Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomised and Nonrandomised Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association*, 74, 318-328.
- Schneider, J. (2008): "The Effect of Unemployment Benefit II Sanctions on Reservation Wages", IAB Discussion Paper 200819.
- Torfig, J. (1999): "Workfare with welfare: recent reforms of the Danish welfare state", *Journal of European Social Policy*, 9, 5-28.
- Van Ours, J. (2004): "The Locking-in Effect of Subsidized Jobs", *Journal of Comparative Economics*, 32, 37-52.
- Wolff, J., and E. Jozwiak (2007): "Does short-term training activate means-tested unemployment benefit recipients in Germany", IAB Discussion Paper 29/2007.
- Wolff, J., and A. Nivorozhkin (2008): "Start me up: The effectiveness of a self-employment programme for needy unemployed people in Germany," IAB Discussion Paper 200820.
- Wunsch, C. (2008): "Labour Market Policy in Germany: Institutions, Instruments and Reforms since Unification", revised version of Discussion Paper 2005-06, Department of Economics, University of St.Gallen.
- Wunsch, C., and M. Lechner (2008): "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes", *Kyklos*, 61, 134-174.

Appendix A: Further details on the data

Table A.1: Descriptive statistics for the main individual characteristics (means)

	Non-participation	One-Euro-Job	Short training	Further training
age	38.66	39.87	33.48	34.24
woman	0.59	0.46	0.47	0.49
German	0.85	0.90	0.80	0.78
migration background	0.30	0.23	0.34	0.33
child below age 3 in household	0.24	0.11	0.15	0.17
married	0.38	0.31	0.35	0.35
single	0.33	0.47	0.41	0.45
cohabiting	0.08	0.07	0.07	0.05
lone parent	0.22	0.15	0.15	0.15
no school-leaving qualifications	0.17	0.16	0.16	0.17
secondary schooling (Hauptschule)	0.41	0.50	0.49	0.45
secondary schooling (Realschule)	0.25	0.25	0.22	0.23
high school graduate	0.09	0.08	0.08	0.09
no professional degree	0.42	0.45	0.45	0.48
vocational education	0.40	0.45	0.42	0.34
technical school	0.04	0.05	0.03	0.04
college or university	0.04	0.03	0.02	0.05
health limitations	0.15	0.17	0.13	0.13
health limitations impact on job placement	0.10	0.12	0.09	0.08
returning to the labour market	0.04	0.04	0.04	0.03
Region				
Berlin	0.03	0.02	0.02	0.04
Eastern Germany	0.19	0.23	0.16	0.18
Mecklenburg-Western Pomerania	0.03	0.03	0.01	0.02
Brandenburg	0.04	0.05	0.04	0.03
Saxony-Anhalt	0.05	0.05	0.06	0.08
Saxony	0.05	0.06	0.04	0.02
Thuringia	0.03	0.03	0.01	0.02
Schleswig-Holstein and Hamburg	0.03	0.04	0.03	0.03
Lower Saxony and Bremen	0.12	0.13	0.15	0.21
North Rhine-Westphalia	0.20	0.15	0.18	0.12
Hesse	0.07	0.02	0.06	0.05
Rhineland-Palatinate and Saarland	0.08	0.14	0.07	0.09
Baden-Wuerttemberg	0.11	0.12	0.11	0.10
Bavaria	0.16	0.15	0.21	0.18
Desired occupation				
agriculture, forestry, horticulture, mining	0.02	0.04	0.04	0.03
production and processing	0.03	0.04	0.05	0.02
technical occupation, engineering	0.06	0.08	0.09	0.08
construction	0.05	0.09	0.07	0.07
unskilled worker	0.04	0.04	0.04	0.03
services	0.25	0.26	0.25	0.24
office management and administration	0.10	0.12	0.10	0.13
artist	0.01	0.01	0.01	0.01
health sector	0.03	0.02	0.02	0.01

teaching	0.04	0.05	0.03	0.03
other occupations	0.17	0.14	0.13	0.13
Role in household				
head of household	0.76	0.82	0.74	0.76
partner	0.16	0.13	0.13	0.15
underage, unmarried child	0.05	0.01	0.07	0.04
unmarried person of full age and under 25	0.03	0.04	0.07	0.05
Additional sources of household income				
earned income	0.14	0.15	0.19	0.14
earned income and other sources of income	0.28	0.24	0.30	0.29
no income	0.33	0.38	0.30	0.34
other sources of income	0.24	0.23	0.22	0.22
Welfare payments in EUR				
baseline benefits	246.42	282.48	272.31	260.63
housing costs	160.29	164.23	153.72	159.08
further costs for special needs	23.51	14.19	14.05	16.85
Source of income before welfare receipt				
unemployment benefits	0.25	0.29	0.29	0.25
unemployment assistance	0.21	0.27	0.18	0.16
social assistance	0.23	0.21	0.22	0.21
earned income	0.19	0.13	0.20	0.23
mini-job (minor employment)	0.11	0.06	0.09	0.10
self-employment	0.05	0.02	0.05	0.03
support for professional training and education	0.03	0.04	0.03	0.04
lived on savings	0.16	0.14	0.20	0.16
lived on partner's income	0.14	0.09	0.14	0.13
lived with parents	0.15	0.14	0.24	0.18
other sources of income	0.11	0.10	0.11	0.10
Reason for applying for welfare				
finished education/ professional training	0.04	0.07	0.07	0.06
gave up self-employment	0.03	0.01	0.03	0.02
unemployment benefit entitlements expired	0.23	0.28	0.28	0.24
became unemployed without entitlement to unemployment benefits	0.10	0.10	0.14	0.15
unemployment benefits were insufficient to make a living	0.05	0.05	0.07	0.04
became incapable of working, disabled	0.05	0.05	0.06	0.06
familial and/or housing situation changed	0.14	0.10	0.12	0.12
other household members became unemployed	0.07	0.04	0.08	0.06
income of other household members decreased	0.06	0.04	0.07	0.07
parents applied for welfare	0.03	0.01	0.04	0.04
other reasons	0.05	0.04	0.04	0.06
savings were eaten up	0.10	0.09	0.11	0.12
moved out and founded an own household	0.03	0.03	0.05	0.05
Employment history since 1998				
months employed	14.42	14.99	13.71	12.84
months unemployed	28.22	29.09	22.63	22.30
months in programme	5.40	9.94	6.90	6.48
months out of labour force	40.11	36.63	38.84	40.00
months since last employment	23.93	28.56	22.75	19.07
months since last unemployment	1.77	2.31	1.87	1.38
months since last programme	12.95	7.52	7.87	9.10
months since last out of labour force spell	23.76	28.78	22.00	22.58

number of employment spells	1.30	1.48	1.22	1.33
number of unemployment spells	1.87	2.41	1.87	2.01
number of programmes	1.05	1.74	1.30	1.33
num. of out of labour force spells	2.32	2.61	2.30	2.49
mean employment duration	8.45	8.36	8.88	7.61
mean unemployment duration	17.27	14.33	13.19	12.52
mean programme duration	2.90	4.91	3.50	3.42
mean out of labour force duration	22.96	19.21	22.13	22.54
share of employment	0.15	0.15	0.15	0.14
share of unemployment	0.31	0.31	0.26	0.27
share in programme	0.06	0.11	0.08	0.07
share out of labour force	0.44	0.40	0.46	0.48
duration of last employment	7.26	7.32	7.36	6.49
duration of last unemployment	15.48	11.46	10.66	10.20
duration of last programme	2.81	3.95	2.74	3.00
duration of last out of labour force spell	18.34	15.56	15.95	13.55
Characteristics of last employment				
last monthly earnings in EUR	613.78	683.66	591.90	629.18
employee, clerk	0.11	0.09	0.08	0.10
skilled worker, master craftsman, foreman	0.08	0.09	0.06	0.07
worker	0.17	0.23	0.22	0.22
apprentice	0.06	0.06	0.10	0.07
part time employment	0.23	0.23	0.19	0.16
employed in production and processing industry	0.28	0.30	0.28	0.29
employed in service industry	0.28	0.29	0.29	0.28
employed in other industries	0.05	0.08	0.07	0.05
occupation: agri-/ horticulture, forestry, mining	0.02	0.05	0.03	0.03
occupation: unskilled worker	0.05	0.07	0.04	0.07
occupation: technical occupation, engineering	0.01	0.01	0.01	0.01
occupation: office management, admin., teaching	0.18	0.15	0.16	0.17
occupation: logistics	0.07	0.07	0.10	0.08
occupation: services	0.16	0.13	0.16	0.10
occupation: construction	0.04	0.06	0.04	0.06
occupation : metal working	0.03	0.03	0.03	0.02
occupation : other production and processing	0.07	0.10	0.08	0.08

Note: All variables are measured at the sampling date. If not stated otherwise, entries are fractions. In addition to the variables in the table, a rich set of regional variables as well as variables that further disaggregate the information contained in the employment histories have been used in the estimation.

Appendix B: Technical details of the matching estimator used

Table B.1: A matching protocol for the estimation of a counterfactual outcome and the effects

Step 1	Specify a reference distribution defined by X .
Step 2	Pool the observations forming the reference distribution and the participants in the respective period. Code an indicator variable W , which is 1 if the observation belongs to the reference distribution. All indices, 0 or 1, used below relate to the actual or potential values of W .
Step 3	Specify and estimate a binary probit for $p(x) := P(W = 1 X = x)$
Step 4	Restrict sample to common support: Delete all observations with probabilities larger than the smallest maximum and smaller than the largest minimum of all subsamples defined by W .
Step 4	<p><i>Estimate the respective (counterfactual) expectations of the outcome variables.</i></p> <p>Standard propensity score matching step (multiple treatments)</p> <p>a-1) Choose one observation in the subsample defined by $W=1$ and delete it from that pool.</p> <p>b-1) Find an observation in the subsample defined by $W=0$ that is as close as possible to the one chosen in step a-1) in terms of $p(x), \tilde{x}$. 'Closeness' is based on the Mahalanobis distance. Do not remove that observation, so that it can be used again.</p> <p>c-1) Repeat a-1) and b-1) until no observation with $W=1$ is left.</p> <p>Exploit thick support of X to increase efficiency (radius matching step)</p> <p>d-1) Compute the maximum distance (d) obtained for any comparison between a member of the reference distribution and matched comparison observations.</p> <p>a-2) Repeat a-1).</p> <p>b-2) Repeat b-1). If possible, find other observations in the subsample of $W=0$ that are at least as close as $R \cdot d$ to the one chosen in step a-2) (to gain efficiency). Do not remove these observations, so that they can be used again. Compute weights for all chosen comparisons observations that are proportional to their distance. Normalise the weights such that they add to one.</p> <p>c-2) Repeat a-2) and b-2) until no participant in $W=1$ is left.</p> <p>d-2) For any potential comparison observation, add the weights obtained in a-2) and b-2).</p> <p>Exploit double robustness properties to adjust small mismatches by regression</p> <p>e) Using the weights $w(x_i)$ obtained in d-2), run a weighted linear regression of the outcome variable on the variables used to define the distance (and an intercept).</p> <p>f-1) Predict the potential outcome $y^0(x_i)$ of every observation using the coefficients of this regression: $\hat{y}^0(x_i)$.</p> <p>f-2) Estimate the bias of the matching estimator for $E(Y^0 W = 1)$ as: $\sum_{i=1}^N \frac{1(W=1)\hat{y}^0(x_i)}{N^1} - \frac{1(W=0)w_i\hat{y}^0(x_i)}{N^0}$.</p> <p>g) Using the weights obtained by weighted matching in d-2), compute a weighted mean of the outcome variables in $W=0$. Subtract the bias from this estimate to get $\widehat{E(Y^0 W = 1)}$.</p>
Step 5	Repeat Steps 2 to 4 with the nonparticipants playing the role of participants before. This gives the desired estimate of the counterfactual nonparticipation outcome.
Step 6	The difference of the potential outcomes is the desired estimate of the effect with respect to the reference distribution specified in Step 1.

Note: \tilde{x} includes gender, elapsed unemployment duration until programme start, and whether a person is employed in month 12 or month 24 before programme start. In some specifications, we also match on education. In the specification where programme composition is held constant, we also match on the type of programme and planned programme duration. \tilde{x} is included to ensure a high match quality with respect to these critical variables.

The parameter used to define the radius for the distance-weighted radius matching (R) is set to 90%. This value refers to the distance of the worst match in a one-to-one matching and is defined in terms of the propensity score. Different values for R are checked in the sensitivity analysis in Lechner, Miquel, and Wunsch (2006). The results were robust as long as R did not become 'too large'.