OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 70, 3 (2008) 0305-9049
doi: 10.1111/j.1468-0084.2008.00505.x

Turning Unemployment into Self-Employment: Effectiveness of Two Start-Up Programmes*

Hans J. Baumgartner† and Marco Caliendo‡,§

†*DIW Berlin, Berlin, Germany*
‡*Institute for the Study of Labor (IZA), Bonn, Germany (e-mail: caliendo@iza.org)*
§*IAB, Nuremberg, Germany*

Abstract

Turning unemployment into self-employment has become a major focus of German active labour market policy (ALMP) in recent years. If effective, this would not only reduce Germany's persistently high unemployment rate, but also increase its notoriously low self-employment rate. Empirical evidence on the effectiveness of such programmes is scarce. We evaluate the effectiveness of two start-up programmes for the unemployed, where we include the probability of being employed, the probability of being unemployed and personal income as outcome variables. Our results show that at the end of the observation period, both programmes are effective. The considerable positive effects present a stark contrast to findings from evaluations of other German ALMP programmes in recent years. Hence, ALMP programmes aimed at moving the unemployed into self-employment may prove to be among the most effective, both in Germany and elsewhere.

I. Introduction

Turning unemployment into self-employment has become a major focus of German active labour market policy (ALMP) in recent years. Whereas the Federal Employment Agency (FEA) funded only 37,000 business start-ups by formerly unemployed individuals in 1994, the number was already above 350,000 in 2004 (approximately

*The authors thank Arne Uhlendorff, Viktor Steiner, one anonymous referee and the editors for valuable comments and Steffen Künn for excellent research assistance. Helpful comments were also received from seminar and conference participants in Berlin (BeNa), Stockholm (RATIO Colloquium), Budapest (ESEM), Munich (Verein für Socialpolitik), Nuremberg (CAPE) and Bonn (SPEAC/IZA). Hans J. Baumgartner left DIW Berlin before this paper was accepted for publication in the *Bulletin*.

JEL Classification numbers: J68, C14, H43, M13.

250,000 in West Germany). This increase was driven, among other things, by a new programme known as the 'start-up subsidy' (SUS, *Existenzgründungszuschuss*), which was introduced in 2003 as part of the 'Hartz reforms'.[1] Unemployed individuals can now choose between this and a second programme, the 'bridging allowance' (BA, *Überbrückungsgeld*), which was already implemented in the late 1980s. The two programmes differ in their design, most importantly regarding the amount and duration of the subsidy. Whereas the BA pays recipients the same amount that they would have received in unemployment benefits for a period of 6 months (plus a lump sum to cover social security contributions), the SUS runs for 3 years, paying a lump sum of €600/month for the first year, €360/month for the second and €240/month for the third. If successful, these programmes could potentially not only decrease Germany's persistently high unemployment rate, but increase its notoriously low self-employment rate as well. Looking at the FEA's spending on ALMP, we clearly see the increasing priority assigned to these programmes within the overall ALMP strategy. Whereas in 1994 only 0.6% of ALMP resources were allocated to these measures, in 2004 this number was 17.2%. This corresponds to an annual spending of over €2.7 billion.

For all the aforementioned reasons, the high research interest in evaluating these programmes is unsurprising. However, empirical evidence on start-up aid is very rare, not only in Germany but also internationally. Meager (1996) summarizes findings for five countries (Denmark, France, West Germany, UK and US) and concludes that the evidence presented does not allow a conclusive assessment of the overall effectiveness of such schemes. Existing papers usually focus either on survival rates of subsidised businesses (e.g. Cueto and Mato, 2006), or compare start-ups by formerly unemployed people with start-ups which were not created out of unemployment (see, e.g. Pfeiffer and Reize, 2000). The present paper takes a different approach. Instead of comparing business start-ups by formerly unemployed individuals with other start-ups, we compare the labour market outcomes of the formerly unemployed entrepreneurs with other unemployed individuals. This approach is driven by the consideration that start-up subsidies form one component of ALMP, and their effectiveness should thus be compared with other ALMP programmes. In recent years, empirical evidence on the effectiveness of German ALMP has been constantly growing. Following the introduction of new legislation at the end of the 1990s (*Sozialgesetzbuch III*, Social Code III) and especially the Hartz reforms in 2002, the FEA was required to evaluate the effectiveness of its ALMP programmes. To fulfil this obligation, researchers were provided access to the FEA's administrative data and several programmes were evaluated. For example, Lechner, Miquel and Wunsch (2005) and Biewen *et al.* (2006) evaluate the effectiveness of vocational training (VT) programmes, whereas Caliendo, Hujer and Thomsen (2008) concentrate on job-creation schemes (JCS). The findings are negative for JCS and mixed for VT programmes, where because of high locking-in effects at the beginning of VT, positive effects appear only after some time.

[1] See Caliendo and Steiner (2005) for a recent overview of German active labour market policy, including the 'Hartz reforms'.

The contribution of this paper is as follows: we evaluate the effectiveness of the two start-up programmes. As the major goal of German ALMP is to avoid future unemployment and integrate unemployed individuals into the primary labour market, we concentrate on the outcome variables 'not unemployed' and 'in paid or self-employment'. In addition, we analyse the programmes' effects on personal income. It should be clear that the aim of this paper is not to compare the relative success of the two programmes, e.g. with respect to the success of the businesses themselves (such as, number of employees). This is left to future studies as is the analysis of the monetary efficiency of these programmes (for the latter, see, Caliendo and Steiner, 2007).

Our analysis is based on a combination of administrative data from the FEA and a follow-up survey. The follow-up survey was necessary because: (1) administrative data are only available with a certain time lag; and (2) more importantly, they only contain information about employment for which social security contributions are compulsory, which is not the case for self-employment. The data contain approximately 3,100 participants in both programmes who founded a business in the third quarter of 2003 in West Germany.[2] The interviews took place at the beginning of 2005 and 2006, such that we observe individuals at least 28 months after the programmes started. Whereas for BA this means we can monitor the employment paths of individuals for at least 22 months after the programme has ended, SUS was still ongoing at the end of our observation period. At this stage, participants in SUS were in their third year of participation and were receiving a reduced transfer payment. Hence, results for this programme are only preliminary and interpretation hinges on this drawback. Additionally, we have a group of unemployed individuals (approximately 2,300) who were eligible for either programme but did not choose to participate in the third quarter of 2003. This non-participant group will function as our comparison group.

Given this informative data set, we base our analysis on the conditional independence assumption and use kernel-matching estimators to estimate the treatment effects. To test the sensitivity of the results with respect to unobserved differences we also use a conditional difference-in-differences strategy as suggested by Heckman *et al.* (1998b). The results show that at the end of our observation period both programmes are effective in terms of the above-mentioned outcome variables. Unemployment rates of participants are lower, and employment rates and personal income are higher when compared with non-participants.

The paper proceeds as follows. Section II gives a brief overview of the German labour market in the last decade, focusing on self-employment, unemployment and active labour market policies, whereas section III summarizes previous empirical findings. Section IV outlines our evaluation approach, while section V describes the data used for the analysis and discusses some implementation issues. Section VI presents the results and section VII concludes.

[2]We concentrate on West Germany in this paper because the labour market and especially self-employment dynamics in East Germany are quite different and have to be analysed separately (see Caliendo, 2008).

II. Unemployment, self-employment and start-up subsidies in Germany

Table 1 contains some summary statistics of the West German labour market. It can be seen that the self-employment rate has remained relatively stable over the last decade, fluctuating between 10% and 11% (relative to the workforce). Compared with other Organization for Economic Co-operation and Development (OECD) countries, this is relatively low. Blanchflower (2000) refers to numbers for 1996 and shows that only Denmark, Luxembourg, Norway and the United States have lower rates. On the other hand, the unemployment rate is persistently high, fluctuating between 7.3% and 9.1%.

To overcome this unemployment problem, the German government spends significant amounts on ALMP (approximately €12 billion in West Germany in 2004), including measures like vocational training programmes, job creation schemes, employment subsidies and self-employment of formerly unemployed individuals.

From 1986 to 2002, BA was the only programme providing support to unemployed individuals who wanted to start their own business. Its main goal is to cover basic costs of living and social security contributions during the initial stage of self-employment. BA supports the first 6 months of self-employment by providing the same amount that the recipient of a BA would have got if he or she had remained unemployed. As the unemployment scheme also covers social security contributions including health insurance, retirement insurance, etc., a lump sum for social security is granted, equivalent to 68.5% of the unemployment support that would have been received in 2003. Unemployed people are entitled to BA conditional on their business plan being approved externally, usually by the regional chamber of commerce. Thus,

TABLE 1

Self-employment, unemployment and start-up subsidies in West Germany, 1994–2004

	1994	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*
Self-employed* (in %)	10.4	10.6	10.4	10.2	10.3	10.4	10.6	11.0
Unemployed* (in %)	8.1	9.1	8.4	7.5	7.3	7.9	8.8	8.8
Supported self-employment (entries)								
BA (in thousand)	22.2	66.2	65.9	59.3	62.0	86.9	115.5	137.4
SUS (in thousand)	—	—	—	—	—	—	68.0	113.8
Total (in thousand)	22.2	66.2	65.9	59.3	62.0	86.9	183.5	251.1
Total† (in %)	0.9	2.4	2.5	2.5	2.7	3.5	6.7	9.0
ALMP expenditure (in bn Euro)								
ALMP – total	9.84	9.86	11.75	12.23	12.42	12.15	12.28	11.89
BA‡	0.06	0.43	0.55	0.53	0.58	0.73	1.09	1.37
SUS	—	—	—	—	—	—	0.18	0.67
Sup. self-empl. (total)	0.06	0.43	0.55	0.53	0.58	0.73	1.27	2.05
Sup. self-empl. (in %)	0.6	4.4	4.7	4.4	4.6	6.0	10.3	17.2

Notes: *Relative to the workforce.
†Relative to all unemployed.
‡The figures for the years 1994–98 are approximated.
Source: Bundesagentur für Arbeit (various issues).

TABLE 2

Design of the programmes

	Bridging allowance	*Start-up subsidy*
Entry conditions	Unemployment benefit *entitlement* Approval of the business plan by an external source (e.g. chamber of commerce)	Unemployment benefit *receipt* Approval of the business required since November 2004
Support	Participant receives UB for 6 months; to cover social security liabilities, an additional lump sum of approx. 70% is granted	Participants receive a fixed sum of €600/month in the first year, €360/month (€240/month) in the second (third) year Claim has to be renewed every year, income is not allowed to exceed €25,000 per year
Other	Social security is left at the individual's discretion	Participants are required to join the legal pension insurance and receive a reduced rate on the legal health insurance
Details	Section 57(1) Social Code III	Section 421(l) Social Code III

approval of an individual's application does not depend on the case manager at the local labour office.

In January 2003, an additional programme was introduced to support unemployed people in starting a new business. This 'start-up subsidy' was introduced as part of a large package of ALMP programmes introduced through the 'Hartz reforms'. The main goal of SUS is to secure the initial phase of self-employment. It focuses on the provision of social security to the newly self-employed person. The support is a lump sum of €600/month in the first year. A growth barrier is implemented in SUS such that the support is only granted if income does not exceed €25,000 per year. The support shrinks to €360/month in the second year and €240/month in the third. In contrast to the BA, SUS recipients are obligated to contribute to the statutory pension insurance fund, and may claim a reduced rate for statutory health insurance (Koch and Wießner, 2003). When the SUS was introduced in 2003, applicants did not have to submit business plans for prior approval, but have been required to do so since November 2004, as is the case with the BA as well (see Table 2 for more details on both programmes).

Hence, unemployed individuals can now choose between two programmes for help in starting their own business. Table 1 contains some information on participants and spending in measures promoting self-employment from 1994 to 2004. In 1994, about 1% of all unemployed individuals participated in BA, and the FEA spent 0.6% of their total resources for ALMP on BA. Because of a change in law in 1995 that made it easier to receive a BA, these numbers increased steadily up to 2002, when 3.5% of the unemployed received a BA (6.0% of the spending). Table 1 also shows that the SUS, when introduced, did not replace the BA, but did make self-employment significantly more attractive for the unemployed. In 2004, as much as 9% of Germany's unemployed participated in these two programmes together, thus absorbing a share of 17.2% of the total spending for ALMP.

Individuals planning to exit unemployment by entering self-employment can now choose between two alternative forms of start-up aid. One supports the first 6 months of self-employment by providing what the individual would have received in unemployment benefits plus a lump sum for social security contributions (BA), and the other provides a fixed and declining amount for the first 3 years of self-employment with the risk of losing the support if the growth barrier is exceeded (SUS). In this institutional framework, rational programme choice favours a BA if the unemployment benefits are fairly high, and/or if the income generated through the start-up firm is expected to exceed €25,000.

III. Previous empirical findings

In contrast to other ALMP programmes such as VT or JCS, the empirical evidence on the effectiveness of start-up subsidies for the unemployed is rather scarce. This might be explained by the fact that in most countries start-up subsidies usually form only one small component of ALMP. In 2003, the EU-15 countries spent an average of 0.697% of their GDP on ALMP, but only 0.034% of GDP on start-up subsidies. That is, of the total spending on ALMP, only 4.8% was used for these incentives (European Commission, 2005). The numbers in the last section have shown that this has changed substantially in Germany.

The main indicators used for evaluating self-employment programmes are the survival rate, the number of jobs created directly by the new business, and the employability and income of participants. Additionally, it is usually of interest whether there have been deadweight losses or displacement effects.[3] One also has to define the comparison group. Some studies do not have a comparison group at all (and focus, e.g. solely on survival rates); others use start-ups by those who were not previously unemployed as a benchmark or compare the income of self-employed programme participants with the income of individuals in paid employment. We have already pointed out that we use a different approach in this paper, comparing the outcomes of participants with other unemployed individuals. In the following, we give a brief overview of the findings in the literature on start-up subsidies for the unemployed, starting with some international evidence before turning to the results for Germany.

Meager, Bates and Cowling (2003) evaluate business start-up subsidies to young people in the UK. They not only look at the characteristics and survival of the start-ups but also compare the labour market outcomes of the participants with those of

[3]A deadweight loss occurs when behaviour is not changed due to the programme, e.g. when unemployed individuals would also have entered self-employment in the absence of the subsidy. Displacement effects take place, e.g. when the businesses set up by the participants drive other existing (unsubsidised) businesses out of the market.

a comparison group (similar in terms of age, gender, region and previous employment status). Based on multinomial and standard logistic regressions the authors conclude that participating in the programme does not have any significant impacts on subsequent employment or earnings chances. Perry (2006) uses difference-in-differences propensity score matching to evaluate the impact on males receiving an Enterprise Allowance grant – an integrated programme that provides business skills training as well as financial aid – in New Zealand between 1993 and 1995. The author's results (measured up to 2 years after participation) indicate statistically significant beneficial effects for the participants, where the outcome variable is 'not registered unemployed'. Cueto and Mato (2006) analyse the success of self-employment subsidies in one region of Spain using a Cox proportional hazards model. They look at the determinants of survival (duration) in self-employment and also estimate a competing risk model to distinguish between business failures and other reasons why businesses were closed. Based on data for individuals who received the subsidy between 1996 and 2000 and their labour market outcomes measured in December 2001, survival rates for 2–5 years can be observed and the survival is approximately 93% after 2 years and 76% after 5 years.

Comparisons are difficult because of the heterogeneity of the institutional settings of the different programmes, the economic conditions in the respective countries and the indicators used. The assumed deadweight losses range from low to high and are usually based on survey information of the participants (Meager, 1996). What should be kept in mind here is that even if a participant would have started a business anyway – that is even without a subsidy – it is unclear whether it would have been equally successful. Displacement effects are hardly ever analysed and would require a macroeconomic framework.

Conclusive evidence for Germany is even harder to find. Pfeiffer and Reize (2000) use the ZEW Firm Start-Up Panel in their study to compare a group of start-ups founded between 1993 and 1995 by formerly unemployed recipients of a BA with a group of start-ups not subsidized by a BA. Assessing business survival and employment growth, they find different effects for West and East Germany. Whereas start-ups by the unemployed in the East German regions have a 6% lower 1-year survival probability, no significant differences can be detected in West Germany. In terms of employment growth, subsidized start-ups by the unemployed are not different from non-subsidized start-ups. Reize (2004) uses the German Socio-Economic Panel (SOEP) and estimates competing risk models to model the paths out of unemployment. Comparing individuals moving into self-employment with those moving into paid employment shows that after 4 years, the unemployment risk is lower for the self-employed than for the other group. Both studies focus on the BA and have the problem of a rather small group of participants. Empirical evidence on the effectiveness of the SUS has not yet been produced, as the programme is relatively new. In the next section, we turn to a description of our evaluation approach.

IV. Identifying average treatment effects

We base our analysis on the potential outcome framework, also known as the Roy (1951)–Rubin (1974) model. The two potential outcomes are Y^1 (individual receives treatment, $D=1$) and Y^0 (individual does not receive treatment, $D=0$). The actually observed outcome for any individual i can be written as: $Y_i = Y_i^1 \cdot D_i + (1 - D_i) \cdot Y_i^0$. The treatment effect for each individual i is then defined as the difference between his/her potential outcomes: $\tau_i = Y_i^1 - Y_i^0$. As we can never observe both potential outcomes for the same individual at the same time, the fundamental evaluation problem arises. We will focus on the most prominent evaluation parameter, which is the average treatment effect on the treated (ATT), and is given by:

$$\tau_{\text{ATT}} = E(Y^1 \mid D=1) - E(Y^0 \mid D=1). \tag{1}$$

Given equation (1), the problem of selection bias can be straightforwardly seen, as the second term on the right-hand side is unobservable. It describes the hypothetical outcome without treatment for those individuals who received treatment. As with non-experimental data the condition $E(Y^0|D=1) = E(Y^0|D=0)$ is usually not satisfied, estimating ATT by the difference in sub-population means of participants $E(Y^1|D=1)$ and that of non-participants $E(Y^0|D=0)$ will lead to a selection bias. This bias arises because participants and non-participants are selected groups that would have different outcomes, even in the absence of the programme and might be caused by observable or unobservable factors.[4] We will combine two evaluation methods – matching and difference-in-differences – to cover both possible sources of selection bias.

4.1. Matching under unconfoundedness

Matching is based on the conditional independence (or unconfoundedness) assumption, which states that conditional on some covariates $W=(X, Z)$, the potential outcomes (Y^1, Y^0) are independent of D.[5] As we are interested in ATT only, we only need to assume that Y^0 is independent of D, because the moments of the distribution of Y^1 for the treatment group are directly estimable. That is:

Assumption 1. Unconfoundedness for Comparison Group: $Y^0 \amalg D|W$, where $\amalg$ denotes independence.

Clearly, this assumption may be a very strong one and has to be justified on a case-by-case basis, as the researcher needs to observe all variables that simultaneously influence participation and outcomes. We do so in section 5.2. Additionally, it has to be assumed that:

Assumption 2. Weak overlap: $\Pr(D=1|W) < 1$, for all W.

[4] See, e.g. Caliendo and Hujer (2006) for further discussion.
[5] See Imbens (2004) or Smith and Todd (2005) for recent overviews regarding matching methods.

This implies that there is a positive probability for all W of not participating, i.e. that there are no perfect predictors which determine participation. These assumptions are sufficient for identification of the ATT, which can be written as:

$$\tau_{\mathrm{ATT}}^{\mathrm{MAT}} = E(Y^1|W, D=1) - E_W[E(Y^0|W, D=0)|D=1], \tag{2}$$

where the first term can be estimated from the treatment group and the second term from the mean outcomes of the matched comparison group. The outer expectation is taken over the distribution of W in the treatment group.

As matching on W can become hazardous when W is of high dimension (curse of dimensionality), Rosenbaum and Rubin (1983) suggest the use of balancing scores $b(W)$. These are functions of the relevant observed covariates W such that the conditional distribution of W given $b(W)$ is independent of the assignment to treatment, i.e. $W \amalg D|b(W)$. The propensity score $P(W)$, i.e. the probability of participating in a programme, is one possible balancing score. For participants and non-participants with the same balancing score, the distributions of the covariates W are the same, i.e. they are balanced across the groups. Hence, Assumption 1 can be re-written as $Y^0 \amalg D|P(W)$ and the new overlap condition is given by $\Pr(D=1|P(W)) < 1$.

4.2. Combining matching with difference-in-differences

Even though we argue in section 5.2 that the CIA is most likely to hold in our setting, we test the sensitivity of our results with respect to unobserved heterogeneity. The matching estimator described so far assumes that after conditioning on a set of observable characteristics, (mean) outcomes are independent of programme participation. The conditional DID or DID matching estimator relaxes this assumption and allows for unobservable but temporally invariant differences in outcomes between participants and non-participants. This is achieved by comparing the conditional before/after outcomes of participants with those of non-participants. DID matching was first suggested by Heckman *et al.* (1998b). It extends the conventional DID estimator by defining outcomes conditional on the propensity score and using semiparametric methods to construct the differences. Therefore it is superior to DID as it does not impose linear functional form restrictions in estimating the conditional expectations of the outcome variable, and it re-weights the observations according to the weighting function of the matching estimator (Smith and Todd, 2005). If the parameter of interest is ATT, the DID propensity score matching estimator is based on the following identifying assumption:

$$E[Y_t^0 - Y_{t'}^0|P(W), D=1] = E[Y_t^0 - Y_{t'}^0|P(W), D=0], \tag{3}$$

where (t) is the post-treatment and (t') the pre-treatment period. It also requires the common support condition to hold and can be written as:

$$\tau_{\mathrm{ATT}}^{\mathrm{CDID}} = E(Y_t^1 - Y_{t'}^0|P(W), D=1) - E(Y_t^0 - Y_{t'}^0|P(W), D=0). \tag{4}$$

V. Implementing the estimators

Having discussed our evaluation approach in the previous section, we now present details on the implementation of the propensity score matching estimator. Caliendo and Kopeinig (2008) provide an extensive overview of the issues arising when implementing matching estimators. They point out that a crucial step is to discuss the likely validity of the underlying CIA. Hence, we deal with this issue in section 5.2, after having presented the data and some sample characteristics in section 5.1. This will be followed by the estimation of the propensity score and a discussion of matching details in section 5.3.

5.1. Data and some descriptives

We use a unique data set which combines administrative data from the FEA with survey data. For the administrative part we use data based on the 'Integrated Labour Market Biographies' (ILMB, *Integrierte Erwerbs-Biographien*) of the FEA, containing relevant register data from four sources: employment history, unemployment support recipience, participation in active labour market measures, and job seeker history. As the administrative data are only available with a certain time lag and more importantly do not provide any information on the employment status and/or income of self-employed individuals, we enriched the ILMB data with information from a computer-assisted telephone interview. To do so, we randomly drew participants from each programme who became self-employed in the third quarter of 2003. As we wanted to compare them with non-participants, we had to choose a comparison group. Choosing such a group is a heavily discussed topic in the recent evaluation literature. Although participation in ALMP programmes is not mandatory in Germany, the majority of unemployed persons participates at some point in time. Thus, comparing participants with individuals who never participate is inadequate, as it can be assumed that the latter group is particularly selective.[6] Sianesi (2004) discusses this problem for Sweden and argues that those who never participate did not enter a programme because they had already found a job. Additionally, as we did not know the future employment/participation status of the comparison group before the interviews took place, we restricted this comparison group to those who were unemployed in the third quarter of 2003, eligible for participation in either of the two programmes, but did not join a programme in this quarter. What should be kept in mind is that these comparison group members might participate in some ALMP programme after this quarter.[7]

[6]Furthermore, it should be noted that using individuals who never participate in the programmes as the comparison group may invalidate the conditional independence assumption due to conditioning on future outcomes (see discussion in Fredriksson and Johansson, 2004).

[7]The actual number of non-participants who participated in any ALMP programme after this quarter is rather low. It is approximately 5% after 12 months, 7% after 18 months and around 10% after 24 months.

TABLE 3

Selected descriptives

	Men			*Women*		
Variable	*NP*	*SUS*	*BA*	*NP*	*SUS*	*BA*
Number of observations	1,448	811	1,207	848	704	378
Qualificational variables						
School degree						
No degree	0.02	0.04	0.01	0.01	0.01	0.01
	(0.14)	(0.19)	(0.12)	(0.10)	(0.08)	(0.07)
Upper secondary schooling	0.24	0.18	0.26	0.32	0.27	0.40
	(0.43)	(0.38)	(0.44)	(0.47)	(0.44)	(0.49)
Job qualification						
High-qualified	0.20	0.12	0.24	0.22	0.17	0.33
	(0.40)	(0.32)	(0.42)	(0.41)	(0.37)	(0.47)
Unskilled	0.18	0.27	0.14	0.15	0.20	0.08
	(0.38)	(0.44)	(0.35)	(0.36)	(0.40)	(0.27)
Labour market history						
Previous unemployment duration						
<3 months	0.24	0.30	0.32	0.24	0.34	0.61
	(0.42)	(0.46)	(0.47)	(0.43)	(0.47)	(0.47)
>12 months	0.17	0.21	0.13	0.18	0.16	0.12
	(0.38)	(0.41)	(0.33)	(0.39)	(0.37)	(0.32)
No. months in employment in 2002	6.69	5.52	7.79	6.35	6.02	7.65
	(5.03)	(4.93)	(4.66)	(5.15)	(5.04)	(4.73)
Average daily earnings in 2002 (in €)	46.02	27.39	64.07	30.75	22.25	50.12
	(43.85)	(29.69)	(47.77)	(34.27)	(25.13)	(42.69)
Daily unemployment transfer (in €)	31.92	23.33	38.82	21.53	17.25	29.76
	(14.03)	(10.99)	(14.97)	(11.45)	(8.97)	(13.16)
Remaining time of UB (in months)	6.32	4.72	7.31	5.57	5.02	6.83
	(6.34)	(5.55)	(6.24)	(5.99)	(5.88)	(6.07)

Note: All variables are measured 1 month before start of programme. Values are given as mean [standard deviations (in parentheses)].

To minimize the survey costs we used a crude propensity score matching approach to select somewhat similar unemployed individuals.[8] These individuals were interviewed twice. The first interview took place in January/February 2005 and the second in January/February 2006. This enables us to observe the labour market activity of individuals for at least 28 months after programmes started. We compiled a sample of 3,100 individuals who had started a new business out of unemployment. Of these, 1,082 individuals received SUS and 2,018 received BA. Additionally, a control group of 2,296 non-participants was assembled.

[8]The potential comparison group consisted of roughly 640,000 individuals. Control individuals (for the interview) were chosen to resemble the distribution of some key variables – including gender, region, age, previous unemployment duration, qualification and nationality – in the population of the treated individuals. To do so, we estimated a 'crude propensity score' based on these variables and chose for every participant non-participants with a similar propensity score as interviewees.

A full list of the available variables can be found in Table A.1 in the Appendix; Table 3 contains sample mean values of the most relevant variables. What should be kept in mind is the non-random sample of non-participants. As we used a crude matching approach to make individuals similar, the non-participant sample does not represent a random sample of unemployed individuals. Clearly, this does not affect our estimation and interpretation strategy but should be kept in mind when interpreting the differences.

A first glance at the number of observations reveals clear gender differences in participation in both programmes. Whereas the male–female ratio is about 3:1 for BA, it is nearly 1:1 for the SUS. Further differences arise when looking at qualifications. In general it can be stated that participants in SUS are less qualified (when compared with BA participants). This is true for the comparison of the participants' qualifications either by highest school-leaving degree or the variable 'job qualifications', an assessment by the placement officer in the local labour office. Based on that, it is hardly surprising that participants in BA programmes also have a more favourable labour market history. Not only were they less frequently found among the long-term unemployed before starting a programme; they also had higher and longer claims for unemployment benefits. We discuss the available variables in more detail in the next section, where we also discuss the validity of the CIA.

5.2. Validity of the CIA and propensity score estimation

The CIA is in general a very strong assumption and the applicability of the matching estimator depends crucially on its plausibility. Blundell, Dearden and Sianesi (2005) argue that the plausibility of such an assumption should always be discussed on a case-by-case basis. Only variables that influence the participation decision and the outcome variable simultaneously should be included in the matching procedure. Hence, economic theory, a sound knowledge of previous research and information about the institutional setting should guide the researcher in specifying the model (see, e.g. Smith and Todd, 2005 or Sianesi, 2004).

Both economic theory and previous empirical evaluation studies highlight the importance of socio-demographic and qualificational variables. Regarding the first category we can use variables such as age, marital status, number of children, nationality (German or foreigner) and health restrictions. Additionally, we also use information whether individuals want to work full-time or part-time, and hence we might be able to approximate the labour market flexibility of these individuals. A second class of variables (qualification variables) refers to the human capital of the individual, which is also a crucial determinant of labour market prospects. The attributes available are school degree, job qualification and work experience. Furthermore, previous evaluation studies also point out that unemployment dynamics and labour market history play a major role in driving outcomes and programme participation. Hence, we use career variables describing the individual's labour market history. The available data in this regard are quite extensive (*inter alia* nearly complete 7-year labour market history; daily earnings from employment; amount of daily unemployment benefits; duration of last unemployment spell, employment status before unemployment, previous profession, etc.). Heckman *et al.* (1998b) also emphasize the importance of drawing treatment and comparison groups from the same local labour market and giving them the same questionnaire. To account for the situation on the local labour market, we use a classification of similar and comparable labour office districts derived by the FEA (see Blien *et al.*, 2004, for details). Additionally, participants and non-participants received the same questionnaire. As we have seen from the discussion in section II, the two programmes differ among other things in the size of the subsidy. Whereas the SUS is a lump sum, the BA depends on the amount of the unemployment benefits. Hence, we include the daily unemployment transfer payment before the start of the programme as an explanatory variable. In contrast to many other studies we are also able to include the remaining duration of unemployment benefits, which probably plays a deterministic role in these individuals' decision.[9]

Based on this exhaustive data, we argue that the CIA holds in our application. The set of variables is extensive and covers nearly all variables which have been identified to be important in previous evaluation studies of labour market policies. However, it should also be clear, that some variables which might influence self-employment dynamics are absent in our data (see Georgellis, Sessions and Tsisianis, 2005, for a recent overview). Even though one might argue that these variables, e.g. intergenerational links, are less important in our context (as we compare participants with other unemployed individuals), we test the sensitivity of the results with respect to time-invariant unobserved differences between participants and non-participants.

5.3. Propensity score estimation and matching details

As the choice probabilities are not known *a priori*, we have to replace them with an estimate. To do so, we estimate binary conditional probabilities for both programmes vs. non-participation. As we estimate the effects separately for men and women, we are left with four logit estimations. The results can be found in Table A.1 in the Appendix. To ensure comparability between the estimates we choose the same covariates for each combination and both genders. We do not interpret the results of the propensity score estimation, as we only use this estimation to reduce the dimensionality problem and the group of participants and non-participants are already quite similar because of the nature of the data.

The distribution of the propensity score is depicted in Figure 1. A visual analysis immediately suggests that the overlap between the group of participants and non-participants is sufficient in general. Nevertheless, there are some parts of the distribution (starting approximately at a propensity score value of 0.7) where the density of comparison individuals is quite thin. This is especially true for female participants in BA. However, by using the usual 'Minmax' criterion, where treated individuals are

[9]Lechner and Wunsch (2006) evaluate the effectiveness of ALMP (excluding start-up subsidies) in East Germany using a very similar set of variables.

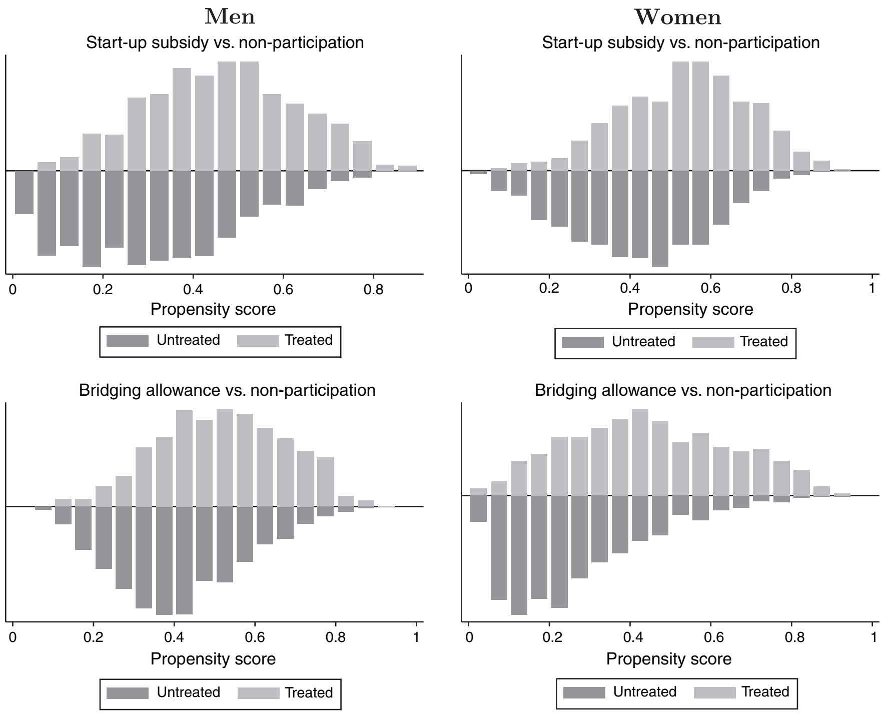

Figure 1. Distribution of the propensity scores – common support. Propensity score is estimated according to the specification in Table A.1. Participants are depicted in the upper half, and non-participants in the lower half of each figure.

excluded from the sample whose propensity score lies above the highest propensity score in the comparison group, only 13 individuals are dropped overall.[10]

Several matching procedures have been suggested in the literature, such as nearest-neighbour (NN) or kernel matching (KM).[11] To introduce them, a more general notation is needed: let I_0 and I_1 denote the set of indices for non-participants and participants respectively. We estimate the effect of treatment for each treated observation $i \in I_1$ in the treatment group by contrasting his/her outcome with treatment with a weighted average of control group observations $j \in I_0$ in the following way:

$$\Delta^{\mathrm{MAT}} = \frac{1}{N_1} \sum_{i \in I_1} \left[Y_i^1 - \sum_{j \in I_0} \omega_{N_0}(i,j) Y_j^0 \right], \tag{5}$$

where N_0 is the number of observations in the control group I_0 and N_1 is the number of observations in the treatment group I_1. Matching estimators differ in the weights

[10]We also test the sensitivity of the results with respect to more strict impositions of the common support requirement, e.g. by dropping 5% (10%) of the individuals where the overlap between participants and non-participants is especially low. It turns out that the results are not sensitive.

[11]See Heckman *et al.* (1998b), Smith and Todd (2005), and Imbens (2004) for overviews.

attached to the members of the comparison group, where $\omega_{N_0}(i, j)$ is the weight placed on the jth individual from the comparison group in constructing the counterfactual for the ith individual of the treatment group (Heckman *et al.*, 1998b). For example, with NN matching, only the closest neighbour is used to construct the counterfactual outcome, while KM is a non-parametric estimator that uses (nearly) all units in the control group. One major advantage of KM is the lower variance which is achieved because more information is used for constructing counterfactual outcomes. As our treatment and comparison groups are rather small, we focus in the later empirical application on this method.[12] An additional advantage of KM comes from the results of Heckman, Ichimura and Todd (1998a) who derive the asymptotic distribution of these estimators and show that bootstrapping is valid to draw inference for this matching method. This allows us to circumvent the issues raised by Abadie and Imbens (2006), pointing out that bootstrap methods are invalid for NN matching.

Before applying KM, assumptions have to be made regarding the choice of the kernel function and the bandwidth parameter h. The choice of the kernel appears to be relatively unimportant in practice (see, e.g. Jones, Marron and Sheather, 1996; Pagan and Ullah, 1999; DiNardo and Tobias, 2001). What is seen as more important is the choice of the bandwidth parameter h with the following trade-off arising: high values of h yield a smoother estimated density function, producing a better fit and a decreasing variance between the estimated and the true underlying density function. On the other hand, underlying features may be smoothed away by a large h, leading to a biased estimate. The choice of h is therefore a compromise between a small variance and an unbiased estimate of the true density function. Instead of using a 'rule of thumb' as proposed by Silverman (1986), we use 'leave-one-out' cross-validation (CV) as suggested in Black and Smith (2004) and Galdo (2005) to choose h. More details and most importantly, the chosen bandwidth parameters can be found in Table A.2 in the Appendix. We use these bandwidth parameters for further empirical analysis.[13]

To test if the matching procedure is able to balance all the covariates, we ran a standardized difference (SD) test (Rosenbaum and Rubin, 1985). This is a suitable indicator to assess the distance in marginal distributions of the W-variables. For each covariate W it is defined as the difference of sample mean values in the treated and matched control subsamples as a percentage of the square root of the average of sample variances in both groups. This is a common approach used in many evaluation studies, including those by Lechner (1999), Sianesi (2004) and Caliendo *et al.* (2008). Table 4 shows the mean standardized difference (MSD), i.e. the mean of the SD over all covariates before and after the matching.

It can be seen that the MSD before matching lies between 7.8% for women and 13.0% for men in SUS and even between 12.7% (men) and 18.6% (women) in BA.

[12] However, we also show that our results are not sensitive to the matching algorithm chosen (see Table A.3 in the Appendix for details).

[13] Estimations are done using the `PSMATCH2` Stata ado-package by Leuven and Sianesi (2003).

TABLE 4

Matching quality – some indicators

	Start-up subsidy		Bridging allowance	
Variable	Men	Women	Men	Women
MSD – before matching	13.049	7.780	12.658	18.577
MSD – after matching	1.375	2.133	1.303	2.612
R^2 – before matching	0.127	0.094	0.082	0.150
R^2 – after matching	0.003	0.007	0.003	0.008
χ^2 – before matching	0.000	0.000	0.000	0.000
χ^2 – after matching	1.000	1.000	1.000	1.000
Participants off support	1	7	4	1

Notes: Mean standardized difference (MSD) has been calculated as an unweighted average of the standardized difference of all covariates. Standardized difference before matching calculated as: $100(\overline{W}_1 - \overline{W}_0)/\{\sqrt{(V_1(W) + V_0(W))/2}\}$ and standardized difference after matching calculated as: $100(\overline{W}_{1M} - \overline{W}_{0M})/\{\sqrt{(V_1(W) + V_0(W))/2}\}$.

The matching procedure is able to balance the distribution of the covariates very well, especially for men, where the MSD after matching lies around 1.3%. For women in SUS, the MSD after matching is 2.1%; for women in BA it is 2.6%. In general, it is not sufficient to look at the MSD if one wants to judge the quality of the matching procedure. Instead a careful look at the SD for each variable is necessary, which, in our case, showed very satisfying results.[14] Additionally, Sianesi (2004) suggests re-estimating the propensity score on the matched sample (i.e. on the participants and matched non-participants) and comparing the pseudo-R^2 values before and after matching. After matching, there should be no systematic differences in the distribution of the covariates between the two groups. Therefore, the pseudo-R^2 after matching should be fairly low. As the results from Table 4 show, this is true for our estimation. The results of the F-tests point in the same direction, indicating a joint significance of all regressors before, but not after matching. Overall, these are satisfying results and show that the matching procedure was successful in balancing the covariates between treated individuals and members from the comparison group. Hence, we move on to the presentation of the results.

VI. Results

We discuss the effectiveness of the two programmes in relation to non-participation based on three outcome variables: first, we want to know if programme participation lowers the risk of returning to unemployment. To this end, we construct a variable that treats registered unemployment as a failure and all other possible states as a success (outcome variable A). As avoiding unemployment is one of the two major goals

[14]Detailed results are available on request from the authors. The highest SD after matching in a single variable lies at 4.0% for men in SUS and 4.0% for men in BA. For women, matching quality is slightly worse and the highest SD after matching lies around 7.5% for women in SUS and 7.4% for women in BA.

of German ALMP, this allows us to compare the effectiveness of the programmes in reaching this goal. A second aim is integration into regular, stable employment. Hence, we construct a second outcome variable which treats ongoing self-employment and regular paid employment as a success (outcome variable B). Finally, we also assess the effects of the programmes on the personal income of participants. We start the discussion with the employment effects over time, before presenting cumulated effects and income effects. For the latter two, we also present conditional difference-in-differences results to test the sensitivity.

6.1. Effects on the employment status over time

Figure 2 presents the treatment effects over time, where the upper panel relates to outcome variable A (not unemployed) and the lower part to outcome variable B (self-employed or in regular employment). Effects for men (women) are depicted on the left (right) side of each row. Rows 1 and 3 show the effects of participating in SUS vs. non-participation, whereas rows 2 and 4 show the effects of BA.

Effects start in the first month after the treatment has begun. Before starting the interpretation one has to note the following: a look at both figures shows a strong positive effect at the beginning of our observation period. This can be seen as a 'positive locking-in effect'. Whereas a locking-in effect usually corresponds to a negative effect during participation in a programme – for example, vocational training – the findings for our programmes are the opposite. Both participants and non-participants are unemployed in the month before the treatment starts, then participants join the programme and change immediately to the 'hoped-for' state. That is, they leave unemployment and become self-employed, which is viewed as a success for both outcome variables. Hence, one should not overemphasize this large effect at the start of the self-employment spell. BA runs out after 6 months, and a reasonable interpretation should start there. Clearly, for the 3-year-long SUS, the problem is that participants may receive aid during the complete observation period, interfering with interpretation. However, after 12 months, the transfer payment is reduced from €600 to €360 and after 24 months it is reduced further to €240. As this reduced payment is hardly sufficient to cover social security contributions, it gives us an initial idea of the success of the newly self-employed.

Let us start the discussion with outcome variable A. In the first months after treatment starts, we have very high positive effects for both programmes, lying well above 60 percentage points, irrespective of programme and gender. This means, for example, that the unemployment probability of participants in SUS or BA is about 60 percentage points lower than the unemployment probability of non-participants. Clearly, results at that point have to be interpreted with care, as both programmes are still ongoing. The effects show a negative time trend, where the paths of the programmes are very similar up to month 6. After that, the transfer payment for participants in BA terminates and the effects plunge. The downward trend continues but the rate of decrease is much lower. At the end of our observation period, that

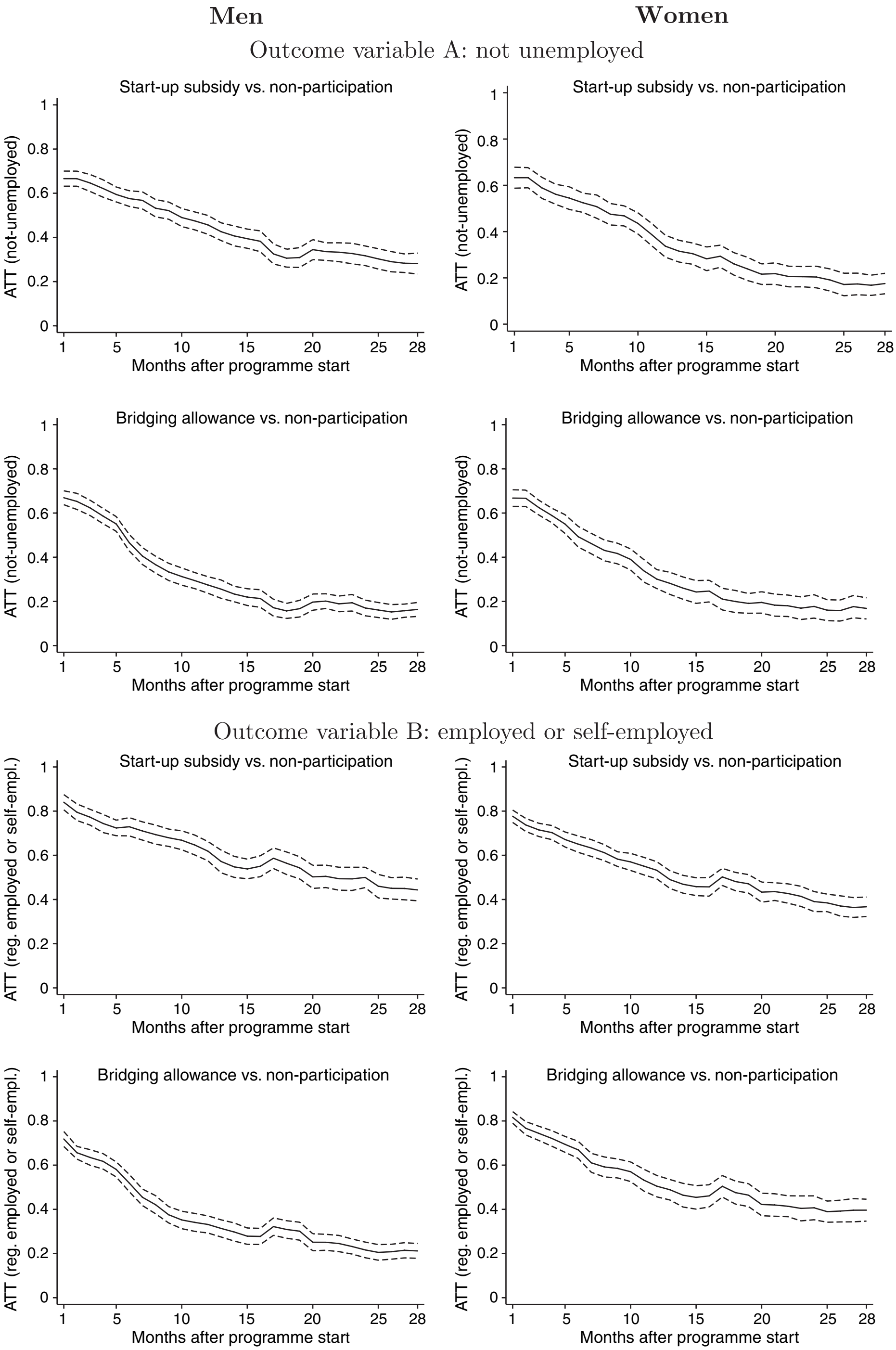

Figure 2. Treatment effects over time. Estimations are based on kernel matching as described in section 5.3. Bootstrapped standard errors are based on 200 replications.

is, 28 months after programmes have started, we get an effect of approximately 17 percentage points for male and female participants in BA. If we look at the effect of SUS vs. non-participation, the downward trend is much smoother, spiking somewhat in month 12, but decreasing relatively constantly to an effect of 28.2 percentage points for males and 17.6 percentage points for females in month 28.[15] A similar pattern – but on a higher level – can be found for outcome variable B (see lower part of Figure 2). This is a strong indication that both programmes are not only effective in avoiding unemployment but that they also give individuals much higher chances of remaining employed (either in paid or self-employment). The differences in both outcome variables can be explained by the fact that outcome variable A only treats registered unemployment as a failure. When individuals retreat from the labour market – and this might be especially relevant for women – they are not counted as a failure. Hence, the second outcome variable, only treating individuals as a success if they are in employment, has more explanatory power.

6.2. Cumulated effects

Table 5 contains the cumulative effects over time, i.e. the cumulative monthly effects over the observation period. For the outcome variable 'not unemployed' this shows the difference in months spent in unemployment between participants and non-participants. It can be seen that male participants in SUS spend roughly 12.2 months less in unemployment than non-participants. For female participants in SUS the effect is approximately 9.7 months. The cumulative effect for participants in BA is slightly lower, at 8.6 months for men and 9.1 months for women. We have already discussed that the effects for the outcome variable 'self-employment or paid employment' are even higher, which is also reflected by the cumulative effects of around 14.7 (16.9) months for men (women) in SUS and 10.2 (14.8) months for men (women) in BA.

As outlined in section 4.2 we also tested the sensitivity of our results with respect to time-invariant unobserved heterogeneity by using a conditional difference-in-differences approach. Before using such an approach, one has to determine the reference level for the before/after difference. We choose three different time periods for the comparison. In the first approach we use the time period from 1997 to 2002, that is, the 6-year employment history before entering the programme. For the first outcome variable, we sum the months not spent in unemployment, whereas for the second, we sum the months spent in paid employment. Additionally, we restrict the reference period to the latest 3 years (2000–2002) as well as the earliest 3 years (1997–99).

[15]The dip in the effects, especially for men, between months 16 and 20, is caused by a change in the interview information. Individuals were interviewed twice, in 2005 and 2006. Months 16 to 20 might involve a time overlap between the first and second interview and might be prone to recall errors. Hence, information for these months should be interpreted with care. For the overall interpretation, especially when moving towards the end of the observation period, this should not pose any problems.

TABLE 5

Cumulated effects – matching and conditional DID

	Start-up subsidy				*Bridging allowance*			
	Men		*Women*		*Men*		*Women*	
Outcome	*Effect*	*SE*	*Effect*	*SE*	*Effect*	*SE*	*Effect*	*SE*
Outcome variable A: not unemployed (in months)								
Matching	12.19	0.418	9.72	0.493	8.55	0.362	9.13	0.491
DID-1	11.82	0.792	9.26	0.768	8.46	0.403	6.79	0.742
DID-2	11.97	0.503	9.80	0.611	8.63	0.372	7.96	0.521
DID-3	12.03	0.488	9.18	0.631	8.39	0.358	7.96	0.520
Outcome variable B: employed or self-employed (in months)								
Matching	14.66	0.474	16.87	0.496	10.17	0.382	14.76	0.505
DID-1	14.61	0.944	15.85	1.332	9.83	0.930	6.75	1.347
DID-2	14.50	0.636	16.63	0.764	10.17	0.599	11.67	0.917
DID-3	14.77	0.791	16.09	0.906	9.83	0.683	9.83	1.012

Notes: Matching estimates are based on kernel matching as discussed in section 5.3. Standard errors are based on 200 bootstrap replications.

Reference level for DID 1: Total month not spend in unemployment (outcome variable A) and spend in regular employment (outcome variable B) between 1997 and 2002.

Reference level for DID 2: same as DID-1, but for the time period 2000–2002.

Reference level for DID 3: same as DID-1, but for the time period 1997–99.

Looking at the table, we see that the results are remarkably stable. For example, the effect on outcome variable B for men in SUS was 14.66 months with the matching approach and varies between 14.50 and 14.77 months with the CDID approaches. For women in SUS and men in BA the variation is slightly higher, but still negligible. This shows that additionally controlling for possible unobserved differences between participants and non-participants did not add much information for our estimates. This can be seen as evidence of the validity of the CIA in our context. Results are less favourable when looking at the smallest group under observation, that is, women in BA. Here the matching estimates are 9.13 months (not unemployed) and 14.76 months (regular or self-employed). The CDID results, however, vary from 6.8 to 8.0 in the first case and 6.8 to 11.7 in the second case. This indicates that unobservable differences between the group of female participants in BA and non-participants remain even after matching. Given the fact that the CDID estimates are smaller than the matching estimates, one could argue that there are unobserved factors that drive not only the participation probability but also labour market outcomes. Hence, we have to interpret these effects with caution.

6.3. Effects on the personal income

After having established that participants in both programmes are more likely to be employed and less likely to be unemployed than non-participants, we now investigate

TABLE 6

Effects on monthly income – matching and conditional DID

	Start-up subsidy				*Bridging allowance*			
	Men		*Women*		*Men*		*Women*	
Outcome	*Effect*	*SE*	*Effect*	*SE*	*Effect*	*SE*	*Effect*	*SE*
Effect on monthly income from self-employment/regular employment (in €)								
Matching	596.27	68.00	298.96	66.77	770.69	96.22	975.80	115.19
DID 1	601.44	67.04	295.57	67.97	768.72	84.13	738.01	108.72
DID 2	586.95	76.58	289.80	65.28	742.37	95.54	429.69	113.86
DID 3	586.76	72.22	316.32	78.26	770.97	96.72	510.58	117.38
Effect on total monthly income (in €)								
Matching	465.99	64.36	237.46	66.58	639.23	82.20	950.87	111.04
DID 1	471.17	66.83	234.07	68.73	637.26	82.18	713.08	109.23
DID 2	456.67	61.83	228.29	65.19	610.90	91.14	404.76	114.74
DID 3	456.48	72.48	254.81	71.09	639.51	94.35	485.65	125.22

Notes: Matching estimates are based on kernel matching as discussed in section 5.3. Standard errors are based on 200 bootstrap replications.

Reference level for DID 1: unemployment benefit before programme start.

Reference level for DID 2: average monthly income in 2002.

Reference level for DID 3: average monthly income from regular employment in 2002.

whether participants also earn more money. We use two income-related outcome variables: the more relevant variable is monthly income from self-employment or paid employment (labour income). However, as it is often argued that differences between (low) labour income and unemployment benefits are especially low in Germany, we also look at the total personal income of individuals, that is, including support such as unemployment benefits.

Table 6 contains the results for both outcome variables. Once again, we first present the results from matching estimates before presenting CDID results.[16] It is quite striking that all participants have significantly higher incomes than non-participants for both possible outcome variables. The upper half of Table 6 reveals that male participants in SUS earn around €600 per month more than their counterparts in the comparison group. Once again, the CDID does not add much information to the matching estimates as all estimates range between €586 and €601. For female participants, the effect is much lower (around €290) but still significant. The effects for the participants in BA are even higher. Male participants earn about €770 more per month. For females, we once again have the problem that matching and CDID results differ significantly, making it hard to draw relevant policy conclusions. Hence, we can conclude that participating in either of the two programmes has helped

[16]For the DID procedure we use three reference levels: (1) the monthly unemployment benefits before the programme started; (2) the average monthly income in 2002; and (3) the average monthly income from regular employment in 2002.

individuals to earn more money at the end of our observation period. This stays true even if we use the total personal income of individuals as an outcome variable, where we additionally take unemployment benefits and other government transfers into account.

VII. Conclusion

The aim of this paper was to evaluate the effectiveness of two active labour market programmes in Germany designed to encourage unemployed people to become entrepreneurs. These programmes have the potential not only to combat Germany's problem of persistently high unemployment, but also to increase its notoriously low self-employment rate. Our analysis is based on a data set that combines administrative with survey data and allows us to follow the employment paths of individuals for up to 28 months after programmes have started. For the first programme under consideration – the bridging allowance – we observed participants for 22 months after the programme ended. However, participants in the second programme – the start-up subsidy – are in their third year of participation at the end of our observation period, and most likely still receive further support (although at a reduced rate). Therefore, the results for SUS have to be treated as preliminary. Given the relatively stable participant structure in the BA programme since the introduction of the SUS, one can argue that the SUS attracts a different 'clientele' for self-employment. In general it can be stated that participants in SUS are less qualified (when compared with BA participants), and that this programme is frequently used by women.

We have evaluated the effectiveness of both programmes relative to non-participation. To this end we used a KM estimator and a conditional difference-and-differences estimator. Three outcome variables were of major interest. The first was 'not unemployed', corresponding to one of the main aims of the FEA. The second one combines the two possible labour market states 'in self-employment' and 'in paid employment' into one success criterion. The results indicate that both programmes are successful: at the end of our observation period, the unemployment rate of participants in BA was approximately 17 percentage points lower than that of non-participants, and for participants in SUS, around 18 percentage points lower for women and as much as 29 percentage points lower for men. Additionally, both the probability of being in self-employment and/or paid employment and the personal income are significantly higher for participants.

This is one of the first studies that allows inferences to be drawn about the effectiveness of start-up programmes that comprise part of Germany's ALMP. In contrast to other German ALMP programmes that have been evaluated recently (including job creation schemes and vocational training programmes), we find considerable positive effects for these two programmes. Hence, programmes aimed at turning the unemployed into entrepreneurs may be among the most promising for ALMP, both in Germany and elsewhere.

To allow more precise policy recommendations, further research is needed. First of all, the relative effects of both programmes should be estimated, which would allow their respective designs to be judged, as well as their suitability for different target groups. Additionally, it would be of interest to look at the development of the start-ups in terms of turnover and number of jobs directly created. Such an investigation would also enable an extensive cost–benefit analysis taking direct (and indirect) costs and benefits into account.

Final Manuscript Received: February 2008

References

Abadie, A. and Imbens, G. (2006). *On the Failure of the Bootstrap for Matching Estimators*, Working Paper, Harvard University.

Biewen, M., Fitzenberger, B., Osikominu, A. and Waller, M. (2006). *Employment Effects of Short and Medium Term Further Training Programs in Germany in the Early 2000s*, Working Paper, Goethe-University, Frankfurt.

Black, D. and Smith, J. (2004). 'How robust is the evidence on the effects of the college quality? Evidence from matching', *Journal of Econometrics*, Vol. 121, pp. 99–124.

Blanchflower, D. (2000). 'Self-employment in OECD countries', *Labour Economics*, Vol. 7, pp. 471–505.

Blien, U., Hirschenauer, F., Arendt, M., Braun, H. J., Gunst, D.-M., Kilcioglu, S., Kleinschmidt, H., Musati, M., Ross, H., Vollkommer, D. and Wein, J. (2004). 'Typisierung von Bezirken der Agenturen der Arbeit', *Zeitschrift für Arbeitsmarktforschung*, Vol. 37, pp. 146–175.

Blundell, R., Dearden, L. and Sianesi, B. (2005). 'Evaluating the impact of education on earnings in the UK: Models, methods and results from the NCDS', *Journal of the Royal Statistical Society, Series A*, Vol. 168, pp. 473–512.

Bundesagentur für Arbeit (various issues). *Arbeitsmarkt*, Bundesagentur für Arbeit, Nürnberg.

Caliendo, M. (2008). *Start-up Subsidies in East Germany: Finally, a Policy that Works?* Discussion Paper No. 3360, IZA, Bonn.

Caliendo, M. and Hujer, R. (2006). 'The microeconometric estimation of treatment effects – an overview', *Allgemeines Statistisches Archiv*, Vol. 90, pp. 197–212.

Caliendo, M. and Kopeinig, S. (2008). 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys*, Vol. 22, pp. 31–72.

Caliendo, M., Hujer, R. and Thomsen, S. (2008). 'The employment effects of job creation schemes in Germany – a microeconometric evaluation', *Advances in Econometrics*, Vol. 21, pp. 383–430.

Caliendo, M. and Steiner, V. (2005). 'Aktive Arbeitsmarktpolitik in Deutschland: Bestandsaufnahme und Bewertung der mikroökonomischen Evaluationsergebnisse', *Zeitschrift für Arbeitsmarktforschung/Journal for Labour Market Research*, Vol. 38, pp. 396–418.

Caliendo, M. and Steiner, V. (2007). *The Monetary Efficiency of Start-up Subsidies in Germany*, Mimeo, Berlin/Bonn.

Cueto, B. and Mato, J. (2006). 'An analysis of self-employment subsidies with duration models', *Applied Economics*, Vol. 38, pp. 23–32.

DiNardo, J. and Tobias, J. (2001). 'Nonparametric density and regression estimation', *Journal of Economic Perspectives*, Vol. 15, pp. 11–28.

European Commission (2005). *Labour Market Policy – Expenditure and Participants, Data 2003*. Office for Official Publications of the European Communities, Luxembourg.

Fredriksson, P. and Johansson, P. (2004). *Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Data*, Discussion Paper No. 1062, IZA.

Galdo, J. (2005). *Bandwidth Selection and the Estimation of Treatment Effects with Non-experimental Data*, Working Paper, Syracuse University.

Georgellis, Y., Sessions, J. and Tsisianis, N. (2005). *Self-employment Longitudinal Dynamics: A Review of the Literature*, Discussion Paper, Brunel University.

Heckman, J., Ichimura, H. and Todd, P. (1998a). 'Matching as an econometric evaluation estimator', *Review of Economic Studies*, Vol. 65, pp. 261–294.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998b). 'Characterizing selection bias using experimental data', *Econometrica*, Vol. 66, pp. 1017–1098.

Imbens, G. (2004). 'Nonparametric estimation of average treatment effects under exogeneity: a review', *The Review of Economics and Statistics*, Vol. 86, pp. 4–29.

Jones, M., Marron, S. and Sheather, S. (1996). 'A brief survey of bandwidth selection for density estimation', *Journal of the American Statistical Association*, Vol. 91, pp. 401–407.

Koch, S. and Wießner, F. (2003). 'Wer die Wahl hat, hat die Qual', IAB Kurzbericht, 02/2003, Nuremberg.

Lechner, M. (1999). 'Earnings and employment effects of continuous off-the-job training in East Germany after unification', *Journal of Business Economic Statistics*, Vol. 17, pp. 74–90.

Lechner, M., Miquel, R. and Wunsch, C. (2005). *The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany After Unification*, Discussion Paper No. 1684, IZA, Bonn.

Lechner, M. and Wunsch, C. (2006). *Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off*, Discusssion Paper No. 2363, IZA, Bonn.

Leuven, E. and Sianesi, B. (2003). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*. Available at: http://www. ideas.repec.org/c/boc/bocode/s432001.html

Meager, M., Bates, P. and Cowling, M. (2003). 'An evaluation of business start-up support for young people', *National Institute Economic Review*, Vol. 186, pp. 59–72.

Meager, N. (1996). 'From unemployement to self-employment: Labour market policies for business start-up', in Schmidt G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar, Cheltenham, pp. 489–519.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Perry, G. (2006). *Are Business Start-up Subsidies Effective for the Unemployed: Evaluation of Enterprise Allowance*, Working Paper, Auckland University of Technology.

Pfeiffer, F. and Reize, F. (2000). 'Business start-ups by the unemployed – an econometric analysis based on firm data', *Labour Economics*, Vol. 7, pp. 629–663.

Reize, F. (2004). *Leaving Unemployment for Self-employment, An Empirical Study*, ZEW Economic Studies, Vol. 25, Nomos, Baden-Baden, Mannheim.

Rosenbaum, P. and Rubin, D. (1983). 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, Vol. 70, pp. 41–50.

Rosenbaum, P. and Rubin, D. (1985). 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *The American Statistican*, Vol. 39, pp. 33–38.

Roy, A. (1951). 'Some thoughts on the distribution of earnings', *Oxford Economic Papers*, Vol. 3, pp. 135–145.

Rubin, D. (1974). 'Estimating causal effects to treatments in randomised and nonrandomised studies', *Journal of Educational Psychology*, Vol. 66, pp. 688–701.

Sianesi, B. (2004). 'An evaluation of the Swedish system of active labour market programmes in the 1990s', *The Review of Economics and Statistics*, Vol. 86, pp. 133–155.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Smith, J. and Todd, P. (2005). 'Does matching overcome LaLonde's critique of nonexperimental estimators?' *Journal of Econometrics*, Vol. 125, pp. 305–353.

Appendix

TABLE A1

Propensity score estimation results – coefficients

	SUS vs. non-participation		*BA vs. non-participation*	
	Men	*Women*	*Men*	*Women*
Socio-demographic characteristics				
Age category (Ref.: 18–24)				
25–29	0.617*	1.030*	0.298	0.526
30–34	0.871*	0.834**	0.274	0.419
35–39	0.481***	0.678***	0.252	0.486
40–44	0.669*	0.861**	0.105	0.427
45–49	0.652**	0.824**	0.165	0.425
50–64	1.223*	1.369*	0.356	0.803***
Children (Ref.: no children)				
One child	0.183	0.097	−0.133	0.272
Two or more children	0.000	−0.027	−0.257***	0.036
Nationality: German	0.016	−0.074	0.180***	−0.127
Desired working time: full-time	−0.209	−0.023	−0.085	0.682*
Qualification variables				
School degree (Ref.: no degree)				
Lower secondary schooling	−0.064	0.903	0.370	0.165
Middle secondary schooling	−0.024	0.864	0.443	0.536
Specialised upper sec. schooling	−0.036	0.764	0.411	0.589
Upper secondary schooling	−0.160	1.019***	0.380	0.323
Occupational group in previous profession (Ref.: manufacturing)				
Agriculture	0.517	−0.232	0.229	−0.376
Technical	−0.566**	0.491	0.235	−0.090
Services	−0.113	−0.101	−0.047	0.084
Other	−0.476**	−0.408	−0.715*	−0.535
Labour market history				
Duration of last unemployment (Ref.: <3 months)				
3 months to <6 months	−0.353**	−0.907*	−0.406*	−0.748*
6 months to <1 year	−0.436*	−0.450*	−0.459*	−0.284
≥1 year	−0.517*	−0.696*	−0.629*	−1.140*
With work experiences	−0.129	−0.340***	−0.169	−0.585*
Number of placement propositions	−0.004	−0.015	−0.015**	−0.019
Unemployment benefits	−0.046*	−0.032*	0.022*	0.035*
Remaining benefit entitlement	−0.041*	−0.065*	−0.056*	−0.050**
Daily income from regular employment				
1999	0.002	0.002	0.005**	0.008***
2000	−0.003	−0.000	−0.005	−0.004
2001	0.005	−0.004	0.008**	−0.001
2002	−0.012*	−0.006	0.002	0.002

continued overleaf

TABLE A1

(continued)

	SUS vs. non-participation		*BA vs. non-participation*	
	Men	*Women*	*Men*	*Women*
Constant	−0.393	−1.537	−0.500	−1.224
Log-likelihood	−1,196.329	−885.819	−1,546.651	−596.322
Hit-rate	40.133	48.135	47.317	39.896

Note: *,**,*** Significance at the 1%, 5% and 10% level respectively.

Additional variables included: family status, health restrictions, job qualification, months spend in regular employment and unemployment in the years 1999, 2000, 2001 and 2002, employment status before unemployment, and dummy variables for the regional labour market context (strategy clusters). Full estimation results and marginal effects are available on request from the authors.

TABLE A2

Cross-validation for the bandwidth selection

Start-up subsidy				*Bridging allowance*			
Men		*Women*		*Men*		*Women*	
h	*RMSE*	*h*	*RMSE*	*h*	*RMSE*	*h*	*RMSE*
0.00558	0.45009	*0.04673*	*0.38910*	*0.09087*	*0.42773*	1.04953	0.36659
0.01558	0.45110	0.05673	0.38956	0.10087	0.42788	1.05953	0.36659
0.02558	0.45016	0.06673	0.38955	0.11087	0.42795	1.06953	0.36658
0.03558	0.44909	0.07673	0.38962	0.12087	0.42803	1.07953	0.36658
0.04558	0.44878	0.08673	0.38963	0.13087	0.42810	1.08953	0.36658
0.05558	*0.44870*	0.09673	0.38963	0.14087	0.42823	1.09953	0.36658
0.06558	0.44887	0.10673	0.38972	0.15087	0.42837	1.10953	0.36658
0.07558	0.44916	0.11673	0.38976	0.16087	0.42851	1.11953	0.36658
0.07558	0.44916	0.12673	0.38973	0.17087	0.42866	1.12953	0.36658
0.07558	0.44916	0.13673	0.38970	0.18087	0.42882	1.13953	0.36658
0.07558	0.44916	0.14673	0.38963	0.19087	0.42897	*1.14953*	*0.36657*

Note: We implement leave-one out cross-validation in a five-step procedure (see, e.g. Galdo, 2005), the chosen values are printed in *italics*.

Step 1. Define a bandwidth search grid. Here, we use $l_{\mathrm{bw}} + 0.05 \times g$ for $g = 0, 1, 2, \ldots, 20$, where $l_{\mathrm{bw}} = \max[\min[|P_{0i} - P_{0-i}|, |P_{0i} - P_{0+i}|]]$ is a lower bound defined by the propensity score values of comparison group members in the support region.

Step 2. Starting with the lowest bandwidth and using only the comparison sample, estimate the counterfactual outcome of each comparison unit using kernel matching on the remaining $N_0 - 1$ observations. Find the weighted MISE for that particular bandwidth.

Step 3. Repeat step 2 for each of the remaining bandwidth values. Find the particular bandwidth h^+ that minimizes the weighted MISE across all estimations.

Step 4. Refine the bandwidth h^+ by defining a +0.05 neighbourhood around h^+ and select a new search grid.

Step 5. Repeat steps 2 and 3 and select the bandwidth that yields the minimum weighted MISE among all estimations.

TABLE A3

Sensitivity of the results with respect to different matching algorithms – outcome variable: cumulated effects 'not unemployed'

	Start-up subsidy						*Bridging allowance*					
	Men			*Women*			*Men*			*Women*		
	Effect	*SE*	*OS*	*Effect*	*SE*	*OS*	*Effect*	*SE*	*OS*	*Effect*	*SE*	*OS*
Kernel matching*												
epan nocommon	12.19	0.418	0	9.72	0.493	1	8.55	0.362	0	9.13	0.491	0
normal nocommon	12.09	0.453	0	9.82	0.497	0	8.79	0.352	0	9.16	0.437	0
epan common	12.19	0.425	1	9.83	0.520	7	8.55	0.346	4	9.13	0.411	1
normal common	12.09	0.414	1	9.92	0.494	7	8.78	0.370	4	9.15	0.461	1
epan common trim(10)	12.27	0.436	76	10.02	0.472	66	8.65	0.356	112	9.01	0.464	36
normal common trim(10)	12.14	0.428	76	10.17	0.462	66	8.91	0.330	112	9.02	0.521	36
epan common trim(5)	12.19	0.413	38	10.10	0.475	33	8.59	0.380	56	9.07	0.462	18
normal common trim(5)	12.10	0.407	38	10.13	0.480	33	8.85	0.333	56	9.09	0.476	18
Nearest-neighbour matching†												
withrep nocommon	12.08	0.630	0	8.79	0.753	0	8.38	0.543	0	8.13	0.854	0
norep nocommon	12.13	0.452	0	10.36	0.443	0	9.43	0.355	0	8.79	0.607	0
withrep common	12.10	0.645	1	8.93	0.697	7	8.32	0.485	4	8.15	0.885	1
norep common	12.11	0.441	1	10.33	0.429	7	9.40	0.376	4	8.79	0.552	1
Caliper 0.01	12.06	0.612	5	8.91	0.782	7	8.37	0.458	1	8.04	0.892	6
Caliper 0.02	12.08	0.642	0	8.84	0.764	3	8.37	0.501	1	8.13	0.834	0

Note: Standard errors are based on 200 bootstrap replications. OS (off support) indicates the number of treated individuals discarded because of missing common support.

*Kernel matching algorithms are implemented with two kernel functions (normal/Gaussian and Epanechnikov), with and without common support (common and nocommon) and partly with the additional imposition of a trimming level of 5% and 10%.

†Nearest-neighbour algorithms are implemented with and without common support (common and nocommon), with and without replacement (withrep and norep). Additionally, caliper matching with two caliper levels (0.01 and 0.02) is implemented.