DO GERMAN WELFARE-TO-WORK PROGRAMMES REDUCE WELFARE DEPENDENCY AND INCREASE EMPLOYMENT?

Martin Huber[+], Michael Lechner[+], Conny Wunsch[+], and Thomas Walter*

[+] University of St. Gallen, Swiss Institute for Empirical Economic Research (SEW)
* Centre for European Economic Research (ZEW)

During the last decade many Western economies reformed their welfare system with the aim of activating welfare recipients by increasing welfare-to-work programmes and job search enforcement. We evaluate the short term effects of three important German welfare-to-work programmes implemented after a major reform in January 2005 ("Hartz IV"), namely short training, further training with a planned duration of up to three months, and public workfare programmes ("One-Euro-Jobs"). Our analysis is based on a combination of large scale survey and administrative data that is rich with respect to individual, household, agency level, and regional information. We use this richness of the data to base the econometric evaluation on a selection-on-observables approach. We find that short-term training programmes on average increase their participants' employment perspectives. There is also 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.

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

JEL classification: J68

Addresses for correspondence

Martin Huber, Michael Lechner, Conny Wunsch

University of St. Gallen, Swiss Institute for Empirical Economic Research (SEW)
Varnbüelstr. 14, CH-9000 St. Gallen, Switzerland
martin.huber@unisg.ch, michael.lechner@unisg.ch, conny.wunsch@unisg.ch, www.sew.unisg.ch/lechner

Thomas Walter
Centre for European Economic Research (ZEW), L 7, 1, DE-68161 Mannheim, Germany
walter@zew.de, www.zew.de

1   Introduction[*]

Over the last decade many OECD countries faced increasing numbers of welfare claimants and reacted by conducting welfare reforms (e.g. US, Canada, UK, and Germany). Most resulted in a shift from passive benefit payment towards increased job search and work requirements imposed on welfare recipients with the objective of encouraging employment uptake and reducing welfare dependency. Welfare recipients are obliged to participate in welfare-to-work programmes, and they can be sanctioned by means of benefit cuts in case of non-compliance.

Welfare research has traditionally focused on North America where welfare-to-work efforts were considerably increased by the US states and the federal government over the 1990's. They also played a key role in the Canadian Self-Sufficiency Project (SSP). In the course of the reforms, an extensive literature evaluating the various welfare programmes and reforms has evolved: See for example Blank (2002), Moffitt (2002) and Grogger and Karoly (2005) for a review of the US welfare reforms and the related empirical literature, and Bitler, Gelbach and Hoynes (2008) for the SSP.

In Europe, where unemployment insurance (UI) is usually more generous and relative numbers of UI claimants are generally higher than in North America, the literature has almost exclusively focused on the evaluation of programmes targeted at UI rather than welfare recipients.[1] However, the results are not easily extendable to welfare recipients, because due to UI eligibility rules welfare recipients differ systematically from UI claimants with respect to their labour market characteristics: They either do not have enough work experience to claim UI or they exhausted their UI claim because of long-term unemployment. These differences may be

[1]   See for example the surveys by Martin and Grubb (2001), Kluve and Schmidt (2002), Kluve (2006), and Wunsch (2006).

particularly relevant as the programmes are shown to exhibit considerable effect heterogeneity with respect to the characteristics of their participants such as gender, education, and local labour market conditions (for Germany see for instance Caliendo, Hujer, and Thomsen, 2005, Lechner, Miquel and Wunsch, 2007, 2010, and Wunsch and Lechner, 2008).

In Germany, a coherent system of welfare benefits and in particular welfare-to-work programmes (WTWP) has been introduced only at the beginning of 2005 with the so-called Hartz IV reform.[2] It constitutes a remarkable change in German welfare policy and has drawn considerable public attention. In this paper, we provide early evidence on the short-run effects of the three most important of the newly introduced welfare-to-work programmes: (i) short training that includes basic job search assistance, work tests and minor adjustment of general skills, (ii) short further training that aims at improving occupation-related skills, and (iii) a workfare programme called *One-Euro-Jobs* that aims at improving the employability of welfare recipients with relatively bad employment prospects.

So far, there exist only two studies that analyse the effects of some of these programmes: Wolff and Jozwiak (2007) investigate the effect of participation of welfare recipients in short-term training, and Hohmeyer and Wolff (2007) evaluate the effectiveness of One-Euro-Jobs.[3] Based on administrative data they analyse programmes starting directly after the reform in early 2005. Among other issues that hamper the evaluation of the future effects of programmes using data from its introduction, this period was characterised by strong data collection problems, which may have affected their results, for instance due to large amounts of missing data raising concerns about representativeness and bias resulting from mismeasured variables.

---

[2] Many other European countries have recently conducted welfare reforms. Surveys on welfare reforms in Europe 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 (2006), Jacobi and Kluve (2007) and Konle-Seidl et al. (2007) for Germany.

[3] The following recent papers investigate 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.

Here, we consider more recent programmes that started between October 2006 and March 2007, when these problems do not occur. We use more informative data than the earlier studies and evaluate short further training as well. Furthermore, we investigate effect heterogeneity in a detailed way and analyse a variety of outcome variables. Finally, we assess the optimality of the allocation process of welfare recipients to the different programmes to investigate whether there is scope for improvement in employment and welfare dependency rates.

Our analysis is based on a combination of rich survey, administrative, and regional data that allows estimation of the programme effects using matching techniques. We use the adjusted calliper propensity score matching estimator proposed by Lechner, Miquel and Wunsch (2010). For up to 17 months after the start of the programme we find no significant effects of the programmes on average welfare dependency rates. With respect to employment, we find positive and significant effects for some programmes and some groups of participants, in particular for short training and for non-migrant welfare recipients. Our assessment of the optimality of programme assignment shows that there is considerable scope for improvement in employment and welfare dependency rates.

The remainder of the paper is organized as follows: Section 2 provides some background on the relevant institutions in Germany. In Section 3 we introduce the data, followed by a discussion of the sample definition, the programmes and participation patterns. Section 4 presents descriptive statistics of the evaluation sample. Identification and estimation of the effects of interest as well as the simulation of alternative treatment allocations 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 welfare policy

In Germany, a coherent system of welfare benefits and welfare-to-work programmes has been introduced only at the beginning of 2005 with the so-called Hartz IV reform. Before that, there existed two parallel systems. On the one hand, unemployed individuals who had exhausted their UI claim were eligible for means-tested unemployment assistance which replaced up to 57% of their previous net earnings. They were administered by the local employment agencies of the federal public employment service (PES) and had access to all labour market programmes available to UI recipients. On the other hand, needy individuals who were never eligible for UI payments received a means-tested lump-sum social assistance payment whose amount depended on household composition and income. They were administered by the municipalities, and welfare-to-work programmes were basically non-existent for this group of people.

The Hartz IV reform removed this asymmetry for needy individuals who do not receive UI payments.[4] Unemployment and social assistance have been combined to one single means-tested welfare payment that is independent of previous earnings. Instead, its level depends on household size, composition and income similar to the former social assistance (so-called unemployment benefits II, UB II).[5] Eligibility for UB II depends on being physically and mentally capable of working for at least 15 hours per week, active job search and willingness to participate in welfare-to-work programmes. Non-compliance with these rules, or the rejection of acceptable job offers, can be sanctioned by means of temporary benefit cuts. The new welfare payments and welfare-to-work programmes are in most cases administered by joint ven-

---

[4] Jacobi and Kluve (2007) provide an excellent survey of the reform package.
[5] UB II amounted to 351 € for a single-person household in January 2009. This is less generous than unemployment assistance (on average 550 € in 2003 in West Germany) but more generous than social assistance (about 300 €, see Ochel (2005). On top of UBII, welfare payments also include rents and housing costs (on average ca. 180 € per person, see http://www.pub.arbeitsagentur.de/hst/services/statistik) and compulsory social insurance contributions. Further costs for special needs might be covered as well.

tures between the local employment office of the PES and the municipality, thus also the asymmetry in administration has been removed. However, in 69 out of 429 offices the agencies are run by the municipality alone, entirely outside of the responsibility and competency of the PES. In this study, we only consider the regular joint ventures for which data availability is much better.

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%).[6] The number of UB II recipients amounted to about 4.5 million in January 2005. About half of these people were unemployed, the rest being the 'working poor'. The number of claimants increased steadily during 2005 and reached a peak of 5.5 million in April 2006. Since then it has declined to just below 5 million in August 2008.[7]

2.2 *German welfare-to-work programmes*

The Hartz IV reform constitutes a remarkable change in German welfare policy. For the first time all welfare recipients are a target group of labour market activation. A well-defined set of welfare-to-work programmes (WTWP) has been introduced for this group of people with the primary objective to (re)integrate welfare claimants into the labour market as quickly as possible and to reduce welfare dependency. Before 2005, no consistent reintegration efforts were made. Thus, there is neither experience nor any evidence on the efficiency of welfare-to-work programmes prior to the reform in 2005.[8] In this paper, we evaluate the effectiveness of the three most important German WTWP.

---

[6] Figures according to the Federal Statistical Office (www.destatis.de).

[7] Figures according to the monthly and annual reports of the FEA (www.arbeitsagentur.de). The increase in claimants in 2005 was partly due to a considerable number of new applicants that did never apply for benefits before the reform because there was some stigma associated with applying for social assistance.

[8] As mentioned above, only unemployment assistance claimants had access to labour market programmes. These were the same as the ones for UI claimants. The existing evaluation studies did not distinguish between the two groups. See e.g. Wunsch (2006).

According to Table 1, which reports entries into German WTWP for the period 2005-2007 and the corresponding expenditures, so-called One-Euro-Jobs are by far most frequently assigned, accounting for a third of overall expenditures, followed by short training courses and further training. *One-Euro-Jobs* are public-sector-related workfare programmes that were specifically introduced for unemployed welfare recipients in 2005. According to the legislator, these programmes should be of public interest and additional in the sense that the assigned work would otherwise not be accomplished by existing companies.[9] The work load typically consists of 20-30 hours per week over a period of 3-12 months. Participants do not receive a (subsidized) wage, but merely a compensation for job-related extra costs which amounts to 1-2.5 € per hour and which is paid in addition to UB II. *One-Euro-Jobs* aim at restoring or improving the employability of their participants rather than direct integration in the labour market. Accordingly, participants differ from participants in other WTWP in terms of worse labour market histories, in particular longer and more frequent welfare dependency and, correspondingly, shorter and less stable employment experience.

Table 1: Entries in and expenditures for selected activation programmes

Category	Entries			Expenditures in million €		
	2005	2006*	2007*	2005	2006	2007
One-Euro-Jobs and other public employment programmes	633,938 (12.7%)	815,380 (15.1%)	798,774 (15.1%)	1,105 (35.3%)	1,381 (36.0%)	1,322 (31.3%)
Short training	410,884 (8.2%)	480,675 (8.9%)	545,960 (10.3%)	158 (5.0%)	164 (4.3%)	163 (3.9%)
Further training	69,906 (1.4%)	124,169 (2.3%)	167,200 (3.2%)	196 (6.3%)	378 (9.8%)	504 (11.9%)
Other programmes	592,682 (11.9%)	849,912 (15.8%)	974,233 (18.5%)	1,666 (53.3%)	1,918 (49.9%)	2,233 (52.9%)
Total	1,707,410	2,270,136	2,486,167	3,125	3,841	4,221

Note: If not stated otherwise, figures are for joint ventures alone. * Includes both joint ventures and agencies controlled by the municipality. In brackets: Ratio of programme participants to the average annual stock of employable welfare recipients (left panel) and ratio to total expenditures (right panel), in percent. Source: Statistics of the Federal Employment Agency at http://www.pub.arbeitsamt.de/hst/services/statistik/detail/e.html, own calculations.

---

[9] 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 e.g. Dostal (2008).

Short training courses are comparably cheap, albeit quantitatively important programmes with durations of usually a few days to two weeks, but in any case no more than 12 weeks. Their content is rather heterogeneous. 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 sample work days as well as job application and job interview trainings. Secondly, they aim at minor adjustments of general job relevant skills. The most common examples are language courses and computer classes.

Further training comprises a more substantial human capital investment and focuses on the adaption of occupation-specific skills to recent labour market developments, e.g. to mitigate skill mismatch due to structural change, rather than general job and search-related skills. The courses either take place as class-room training, potentially in combination with short internships in firms, or in so-called practice firms, that simulate a job in a specific profession. Planned durations vary from a few months to up to three years.

The main components of the residual category 'other programmes' in Table 1 are wage subsidies (paid to firms which employ difficult-to-place workers during the first months of employment), start-up grants (bridging allowances for taking up self-employment), and job placement services of private companies.

3   Data and definition of sample and participation status

3.1   Data

Our analysis is based on a combination of very informative survey, administrative and regional data. The core of these data is a survey of welfare recipients who have been interviewed in two waves at the beginning (January to April) and around the end of 2007 (November to March 2008). It provides information on gender, age, marital status, education, nationality, migration background, employment status, welfare receipt, participation in WTWP, past

performance on the labour market, job search behaviour, and household information such as household composition and employment status of each household member.

We use survey data for a stratified stock sample of 21,000 welfare recipients in October 2006.[10] 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, yielding 11,276 panel cases.[11] To make up for these losses, a refreshment sample of 5,744 cases was drawn from the same population as the original sample. The 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. The data contain sample weights for each individual that take into account both stratification and attrition.

The survey data have been merged with administrative data on welfare recipients provided by Germany's Federal Employment Agency 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 administrative database comprises individual characteristics (education, age, gender, marital status, number of children, profession, nationality, disabilities and health), the type and amount of benefits received, compliance with benefit rules, programme participation (type, duration) and up to 10 years of employment histories (type of employment, industry, occupational status, earnings).

The combined administrative and survey data were linked to further data at the agency and regional level. They include a wide range of indicators reflecting labour market conditions (e.g. share of unemployed, long term unemployed, welfare recipients and migrants, GDP per worker, population density, industry structure) and variables that characterize the agencies'

---

[10] Stratification is based on age (15-24/25-49/50-64), children aged <3 in the household, and lone parent status.

[11] Attrition is not random with respect to various characteristics including gender, education and employment status. See the final project report for a detailed attrition analysis and how this affected the construction of the sample weights. (http://www.bmas.de/portal/30144/property=pdf/f387__forschungsbericht.pdf).

organisational structure (e.g. case management and counselling concept, number and qualification of caseworkers, welfare recipients per caseworker, placement strategy).

3.2 Sample and treatment definition

We restrict our analysis to welfare recipients in the panel and refreshment sample who are administered by joint ventures between the local employment office of the PES and the municipality because this is the most common case and because there are severe problems with the quality of the administrative data for the other type of administration. Moreover, we discard persons who did not agree to merge their survey data with their administrative data. The first issue reduces the sample by about one third to 11,260 observations, the second one by another 585 cases to 10,675 observations.

Since we use a stock sample of welfare recipients in October 2006, we have to restrict the analysis to programmes that start after the sampling date, as the sampling procedure conditions on the outcome (failure) of programmes attended before this date. Furthermore, as the survey data are only available up to the second interview, and administrative records end in December 2007, we restrict attention to the first programme that starts after the sampling date and before April 2007 to have a follow-up period for measuring outcomes which is not too short.

We define program participation on the basis of the administrative data only for programme starts before January 2007 as differences in the survey design between the first wave of the panel and the refreshment sample imply that programme starts can be consistently identified for all individuals in the survey only from January 2007 onwards. Nonparticipants are defined as those individuals not receiving any treatment between the sampling date in October 2006 and March 2007. This treatment definition leads to a sample of 8,091 nonparticipants, 656 participants in *One-Euro-Jobs* (mean duration 7 months), 479 participants in *Short Training* (mean duration 1 month), and 394 participants in *Further Training* with planned durations of

no more than 3 months (mean duration 2 months). The latter restriction is imposed because of our short follow-up period. Consequently, *Short* and *Further Training* have similar durations in our analysis but differ with respect to their contents (improvement of job search or general skills versus occupation-specific skills). We do not impose a similar restriction on the duration of *One-Euro-Jobs* because we only observe actual duration which is potentially endogenous as welfare recipients are obliged to continue job search and accept job offers even while in a programme.

Table 2: Gross stock sample and final evaluation sample

	Nonpar-ticipation	1-Euro Job	Short training	Further training
Stock sample	8,091	656	479	394
Simulated programme start for nonparticipants is not between sampling date 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 & severely disabled	207	7	2	4
Final evaluation sample	5,210	611	415	347

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 characteristics[12] in the pool of participants and (ii) use the coefficient estimates along with randomly drawn residuals to predict the nonparticipants' starting dates.[13] We drop all nonparticipants from the analysis whose simulated programme start date is outside the treatment window. Secondly, we ensure that only individuals receiving welfare at the sampling date and just prior to the programme start re-

---

[12] Variables related to the stratification, gender, education, marital status, variables reflecting the employment state history and benefit receipt, and regional variables are used as predictors.

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

main in the sample to ensure programme eligibility. Thirdly, all individuals not available to the labour market in the period before programme start due to pregnancy, retirement, eased welfare receipt and (contemporaneous) long-term health problems or severe disability are disregarded in the analysis.[14] Table 2 shows how the sample shrinks accordingly to 5,210 non-participants and about 350-600 participants in each group.

4. Descriptive statistics

Table 4 shows the mean characteristics of the four treatment groups in order to investigate selectivity in programme participation.

Women constitute 59% of the nonparticipants but account for less than half of the programme participants. Lone parents and individuals with children younger than three are over-represented among nonparticipants, too. Moreover, with lower shares of unskilled persons nonparticipants are better educated than participants. In other respects nonparticipants and participants in *One-Euro-Jobs* are similar but differ from the other participants: Average age is considerably higher and German citizens are over-represented while individuals with a migration background are under-represented. With respect to regional differences, *One-Euro-Jobs* are more extensively used in East Germany than *Short* and *Further Training*. Participants in One-Euro-Jobs seem to have the worst labour market histories, as is indicated by their frequent welfare receipt, repeated programme participation, fewer periods of employment and a large share of unemployment. Participants in *Short* and *Further Training* overall seem to be relatively similar with the exception of a somewhat larger share of unskilled individuals in *Further Training*.

---

[14] We use both administrative and survey data to do this. Although in the survey part of the information is retrospective (in particular for the refreshment sample) the date when a particular state began and ended is also available, which ensures that we do not condition on outcome variables at this stage.

Table 3: Selected descriptive statistics (shares in % in subsample)

Subsample	Non-participation	One-Euro Job	Short Training	Further Training
Observations	5,210	611	415	347
Individual characteristics				
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
German citizenship	85	90	80	78
Migrant or child of migrant(s)*	30	23	34	33
East Germany	19	23	16	18
Education				
No vocational degree	41	45	45	48
Completed apprenticeship training	44	50	45	38
Polytechnical college or university degree	4	3	2	5
Labour market history				
Months of welfare receipt since 2005	16.7	17.8	16.0	16.2
Months of minor employ. up to 400 € 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. Further descriptive statistics are available from the authors upon request.

Figure 1 displays the proportion of individuals in the sample who receive welfare 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. By construction everyone receives welfare directly before the programme. Moreover, the similarity in dependency rates among participants and nonparticipants in the 6 months before programme start results from the stock sampling of welfare recipients in October 2006 in combination with the treatment window and the procedure to simulate start dates for nonparticipants. However, the rates before this indicate selectivity in programme participation showing that participants in *One-Euro-Jobs* seem to be the worst risks in terms of welfare dependency followed by nonparticipants and participants in the training courses.

Figure 1: Welfare receipt before and after programme start before matching
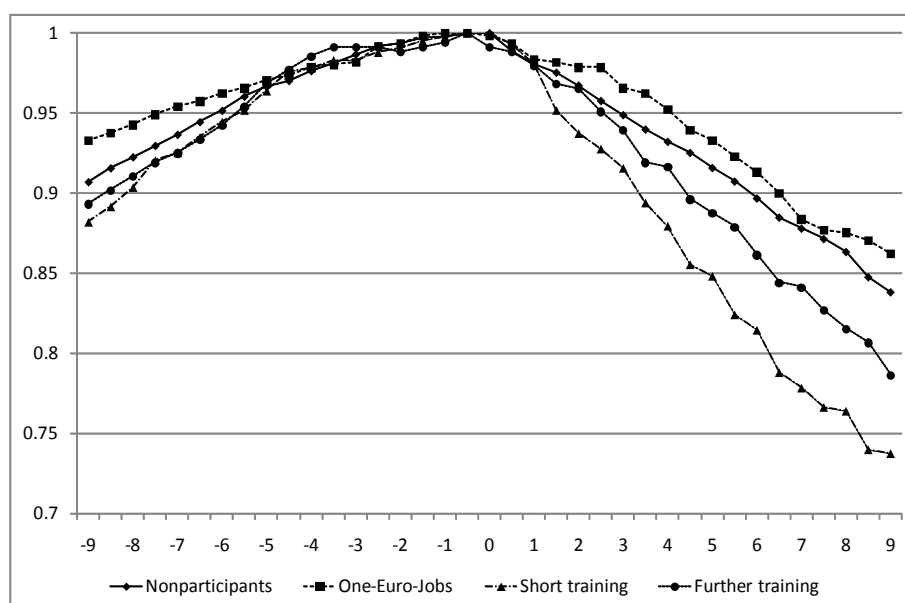
Note: Horizontal axis: months relative to programme start (month 0).

The dependency rates after programme start are not corrected for potential selectivity. Still, they can provide first hints at possible programme effects. The ordering of the four groups is preserved but the difference relative to nonparticipants widens and becomes larger than in the pre-treatment period for participants in both types of training, in particular for *Further Training*. For *One-Euro-Jobs* it remains relatively constant.

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, i.e. the average treatment effect on the treated (ATET). Ideally, we would like to know the potential outcomes for all states and for each individual (see Rubin, 1974, for an early discussion of the potential outcome framework). However, only the outcome under the treatment that was actually received is observed. Therefore, ATETs are generally not identified without additional assumptions.

Our identification strategy is based on the assumption that we observe all factors that jointly affect selection into the programmes and outcomes (conditional independence assumption: CIA). If it holds, then the potential outcomes are independent of the treatment conditional on these observed factors. Instead of directly conditioning on the covariates, Rosenbaum and Rubin (1983) have shown that identification is equivalently obtained by conditioning on a so-called balancing score, such as the conditional probability to participate in a programme given the covariates (propensity score). This is useful to circumvent the curse of dimensionality related to a nonparametric regression using a high-dimensional covariate vector.

5.2 Plausibility of the conditional independence assumption in this study

As the CIA is not testable, we need to establish its plausibility from an analysis of the selection process. In our context the selection process lies formally in the hands of the caseworkers. They assign welfare recipients to WTWP 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 their clients 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. One-Euro-Jobs are targeted at 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 the 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.

5.3   *Estimation of the programme effects*

We use propensity score matching to estimate the programme effects. An advantage of these estimators is that they are semi-parametric and that they allow arbitrary individual effect het-

erogeneity (see Heckman, LaLonde, and Smith, 1999; Imbens, 2004, provides an excellent survey of the recent advances in this field).

We apply a matching procedure that incorporates the improvements suggested by Lechner, Miquel, and Wunsch (2010). Compared to nearest-neighbour matching this procedure is more precise because it incorporates the idea of calliper or radius matching (e.g. Dehejia and Wahba, 2002) into the algorithm. Furthermore, matching quality is improved by exploiting the fact that appropriate weighted regressions that use the sampling weights from matching have the so-called double robustness property: 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 can be obtained from the authors upon request.[15]

As discussed in Section 3.2, our sample is not randomly drawn from the population. Since we are interested in ATETs and since participation is not random, we cannot simply use the sample weights that account for stratification and attrition. Rather, we have to compute the probability of being part of a particular subpopulation conditional on treatment status. When calculating the mean potential outcomes in each treatment state, this probability has to be multiplied with the weight of the individual obtained by matching (1 for treated). We exploit that if stratification and attrition are independent of the participation status it suffices for the consis-

---

[15] We estimate the propensity scores 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 the 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 the 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 One-Euro-Jobs). The exact specifications and results are available upon request.

tency of the first-step estimation of the propensity scores to include all characteristics used to compute the sample weights as explanatory variables, see Manski and Lerman (1977).

We use the fixed-weight standard error estimator proposed by Lechner, Miquel, and Wunsch (2010). 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 (2009) 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 of all treatments (incl. nonparticipation) for each individual in the sample.

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 scores.[16] 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. Fourth, all characteristics used to compute the sample weights have to be included in the pro-

---

[16]   The estimation of the propensity scores is based on the multiple treatment framework by Imbens (2000) and Lechner (2001) using probit models for the pair-wise comparisons of all programmes against each other as well as nonparticipation.

bit specifications, too, for the estimation to be consistent for a sample subject to stratification and attrition. To obtain representative average potential outcomes, the individual potential outcomes are multiplied with the respective sample weight.

Estimation of the coefficients required to predict the potential outcomes is based on the sub-sample in the respective treatment group. 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 outcomes for all individuals in the sample and this is done for all treatments. Based on the predicted potential outcomes, the results for different allocation rules regarding the assignment of welfare recipients into the programmes are simulated.

One cautionary note though: The probit coefficients are estimated rather imprecisely due to small sample sizes in *One-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   The effects of the programmes

From a policy perspective, the main objectives of welfare-to-work programmes are reducing welfare dependency and increasing employment rates. We measure the corresponding outcome variables of interest using both administrative and survey information.

6.1.2   Outcomes constructed from administrative data

From the administrative data we construct half-monthly measurements of welfare dependency 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 as welfare recipients are obliged to continue job search and accept job offers even while in a 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 cost of looking at very recent programmes. However, the half-monthly measurements allow analysing the short-run dynamics of the effects, thus potentially providing first indications of trends 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).

Figure 2: Dynamics of the effects compared to nonparticipation (in %-points/100)
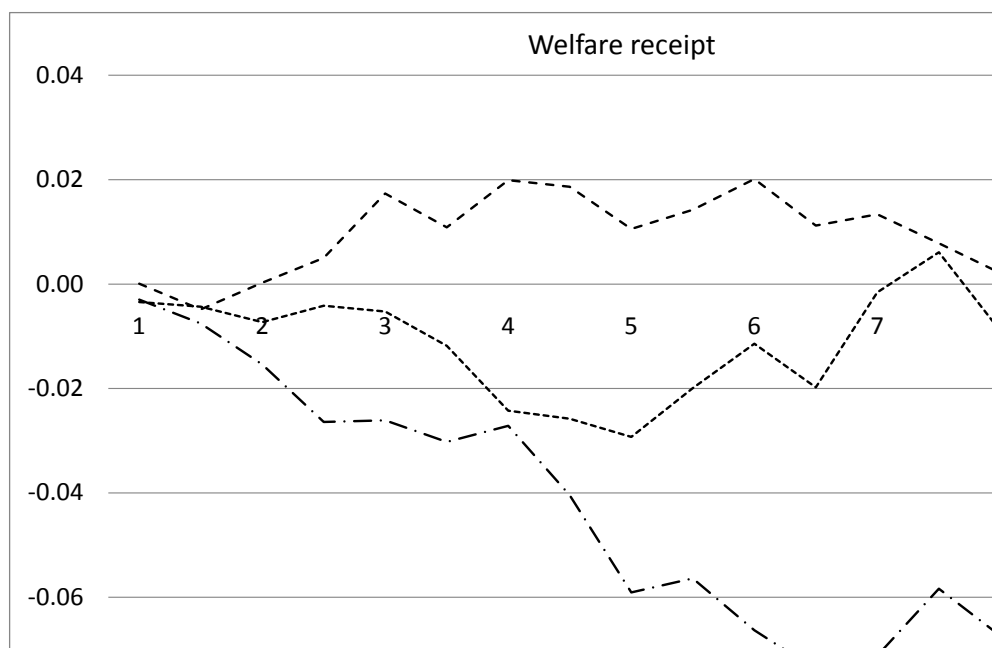
Note: Horizontal axis: months after programme start. Sig.: effect is significant on the 5% level (point-wise). Outcomes are calculated from administrative records from one to nine months after programme start. Standard errors vary between 0.06-0.07.

Figure 2 shows the evolution of the effects of the programmes on welfare dependency compared to nonparticipation for the first 9 months after programme start. 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 percentage points but the effect is still not significant.[17] We have to bear in mind that sample sizes are too small to detect significant effects of the programmes (standard errors vary between 0.06-0.07). Thus, concluding from the results that the programmes are ineffective would not be appropriate. Unfortunately, the administrative information on employment is missing after 2006 so that we cannot consider this outcome based on administrative data.

6.1.2 Outcomes constructed from survey data

The second set of outcomes is constructed from the second wave of the survey, more specifically from the self-reported employment status at the time of the second interview. Even though we observe all outcomes of interest, there are also drawbacks related to the survey data. For each individual the second interview differs with respect to the month it took place and with respect to its distance to the programme start. This distance varies between 7 and 17 months. Thus, when measuring outcomes based on the survey data we pick up a mixture of short (in particular of potential lock-in effects) and longer run effects.

Table 4: Effects of the programmes compared to nonparticipation in %-points/100

	1-Euro-Job			Short Training				Further Training		
Observations	611			415				347		
	Y0	Y1	Y1-Y0	Y0	Y1	Y1-Y0		Y0	Y1	Y1-Y0
Welfare receipt	0.79	0.82	0.030	0.76	0.72	-0.046		0.79	0.75	-0.034
			(0.07)			(0.06)				(0.07)
Insured employment	0.17	0.22	0.056	0.22	0.31	0.091	**	0.22	0.25	0.035
			(0.04)			(0.04)				(0.04)
Minor employment	0.16	0.13	-0.037	0.13	0.12	-0.014		0.12	0.12	-0.005
			(0.04)			(0.03)				(0.04)
Employed or self-employed	0.37	0.39	0.023	0.39	0.45	0.063		0.37	0.36	-0.012
			(0.05)			(0.05)				(0.05)
Employed, self-employ., no welfare receipt	0.13	0.13	0.007	0.14	0.22	0.080	**	0.13	0.18	0.041
			(0.03)			(0.03)				(0.04)

Note: Y0: Mean estimated counterfactual. Y1: Mean outcome of participants. Standard errors in brackets.
***/**/* Effect is significant at the 1/5/10% level. Outcomes are calculated from the self-reported employment status from the second wave of the survey. Insured employment: Regular employment that is subject to social insurance. Minor employment: Monthly salary/wage amounts to no more than 400 €.

---

[17] 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 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 months of our observation period and do not indicate any future change.

The results in Figure 2 concerning welfare dependency are confirmed when looking at the self-reported employment status at the time of the second interview in Table 4: The effects indicate a small (given the levels) reduction in welfare dependency of 3-5 percentage points for the training courses but they are not significant. With respect to employment, we find positive and significant average effects for participants in short training. With 9 %-points and a counterfactual mean employment rate of 22% this effect is large and it seems that the gain is in terms of 'self-sufficient' employment (individuals who are employed but not welfare dependent). For the other programmes, especially for further training, there are some indications of positive employment effects as well but they are not significant.

Potential reasons for the ineffectiveness of *One-Euro-Jobs* in the short run are negative lock-in effects. In our sample *One-Euro-Jobs* have a mean duration of 7 months with several individuals participating 9 to 12 months. Thus, the second interview takes place when a non-negligible fraction of participants is still in the programme, or very shortly thereafter. While accomplishing a lengthy *One-Euro-Job*, individuals are likely to reduce their job search effort relative to nonparticipants who have more time to seek employment. This argument is in line with Graversen (2003), who estimates relatively large lock-in effects for public welfare-to-work programmes in Denmark. Moreover, the objective of *One-Euro-Jobs* is to restore or improve employability rather than direct reemployment so that also for this reason the follow-up period may be too short to detect employment effects.

6.2   *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. For this purpose, we estimate 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 5. Note however, that the samples 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 male participants in *One-Euro-Jobs* 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 dependency. 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.

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 that pay no more than 400 €per month 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.

Table 5: 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 back-ground
			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.

6.3 *Optimal allocation of welfare recipients to programmes*

Given that the programmes exhibit effect heterogeneity with respect to participant characteristics we investigate whether caseworkers send those welfare recipients to the programmes who benefit most. Focusing on the two most important outcome variables, i.e. welfare dependency and self-sufficient (self-) employment, we compare the average outcomes of different allocations of welfare recipients to programmes using predictions of the respective outcome variable as a function of characteristics for each individual in our evaluation sample.

Table 6: Mean outcomes for different allocations

	Welfare receipt	Employment or self-employment. w/o 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.

Table 6 presents the mean outcomes of the actual allocation and three alternatives for cost-neutral reallocations that keep the share of participants in each programme group 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 point higher rate of welfare dependency or an about 4 percentage point 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.

7   Conclusions

We use a unique data set that combines rich survey, administrative and regional data to provide early evidence on the short-run effects of the three most important welfare-to-work pro-

grammes used in Germany since the last major welfare reform in 2005 which constitutes the starting point for labour market activation of welfare recipients in Germany. On the one hand, we investigate short and further training with a planned duration of up to 3 months that aim at increasing employment rates and reducing welfare dependency. On the other hand, we analyse workfare programmes called One-Euro-Jobs that aim at improving the employability of welfare recipients with particularly bad employment prospects with the hope of reducing welfare dependency in the longer run. We look at programmes that were conducted between October 2006 and March 2007, and consider short-run outcomes up to 17 months after programme start.

On average, we do not find significant effects of all three programmes on future welfare dependency within our follow-up period. Only short training, which is a combination of job-search assistance, work-tests and minor improvements of general skills, has a significantly positive effect on self-sufficient employment on average. Thus, one may be tempted to conclude that at least in the short run the newly introduced German welfare-to-work programmes do not reduce welfare dependency and increase employment only in the case of short training. However, there is also some indication that the training courses may reduce welfare dependency and that further training also has some positive employment effects. Our sample sizes are too small to draw definite conclusions based on the estimated effects being insignificant. Moreover, our evaluation window is too short to draw any conclusions with respect to mid and long term effects, especially for the workfare programme.

The findings are in line with the existing literature on similar programmes in other countries, though. Re-evaluating California's Greater Avenues for Independence (GAIN) program, Hotz, Imbens, and Klerman (2000) find that a focus on job search assistance and a "work first" approach is more effective with respect to employment than basic skills training one to three years after the respective programme, even though the superiority decreases in later periods.

For the New Deal for Young People in the UK, Dorsett (2006) finds that two years after programme start job placements into the voluntary sector (which resemble public workfare programmes) and full-time education/training for the individuals without basic qualification appear to be less successful than subsidized employment in the private sector. Graversen (2003) finds for Denmark that placement into the public sector yields a smaller employment effect and entails longer welfare dependency than placement into the private sector.

The average effects hide some considerable effect heterogeneity, though. There are several subgroups of participants that do benefit from the programmes. We find positive and weakly significant employment effects for participants in One-Euro-Jobs who are men, who are not lone parents and who are not migrants. 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 would benefit most from participating in the respective programmes. 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 welfare-to-work programmes used since the Hartz IV legislation. However, sample sizes are 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 activation measures. This will eventually allow judging 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.

Bitler, M., J. Gelbach and H. Hoynes (2008): "Distributional Impacts of the Self-Sufficiency Project", *Journal of Public Economics*, 92(3-4), 748-765.

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.

Dorsett, R. (2006): "The new deal for young people: effect on the labour market status of young men", *Labour Economics,* 13, 405-422.

Dostal, J. M. (2008): "The Workfare Illusion: Re-examining the Concept and the British Case", *Social Policy and 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): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland", *Economic Journal*, 112, 854-893.

Graversen, B. K. (2003): "The Impact of Active Labour Market Programmes on Welfare Dependence in Denmark", mimeo, Danish National Institute of Social Research, Aarhus.

Grogger, J., and L. Karoly (2005): "Welfare Reform: Effects of a Decade of Change", *Harvard University Press*.

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.

Hotz, V. J., G. W. Imbens, and J. Klerman (2000): "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program", NBER working paper 8007.

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", *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", *American Statistician*, 58-4, 272-279.

Knijn, 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 and Economic 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 and Economic Statistics*, 27(1), 71-83.

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., R. Miquel, and C. Wunsch (2010): "Long-Run Effects of Public Sector Sponsored Training in West Germany", *Journal of the European Economic Association*, forthcoming.

Lechner, M., and J. Smith (2007): "What is the value added by caseworkers?", *Labour Economics*, 14(2), 135-151.

Lechner, M., and C. Wunsch (2009): "Are Training Programs More Effective When Unemployment is High?", *Journal of Labor Economics*, forthcoming.

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. (2004): "Welfare-To-Work Experiences With Specific Work-First Programmes In Selected Countries", CESifo Working Paper No. 1153.

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 19/2008.

Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s", *Review of Economics and Statistics,* 86, 133-155.

Torfing, 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 20/2008.

Wunsch, C. (2006): "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.