Journal of Public Economics 95 (2011) 311–331

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity

Marco Caliendo [a,c,d,*], Steffen Künn [a,b]

[a] IZA, Germany
[b] FU Berlin, Germany
[c] DIW Berlin, Germany
[d] IAB, Germany

ARTICLE INFO

Article history:
Received 25 February 2010
Received in revised form 29 September 2010
Accepted 4 November 2010
Available online 19 November 2010

JEL classification:
J68
C14
H43

Keywords:
Start-up subsidies
Self-employment
Evaluation
Long-term effects
Effect heterogeneity

ABSTRACT

Turning unemployment into self-employment has become an increasingly important part of active labor market policies (ALMP) in many OECD countries. Germany is a good example where the spending on start-up subsidies for the unemployed accounted for nearly 17% of the total spending on ALMP in 2004. In contrast to other programs—like vocational training, job creation schemes, or wage subsidies—the empirical evidence on the effectiveness of such schemes is still scarce; especially regarding long-term effects and effect heterogeneity. This paper aims to close this gap. We use administrative and survey data from a large sample of participants in two distinct start-up programs and a control group of unemployed individuals. We find that over 80% of participants are integrated in the labor market and have relatively high labor income five years after start-up. Additionally, participants are much more satisfied with their current occupational situation compared to previous jobs. Based on propensity score matching methods we estimate the long-term effects of the programs against non-participation and take great care in assessing the sensitivity of our results with respect to deviations from the identifying assumption. Our results turn out to be robust and show that both programs are effective with respect to income and employment outcomes in the long-run, i.e., five years after start-up. Moreover, we consider effect heterogeneity with respect to several dimensions and show that start-up subsidies for the unemployed tend to be most effective for disadvantaged groups in the labor market.

1. Introduction

The recent OECD report on income and poverty (OECD, 2008) illustrates an increase in poverty rates over the past decade, where the risk of becoming poor shifted from the elderly in particular towards children and people of working age. The importance of employment in this context is straightforward as poverty among non-working households increased sharply during the last decade. The poverty rate[1] for households where the head is of working age but no household member actually works amounted to 36% and was three (twelve) times higher than for households with one (two or more) worker in the mid-2000s. Despite cross-country variation in terms of the scope of poverty, the negative correlation between employment rates and poverty is throughout valid. In an earlier study, Sen (1997) presents different concepts on how unemployment may cause poverty and inequality due to social exclusion. The main idea is that specific groups of individuals are generally excluded from the labor market, for example low skilled or youth. In addition, economic conditions may also foster social exclusion. He argues that along with the abolishment of social exclusion, unemployment and therefore poverty will be reduced. Governments are fully aware of this concept and therefore spend significant amounts of their budget on active labor market policies (ALMP) to equalize labor market conditions of unemployed individuals, in which a special focus is usually put on disadvantaged groups. By removing severe differences in terms of education, work experience or productivity, existing labor market barriers are to be overcome, consequently reducing unemployment. Several labor market programs have been introduced in which the most popular programs are traditionally training measures such as retraining, classroom training or on-the-job training. Furthermore, employment subsidies, job creation schemes and job-search assistance have also been adapted by almost all OECD countries. These programs are supposed to integrate unemployed individuals in the labor market and are associated with an upward shift in income level to secure one's livelihood and an increase in life and job satisfaction. Much research has been dedicated to investigating the effectiveness of ALMP programs. Although positive results with respect to income and employment prospects were found occasionally, the overall evidence indicates that the effects of those traditional measures are rather disappointing (see Martin and Grubb, 2001; Dar and Gill, 1998; Dar and Tzannatos, 1999; Fay, 1996 for evidence on OECD countries and Kluve and Schmidt, 2002 for the European experience). In particular,

* Corresponding author. Institute for the Study of Labor (IZA), P.O. Box 7240, 53072 Bonn, Germany. Tel.: +49 228 3894 512; fax: +49 228 3894 510.
E-mail addresses: caliendo@iza.org (M. Caliendo), kuenn@iza.org (S. Künn).
[1] The poverty rate is defined as the share of people with an equivalised disposable income below 50% of the median of the entire population.

0047-2727/$ – see front matter
doi:10.1016/j.jpubeco.2010.11.003

job creation schemes turn out to be not appropriate for improving participants' employment perspectives.

On the other hand, it is found that the promotion of self-employment among unemployed individuals is a promising tool. Public authorities usually tie start-up subsidies with the hope for a “double dividend”. Besides creating a job for the self-employed themselves, the newly founded businesses may potentially create further jobs and thus reduce unemployment rates even further. Moreover, individuals who receive support also increase their employability, human capital and labor market networks during the period of self-employment, which, in the case of failure, makes them more able to find regular employment. Start-up subsidies may also be promising from a macroeconomic perspective, since the entry of new firms generally increases competition and consequently productivity of firms. This potentially can promote efficient markets and technology diffusion and might finally lead to economic stability and economic growth, i.e., an increase in wealth (see Storey, 1994; Fritsch, 2008). However, there are also some concerns related to financial promotion of start-ups by the unemployed. First of all, supported individuals may have become self-employed even without financial support. This is referred to as deadweight loss and is usually hard to determine.[2] Another concern addresses crowding out effects, whereby incumbent or non-subsidized firms may be displaced by supported start-ups. Finally, firms may also substitute employees with subsidized self-employed workers. Due to a highly regulated labor market in Germany, however, such substitution effects are likely to play only a minor role in practice.

We focus in our analysis on the effects of start-up subsidies on the participating individuals only, that is we do not address any macroeconomic or general-equilibrium effects. Most of the existing evaluation studies on start-up schemes report positive effects with respect to different labor market outcomes. The evidence varies with respect to countries and institutional design of support. A main shortcoming of previous studies is that they provide short- to medium-run evidence only and—especially in the case of industrialized countries—do not consider effect heterogeneity. If the analysis is conducted at a point at which individuals still receive the support, the results are likely to be upward biased due to locking-in effects. To properly judge the effects of the programs, the observation window needs to be (substantially) longer than the period of support. Furthermore, it can be assumed that there will be heterogeneity in the effects of these programs, which implies that some groups might benefit more and others less from participation. This is of special interest for particular disadvantaged groups, for example low educated or young individuals who are over-represented among the long-term unemployed and socially excluded. Knowing how start-up schemes work for these groups will help to design programs more appropriate and thereby tackle long-term unemployment, social exclusion, and the associated risk of poverty.

The aim of this paper is to close the existing research gap by providing long-term evidence and an extensive analysis with respect to effect heterogeneity for two distinct start-up subsidies for unemployed individuals in Germany. The first program—*bridging allowance* (BA, “Überbrückungsgeld”)—provided relatively high financial support (depending on individuals' previous earnings) to unemployed workers for six months; whereas the second program—*start-up subsidy* (SUS, “Existenzgründungszuschuss”)—consisted of (lower) monthly lump-sum payments for up to three years.[3] Since both schemes differ sharply in terms of financial support and duration, they also attracted different types of individuals. Using a combination of administrative and survey data, we are able to follow individuals for nearly five years after entering the programs. In addition, we also have access to a suitable control group of other unemployed individuals allowing us to use propensity score matching methods for the impact analysis. We take great care in assessing the sensitivity of our results with respect to deviations from the identifying assumption. Our results turn out to be robust and we find strong positive long-run effects nearly five years after start-up for both programs with respect to several labor market outcomes. In addition, we show that they are most effective for individuals at high risk of being excluded from the labor market, i.e., low educated and low qualified individuals.

The paper is organized as follows: Section 2 provides a brief literature review on ALMP in OECD countries, institutional details on start-up programs for the unemployed in Germany and a discussion of previous results on such measures. Afterwards we describe the data, present descriptive results and illustrate the identification and estimation strategy in Section 3. The main results are discussed in Section 4, which also contains an extensive analysis of effect heterogeneity. Finally, we present the sensitivity analysis in Section 5 before we conclude and give an outlook in Section 6.

2. ALMP to reintegrate unemployed individuals

2.1. Previous literature

The OECD reports an average spending of 0.6% of a country's GDP on ALMP among all OECD member states in 2007, and therefore, much research has been conducted investigating the effectiveness of such measures (see OECD, 2009). The main question is whether ALMP programs are appropriate for improving participants' labor market perspectives and in addition whether they also generate income gains for participants. For instance, Martin and Grubb (2001), Dar and Tzannatos (1999) and Fay (1996) review evaluation studies on ALMP across OECD countries and present mixed results for several programs. In fact, they do find some positive results for certain subgroups, for example training for the long-term unemployed, or women. Dar and Gill (1998) consider retraining programs in OECD countries and are not able to identify significant effects. Focusing on Europe, Kluve and Schmidt (2002) find strong heterogenous effects for different programs and subgroups and argue that job search assistance and training might be effective. Card et al. (2010) provide an international meta-analysis of recent evaluation studies on the effectiveness of ALMP programs and confirm the overall ineffectiveness of job creation schemes. Moreover, they find promising effects for classroom or on-the-job training in the medium-run. In an earlier review on the US and European experience, Heckman et al. (1999) point out that benefits from ALMP programs do not significantly reduce poverty or unemployment, however, employment gains are more likely to occur as an increase in income levels. Betcherman et al. (2004) provide an overview on the effectiveness of ALMP in developing and transition countries and find some positive results for employment services while training measures, public works and wage subsidies are rather unsuccessful. For Germany, Fitzenberger et al. (2008) and Lechner et al. (2004) find positive effects for training measures in the long-run. Moreover, Stephan (2008) and Stephan and Pahnke (2008) provide evidence for vocational training, short-term training, wage subsidies and job creation schemes and show consistently negative effects for job creation schemes (in line with Caliendo et al., 2008) and mostly positive but not always significant effects for the other programs under consideration. Lechner and Wunsch (2008) conclude that programs such as vocational training, wage subsidies, short-term training and assessment schemes are overall ineffective for the West German labor market. To sum up, despite occasionally positive results, the overall evidence indicates that traditional measures are rather disappointing. In particular job

[2] Meager (1993) provides an estimate of the deadweight effect related to the bridging allowance in Germany and concludes that the effect is rather small (about 10%).

[3] Both programs were replaced in August 2006 by a single new program—the new start-up subsidy program (*Gründungszuschuss*)—which will not be analyzed here.

creation schemes turned out to be not appropriate for improving participants' employment prospects, and training programs bring modest effects only in the (very) long-run.

In light of these findings, supporting unemployed individuals in becoming self-employed might be a promising tool among active labor market policies. The international evidence is still relatively scarce on such measures but predominantly indicates positive results. For developing countries, for instance, Almeida and Galasso (2007) investigate the impact of financial and technical assistance for welfare beneficiaries on their way to self-employment in Argentina. They observe a period of 12 months in 2004/2005 and find an increase in total working hours but no significant income effects due to the program. However, for young and highly educated individuals they are able to identify positive income effects. Rodriguez-Planas (2008) investigates a start-up program for Romania in which the participants obtained professional assistance through counseling or short-term entrepreneurial training. In addition, working capital loans were offered. She identifies positive employment effects but no income gains for participants compared to non-participants and reveals strong positive employment effects for a subgroup of low educated individuals. O'Leary (1999) considers self-employment schemes for Poland and Hungary. The scheme in Poland provides loans at market interest rates to the unemployed combined with the attractive option that 50% of repayments will be waived if firms survive at least two years. In contrast, the Hungarian program consists of unemployment benefits paid up to 18 months. In addition, it also incurs half of the costs for training and counseling. O'Leary (1999) finds large and positive employment effects for both countries. Whilst he is also able to identify strong positive earning effects for Hungary, the income effect in Poland is negative.[4] Among participants, O'Leary (1999) finds high survival rates in self-employment and additional employment effects in both countries. The findings are similarly positive for developed countries. Carling and Gustafson (1999) provide a comparative study between employment subsidies and self-employment grants for the unemployed in Sweden. They find that individuals in subsidized employment have a higher probability of re-entering unemployment than recipients of self-employment grants. Therefore, they conclude that self-employment grants are more effective in avoiding unemployment. Cueto and Mato (2006) analyze the success of self-employment subsidies for particular districts in Spain. They find survival rates of approximately 93% after two years and 76% after five. For New Zealand, Perry (2006) evaluates enterprise allowance grants, an integrated program that provides business skills training as well as financial aid. The author's results indicate a decrease in time registered as unemployed for participants. Meager et al. (2003) evaluate business start-up subsidies by the Prince Trust to young people in the UK. The authors conclude that participating in the program does not have any significant impact on subsequent employment or earning chances. Nonetheless, descriptively they find a fraction of 69.1% in self-employment among participants after 18 months. Kelly et al. (2002) consider an allowance paid up to 52 weeks as well as training and counseling in Australia. The authors find a survival rate of 56.2% in self-employment after three years following start-up.

In a study for Germany which is very closely related to ours, Baumgartner and Caliendo (2008) provide an evaluation of BA and SUS programs for the short- and medium-run. They find strong positive employment and income effects for participants compared to a group of non-participants. However, the authors underscore the preliminary character of their results, as the majority of start-up subsidy participants still received financial support during the observation period. Therefore, the survival rate is likely to further decrease after financial support completely expires. In an earlier study, Pfeiffer and Reize (2000) analyze the effect of BA on survival rates in self-employment during the first year after entry. They find neither differences in survival probability nor in employment growth between supported and non-subsidized firms in West Germany.

To summarize, the existing literature on start-up schemes for the unemployed mainly reports either positive or insignificant effects with respect to different labor market outcomes; whilst negative impacts are scarce. Effect heterogeneity is considered only by studies on developing countries. The evidence varies with respect to countries, institutional design of the support and entrance conditions. However, the main shortcoming is that existing studies provide evidence for the short to medium-run only. Long-term evidence is indeed highly demanded by the literature but—due to data limitations—still missing. We are now able to observe supported firms up to five years since start-up and hence contribute long-term evidence to the literature. Moreover, we show in an extensive analysis with respect to effect heterogeneity for which subgroups of individuals such programs are most effective.

2.2. Institutional settings in Germany

Before heading to the empirical section, we discuss the institutional settings of start-up programs for the unemployed in Germany. The *bridging allowance*, introduced in 1986, remained the only program providing support to unemployed individuals who wanted to start their own business until 2003. Its main goal was to cover basic costs of living and social security contributions during the initial stages of self-employment. The recipient of BA received the same amount during the first six months he or she would have received if unemployed. Since the unemployment scheme also covers social security contributions (including health insurance, retirement insurance, etc.) a lump sum for social security is granted equal to 68.5% of the unemployment support that would have been received. Unemployed individuals were entitled to BA on condition of their business plan being externally approved, usually by the regional chamber of commerce. Thus, approval of an individual's application did not depend on the case manager at the local labor office. In January 2003, an additional program was initiated to support unemployed people in starting a new business. This *start-up subsidy* was introduced as part of a large package of ALMP programs introduced through the "Hartz reforms".[5] As was the case with BA, 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 comprises of a lump sum payment of €600 per 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 per month in the second year and to €240 per month in the third. In contrast to the BA, SUS recipients have to pay into the statutory pension fund and may claim a reduced rate for statutory health insurance. When the SUS was introduced in 2003, applicants did not have to submit business plans for prior approval, but they have been required to do so since November 2004. Moreover, parallel receipt of BA and SUS is excluded. The important features of both programs are summarized in Table 1. Moreover, it should be mentioned that other institutions such as federal state governments or the chamber of commerce offer general programs to encourage self-employment, for example, counseling, preparatory courses or even capital loans. Additionally, in some professions self-employment is highly restrictive in Germany when compared to other countries. For some "typical" self-employed occupations (physicians, lawyers, etc.) and several handcraft occupations it is required to occupy an advanced certificate in order to be allowed to become self-employed.

[4] O'Leary (1999) primarily attributes the negative earning effect in the case of Poland to firms' reluctance in full disclosure to the tax authorities.

[5] See Caliendo (2009) for an overview of the most relevant elements of the "Hartz reforms".

Table 1
Terms and conditions of programs.
Source: Social Act III, §57 — Bridging Allowance, §421l — Start-up Subsidy.

	Start-up subsidy	Bridging allowance
Entry conditions	-Unemployment benefit *receipt* -Income is restricted to €25,000 per year -Approval of a business plan was subsequently introduced in November 2004 -Below 65 years of age	-Unemployment benefit *entitlement* -No income restrictions -Approval of the business plan -Below 65 years of age
Support	-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	-Participants receive UB for six months -To cover social security liabilities, an additional lump sum of 68.5% is granted
Others	-Participants have to become a member in the state pension insurance and take advantage of a reduced rate in the legal health insurance	-Social security is left to the individual's discretion

However, Cressy (1996) argues that such preconditions for entry into self-employment tend to significantly enhance survival of businesses. In addition to entry restrictions, the German system prevents entrepreneurs from founding a company again if their business went bankrupt once.

Due to the institutional framework, it was rather rational to choose BA if unemployment benefits were fairly high or if the income generated through the start-up firm was expected to exceed €25,000 per year. Both programs were replaced in August 2006 by a single new program—the new start-up subsidy program (*Gründungszuschuss*)—which will not be analyzed here.[6] Table 2 contains number of entries into start-up programs as well as other programs of ALMP in West Germany. It is noticeable that start-up programs are comparable in terms of number of entries to other programs of ALMP, such as *wage subsidies* (WS) or *vocational training* (VT). On the other hand, entries into *training measures* (TM) are more than three times as much; but, of course, one has to keep in mind that TM are rather short-term, i.e., with a maximum duration of three months and an average duration of two weeks. Accordingly, entrance requirements are much lower. As we can see, the scope of the *new start-up subsidy* (New SUS) is below the cumulated number of entries in BA and SUS.

3. Data and empirical strategy

3.1. Data

In our analysis we focus on entries into SUS and BA in the third quarter of 2003[7] and combine administrative data from the "Federal Employment Agency" (FEA) with a survey such that longitudinal as well as cross-section data are available. We draw on data used by Baumgartner and Caliendo (2008) and extend it with an additional interview wave.[8] The administrative part consists of data based on the "Integrated Labour Market Biographies" (ILMB) of the FEA, containing relevant register data from four sources: employment history, unemployment support recipience, participation in active labor market measures, and job seeker history. Since the administrative data do not provide any information on self-employed individuals, the ILMB data are complemented by information from a computer-assisted telephone interview. For this, participants for each program who became self-employed in the third quarter of 2003 are randomly drawn. The comparison group is restricted to those who were unemployed in the third quarter of 2003, eligible to participate in either of the two programs, but did not join a program in this quarter. However, controls are allowed to participate in ALMP programs afterwards.[9] The first two interviews took place in January/February of 2005 and 2006; the final one was conducted in May/June of 2008. This enables us to follow individuals up to five years after start-up.

We restrict our sample to men in West Germany for two reasons. Men are more likely to become full-time self-employed than women; and West Germany is characterized by better labor market conditions compared to East Germany. We are interested in the effectiveness of start-up schemes to integrate former unemployed individuals in the labor market and we avoid several side-effects, such as labor supply decisions, macroeconomic constraints and so on, by excluding women and the East German labor market. Table 3 provides the number of realized interviews in the respective waves. In the case of SUS, for instance, we initially started with 1116 individuals who became self-employed in the third quarter of 2003; 811 responded to the second interview and finally we end up with 486 after the last interview in 2008. Hence, our final sample consists of 486 participants in SUS, 780 recipients of BA and 929 non-participants.

3.2. Descriptive evidence

Table A.1 in the Appendix provides descriptive statistics measured at entry into program in the third quarter of 2003 separately for participants (SUS and BA) and non-participants. Participants in SUS are on average younger and lower educated individuals with less employment duration and lower earnings in the past. This is in line with our expectations, as the financial support in case of BA depends on previous earnings and is only paid for a short period of six months. Hence, individuals with low earnings in the past are only eligible to minor support if they choose BA. It is therefore rational for those individuals to choose SUS because the subsidy is small but it might be extended up to three years. On the other hand, individuals with higher earnings want to secure their high entitlement and, consequently, choose BA. BA participants in our sample received on average €2056 per month and 89% of the SUS participants received the subsidy for three years. Moreover, in terms of location participants seem to be equally distributed throughout West Germany. As pointed out in previous research (e.g. Dunn and Holtz-Eakin, 2000), we find that self-employment is influenced by intergenerational transmission, i.e., the fraction with parental self-employment among participants is higher than among non-participants.

[6] The new start-up subsidy consists of unemployment benefits and a lump-sum payment of €300 per month for social coverage paid for nine months. Afterwards the lump-sum payment of €300 might be extended for further six months if the business is the full-time activity of the applicant. See Caliendo and Kritikos (2009) for information and a critical discussion of the features of the new program.

[7] Having access to only one particular quarter of entrants bears the risk of a selective sample. However, comparing the distribution of certain characteristics (e.g. age and educational background) across different quarters does not show any significant differences.

[8] Therefore, we only briefly discuss the basic data construction and refer to Baumgartner and Caliendo (2008) for a more extensive discussion of the data issues.

[9] The actual number of non-participants who participated in ALMP programs after the third quarter 2003 is rather low. Approximately 15% of all non-participants were assigned to programs of ALMP and only 2% participated in SUS or BA within our observation period.

Table 2
Entries into ALMP programs in West Germany (in thousands).
Source: Federal Employment Agency (various issues).

	BA	SUS	New SUS	VT	TM	WS
2000	92.6	–	–	337.9	285.9	120.4
2001	64.5	–	–	261.2	338.5	101.0
2002	89.0	–	–	273.2	545.4	114.4
2003	114.4	64.2	–	154.0	694.3	96.5
2004	137.3	113.8	–	124.0	788.5	93.9
2005	120.0	57.3	–	91.1	607.2	86.0
2006	83.6	27.0	25.4	173.0	671.1	152.1
2007	–	–	96.5	246.2	719.1	160.7

Note: BA — Bridging Allowance, SUS — Start-up Subsidy, VT — Vocational Training, TM — Training Measures, WS — Wage Subsidy.

Table 3
Number of realized interviews.

	Start-up subsidy	Bridging allowance	Non-participants
January/February 2005	1116	1665	2530
January/February 2006	811	1207	1448
May/June 2008	486	780	929

Note: See Baumgartner and Caliendo (2008) for more details on the construction of the data and detailed information with respect to the first and second interview waves. A minor part of the third wave interviews (4%) took place in July 2008.

In Table 4 we provide the labor market status of participants and non-participants after 28 and 56 months following start-up and the monthly income after 56 months. All descriptive results are weighted using *sequential inverse probability weighting* to adjust for the selection process due to panel attrition (see Wooldridge, 2002).[10] First of all, a closer look at the labor market developments of participants reveals that the fraction of self-employed individuals decreases from 71.5% to 67.9% for former BA recipients and from 67.6% to 59.7% for firms initially supported by SUS. Hence, the decline in self-employment is more than twice as high for SUS (−7.9 percentage points) than for BA (−3.6 percentage points) in the given period. This is mainly due to the fact that SUS expired between the second and third interview; whereas BA support had already stopped after six months, that was before the first interview took place. The sharp drop in self-employment rates after the end of the subsidy period may be seen as indication that some businesses were only able to survive with the help of the subsidy.

However, the main objective of ALMP is not primarily to create self-employment but to integrate unemployed individuals into the labor market. Hence, we now consider the share of individuals either in self-employment or regular employment. After 56 months since start-up, we find about 81% of SUS and 89% of former BA participants well integrated in the labor market. For non-participants, only 63% are either self-employed or regular employed. Hence, we observe a raw difference of employment rates of about 20% between participants and non-participants. These are descriptives only and the gap is potentially caused by differences in key variables. We return to this point in Section 3.3 when discussing the identification strategy and finally present causal effects of the programs in Section 4.1.

With respect to another objective of ALMP, the achievement of certain income levels for participants, we also provide in Table 4 net incomes (measured 56 months after start-up). Next to working income, the total income captures transfer payments such as

[10] As we can see in Table 3 the number of realized interviews decreased partly dramatically. On average, we are only able to observe 45% of all participants and 37% of non-participants for the entire period of 56 months. We checked our results with respect to potential selection process due to panel attrition and find positive selection, i.e., individuals who perform relatively well in terms of labor market outcomes are more likely to respond. Therefore, we use *sequential inverse probability weighting* to adjust for this selective attrition. Our matching results later on rely on unweighted outcome variables because participants and non-participants are similarly affected by selection due to panel attrition.

Table 4
Descriptive evidence on labor market status and income.

	Start-up subsidy	Bridging allowance	Non-participants
Labor market status			
2nd interview (January/February 2006)			
Self-employed	67.6	71.5	12.7
Regular employed	11.7	14.0	35.9
Unemployed or in ALMP	15.2	11.1	35.9
Others	5.6	3.4	15.5
3rd interview (May/June 2008)			
Self-employed	59.7	67.9	14.1
Regularly employed	20.9	21.1	49.1
Unemployed or in ALMP	11.7	6.7	19.9
Others	7.6	4.3	16.9
Income[a] at 3rd interview (May/June 2008)			
Total income	1672.0	2336.0	1581.1
	(1720.4)	(1962.9)	(1601.6)
	[1276.3]	[1942.3]	[1338.0]
Working income	1498.5	2167.4	1302.8
	(1780.2)	(2006.3)	(1662.5)
	[1145.3]	[1815.2]	[1190.1]
Household members	1.6	1.8	1.6
Equivalent income[b]	1678.2	2020.6	1458.4
	(1907.8)	(1809.4)	(1560.4)
	[1236.7]	[1602.6]	[1135.6]

Note: Numbers are percentages unless otherwise stated.
[a] Income is measured as average monthly net income in Euros: standard deviation and median are provided in parentheses and square brackets respectively.
[b] The equivalent income is calculated by adjusting the household income by the number of household members. The household income is divided by the weighted number of household members. Following the actual OECD equivalence scale, the household head achieves a weight of one, all children below the age of 15 are weighted with 0.3 and everybody else with 0.5 (see Whiteford and Adema, 2007). Since we only observe the total number of household members, every household member beside the household head receives a weight of 0.4.

unemployment benefit, pension, or child benefit and the equivalent income takes the number of household members into account.[11] We can see that former BA recipients have higher income in terms of working, total and equivalent income compared to SUS participants. This is not surprising because of the aforementioned selection into BA of highly educated individuals with high earnings in the past. It is also noticeable that non-participants earn on average less than participants; however considering the median of the income distribution, the difference to SUS participants almost vanishes.

Finally, to answer the question whether participants are more satisfied with their employment status compared to previous dependent employment, Table 5 provides some evidence on job satisfaction among participants who are self-employed at the third interview. The respondents were asked to compare their self-employment with the previous employment spell with respect to different aspects. Thereby, positive values indicate an overall improvement while negative values depict a decline. For participants in both programs, the situation improved in terms of type of activity, income and promotion prospects but declined for measures such as workload, working time and social security. However, the improvement among the first three measures is obviously more valued by individuals than the decrease in the latter because of higher absolute values.

3.3. Identification of causal effects

In order to estimate causal effects, we base our analysis on the potential outcome framework, also known as the Roy (1951) and

[11] The equivalent income is calculated by adjusting the household income by the number of household members. According to the actual OECD equivalence scale, the household head achieves a weight of one, all children below the age of 15 are weighted by 0.3 and everybody else with 0.5 (see Whiteford and Adema, 2007). Since we are only able to observe the total number of household members, we assign a weight of 0.4 to every household member beside the household head.

Table 5
Job satisfaction: Comparison to previous dependent employment.

	Start-up subsidy	Bridging allowance
Type of activity	0.6	0.5
Income	0.2	0.2
Promotion prospects	0.5	0.5
Workload	−0.1	−0.1
Working time	−0.2	−0.3
Social security	−0.2	−0.3

Note: Only self-employed individuals after 56 months since start-up. Scale: Improved (1), Unchanged (0), Declined (−1).

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 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 her potential outcomes: $\tau_i = Y_i^1 - Y_i^0$. Since 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:

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

The last term on the right hand side of Eq. (1) describes the hypothetical outcome without treatment for those individuals who received treatment. Since the condition $E(Y^0|D=1) = E(Y^0|D=0)$ is usually not satisfied with non-experimental data, estimating ATT by the difference in sub-population means of participants $E(Y^1|D=1)$ and 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 program due to observable or unobservable factors.[12] We apply propensity score matching and thus rely on the conditional independence assumption (CIA), which states that conditional on observable characteristics (W) the counterfactual outcome is independent of treatment: $Y^0 \amalg D|W$, where $\amalg$ denotes independence. In addition to the CIA, we also assume overlap: $Pr(D=1|W)<1$, for all W. 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 then be written as:

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

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 direct matching on W can become hazardous when W is of high dimension ("curse of dimensionality"), Rosenbaum and Rubin (1983) suggest using 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, that is, $W \amalg D|b(W)$. The propensity score $P(W)$, i.e., the probability of participating in a program, 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, the identifying assumption can be re-written as $Y^0 \amalg D|P(W)$ and the new overlap condition is given by $Pr(D=1|P(W))<1$.

The CIA is clearly a very strong assumption and the applicability of the matching estimator depends crucially on its plausibility. Blundell et al. (2005) argue that the plausibility of such an assumption should always be discussed on a case-by-case basis. Only variables which simultaneously influence the participation decision and the outcome variable 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; Sianesi, 2004). We use both administrative and survey data, which enables us to control for numerous individual information and labor market conditions. Based on this exhaustive data, we argue that the CIA holds in our application. However, we test the sensitivity of the results with respect to time-invariant unobserved differences between participants and non-participants by implementing *conditional difference-in-differences* (CDID). This allows for unobservable but temporally invariant differences in outcomes between participants and non-participants, which obviously relaxes the CIA. Conditional DID was initially suggested by Heckman et al. (1998). It extends the conventional DID estimator by defining outcomes conditional on the propensity score and using semiparametric methods to construct the differences. If the parameter of interest is ATT, the conditional DID 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:

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

For identification of causal effects, any general equilibrium effects need to be excluded, that is treatment participation of one individual cannot have an impact on outcomes of other individuals. This assumption is referred to as *stable-unit-treatment-value-assumption* (SUTVA). Imbens and Wooldridge (2009) argue that the validity of such an assumption depends on the scope of the program as well as on resulting effects. They infer that for the majority of labor market programs, the SUTVA is potentially fulfilled because such programs are usually of small scope with rather limited effects on the individual level. We follow their argumentation and refer to Table 2, where we see that entries into SUS and BA are approximately of the same scope as other ALMP programs and in relation to the total number of entries into unemployment of 5.5 million in 2004 quite small.

3.4. Estimation procedure

After having discussed identification issues, we proceed with the estimation of causal effects. We apply propensity score matching and estimate the propensity scores for participation in the respective program versus non-participation using *probit*-estimation. We test different specifications following economic theory and previous empirical findings as discussed above. But we also check econometric indicators such as significance of parameters or *pseudo*-R^2 to find the final specification.[13] The results of the *probit*-estimation can be found in Table A.2 in the Appendix. Let us briefly discuss the main components that influence the selection into treatment. In particular, variables such as age, duration of previous unemployment, regional cluster, information with respect to previous earnings and the intergenerational transmission turn out to be most important for the selection into SUS. In the case of "BA vs. NP", the duration of previous unemployment, indicators for the labor market history and also parental self-employment have a significant impact. This actually confirms our expectation that individuals with higher previous

[12] See, for example Caliendo and Hujer (2006) for further discussion.

[13] For a more extensive discussion on the estimation of propensity scores, we refer to Heckman et al. (1998) and Caliendo and Kopeinig (2008) among others.

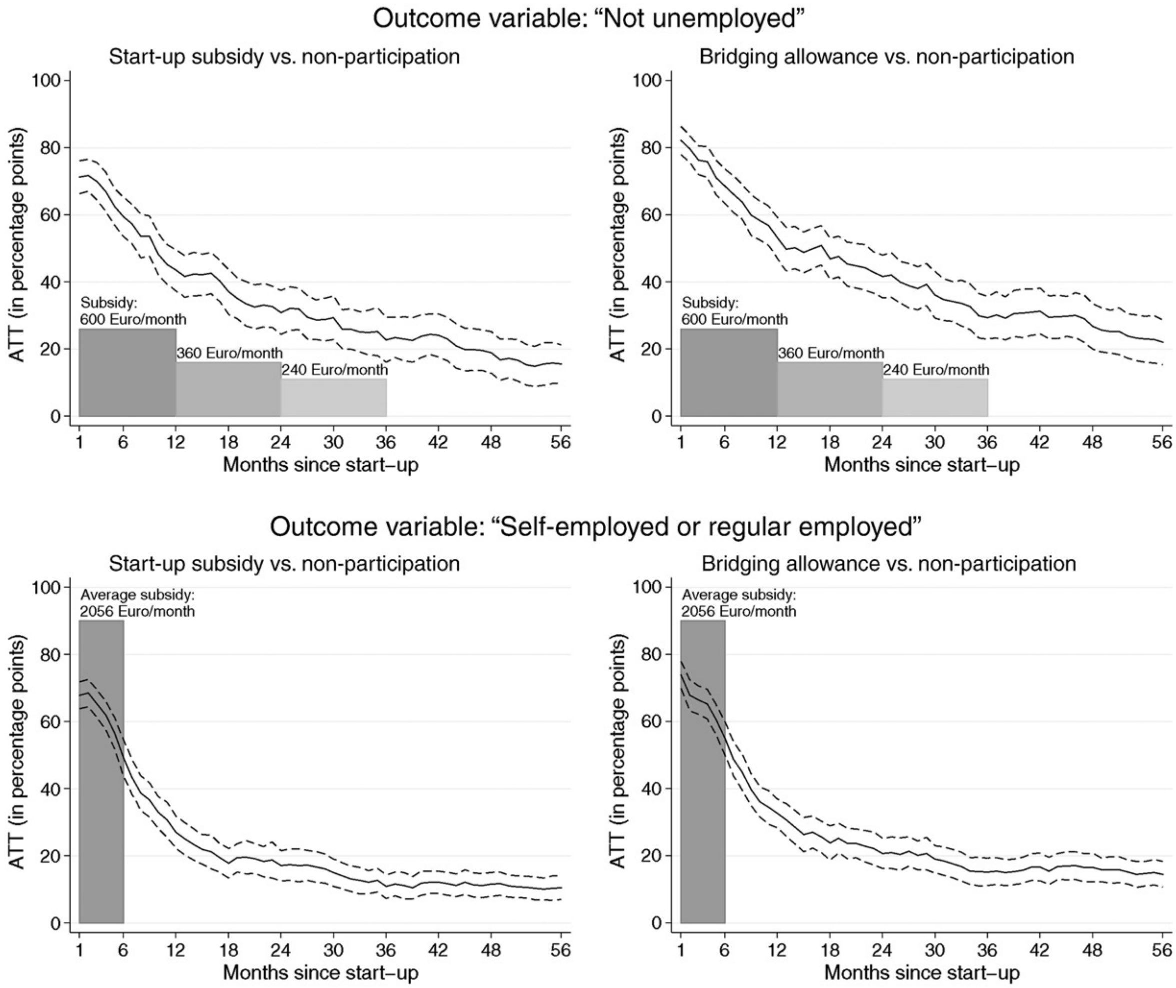

Fig. 1. Causal effects of start-up subsidy and bridging allowance over time.
Note: Depicted are average treatment effects on the treated (solid line), i.e., the difference in outcome variables between participants and non-participants. In addition, we provide 5% confidence intervals (dashed lines), which are based on *bootstrapped* standard errors with 200 replications. The duration and the amount of financial support are indicated by shaded bars. Due to institutional settings, the start-up subsidy amounted to €600/month, €360/month and €240/month in the first, second and third year; while the average subsidy in the case of bridging allowance was €2056/month paid for six months only. Thereby, the average subsidy is calculated by taking the average monthly unemployment benefit level (€40/day times 30.5 days) plus 68.5% for social security liabilities.

earnings are more likely to choose BA. In addition, we also provide the distribution of the estimated propensity scores in Fig. A.1 in the Appendix. As we can see, the distribution of the propensity scores are biased towards the tails, that is participants have a higher probability on overage of becoming self-employed than non-participants. Nevertheless, participant's propensity score distribution overlaps the region of the propensity scores of non-participants completely; therefore, the overlap assumption is fulfilled.

In the next step we estimate the average treatment effects on the treated as depicted in Eq. (2). In order to increase efficiency and being able to apply bootstrapping for inference we use a *kernel* matching algorithm.[14] To assess the matching quality, that is, whether the matching procedure balances the distribution of observable variables between participants and non-participants, Table A.3 summarizes different quality measures.[15] First of all, we provide in the upper part the number of variables which differ significantly between participants and non-participants by using a *t-test*.[16] For instance, we can see that for SUS, 28 variables have a mean that is significantly different between treated and non-treated at the 5% level before matching takes place. In the matched sample in turn, only two variables are significantly different for treated and non-treated individuals. In fact, in the case of BA after matching, we find no significant differences at all. This indicates that matching has been successful. Since using a *t-test* to assess the matching quality does not tell us anything about the bias reduction, we also provide the *mean standardized bias* (MSB) and the number of variables with a standardized bias of a certain amount. It can be seen that in case of "SUS vs. NP" ("BA vs. NP") the MSB declines from initially 14.6% to 3.5% (8.6% to 2.2%) after matching, where a MSB below 3% to 5% generally indicates a success of the matching approach (Caliendo and Kopeinig, 2008). Finally, we also re-estimate the propensity scores within the matched sample, as suggested by Sianesi (2004). The distribution of covariates should be well balanced within the matched sample and hence the resulting *pseudo-*R^2 from the propensity score estimation should be rather low. In fact, we do observe a sharp drop in *pseudo-*R^2 for both programs also suggesting a successful matching.

[14] More specifically, we apply an *Epanechnikov Kernel* with an bandwidth of *0.06*. We run different matching algorithm and find that our results are not sensitive. Furthermore, we applied *inverse probability weighting* (IPW) as an alternative approach for estimating ATT, as suggested by Imbens (2004). This method also relies on the CIA. Using IPW, we find hardly any substantial differences for the employment effects but slightly higher income effects.

[15] For a more intensive discussion with respect to assessing the matching quality, we refer to Caliendo and Kopeinig (2008).

[16] We consider the distribution of observable characteristics between participants and non-participants before and after matching based on 56 variables in total.

4. Results

The aim of the programs is to integrate unemployed individuals in the labor market and to increase income levels. Therefore, we use different outcome variables for the calculation of causal effects. We employ "not unemployed" and "self-employed or regular employed" as binary outcome variables to measure the degree of labor market integration. This is due to two reasons. First, non-participants are less likely to become self-employed than participants; and hence, comparing participants and non-participants with respect to self-employment only would bias the causal effects upwards. Second, the main objective of ALMP is to integrate individuals into the labor market which includes being regular employed as a success. We also want to highlight that being not registered as unemployed captures an upper bound estimation for the degree of labor market integration, i.e., independence of unemployment or social benefits. The binary outcome variables take on the value one if the individual is either "not unemployed" or "self-employed or regular employed" and zero otherwise.[17] Moreover, we examine whether program participation leads to an increase in income levels.

In the following, we first discuss the causal effects of SUS and BA with respect to the predefined outcome variables in Section 4.1. Afterwards, we consider effect heterogeneity in Section 4.2 to investigate for which subgroups both programs are in particular successful. Finally, Section 5 verifies the validity of our results with respect to unobserved heterogeneity.

4.1. Main results

Fig. 1 shows the average treatment effect on the treated as defined in Eq. (2) over time and Table 6 provides the corresponding exact values for selected points in time. As one can see in Fig. 1, the effects are positive and significant at all times for either outcome variable. To be precise, 56 months after start-up, participants in SUS (BA) have a 15.6% (10.6%) higher probability of not being registered as unemployed compared to non-participants. Regarding integration into the labor market, that is being either self-employed or regular employed, we detect that the employment probability of participants is 22.1 percentage points higher for SUS and 14.5 percentage points for BA participants in comparison to non-participants. These strong positive long-run effects are remarkable compared to findings of evaluation studies investigating other programs of ALMP in Germany, such as vocational training or job creation schemes.

Moreover, for BA the positive effect seems to be rather stable after three years following start-up, indicating that either surviving firms or employed individuals are well integrated in the (labor) market. For individuals supported with SUS, we do not find such a convergence. We argue that due to financial support which lasted longer, the adjustment process at the market is still ongoing. Because of this and the fact that the control group for BA participants is more competitive in the labor market than the assigned control group for SUS participants, the higher effects for SUS cannot be directly contrasted to the results of BA participants. In Table 6, we also provide the cumulated effects over time which reveal that within our observation period of 56 months, participants in SUS (BA) spent on average 23.5 (14.6) months more in self-employment or regular employment than non-participants. One may argue that cumulating the effects over the entire period will capture locking-in effects and lead to an overestimation of the effects, since participants received financial support. We take care of this by providing "partly" cumulated effects,

Table 6
Causal effects of start-up subsidy and bridging allowance.

	Start-up subsidy vs. non-participation	Bridging allowance vs. non-participation
Outcome variable: "Not unemployed"		
Difference in percentage points		
After 6 months	59.4 (3.0)	49.3 (2.8)
After 36 months	22.9 (3.4)	10.9 (1.8)
After 56 months	15.6 (2.9)	10.6 (1.8)
Difference in months		
Total cumulated effect ($\sum_{t=1}^{56}$)	18.7 (1.3)	12.2 (0.8)
Partly cumulated effect[a]	3.9 (0.6)	8.5 (0.7)
Outcome variable: "Self-employed or regular employed"		
Difference in percentage points		
After 6 months	68.5 (2.6)	55.0 (2.5)
After 36 months	29.4 (3.3)	15.3 (2.1)
After 56 months	22.1 (3.4)	14.5 (1.9)
Difference in months		
Total cumulated effect ($\sum_{t=1}^{56}$)	23.5 (1.3)	14.6 (0.9)
Partly cumulated effect[a]	5.5 (0.6)	10.8 (0.9)
Outcome variable: "Income 56 months after start-up"		
Difference in €/month		
Working income	435 (135)	618 (110)
Total income	270 (121)	485 (110)
Equivalent income[b]	248 (151)	546 (92)

Note: Depicted are average treatment effects on the treated as the difference in outcome variables between participants and non-participants. We define individuals who are neither registered as unemployed nor in a program of active labor market policy (except the two start-up subsidies) as being "not unemployed". Moreover, individuals who are either employed subject to social security contribution or self-employed are treated as "self-employed or regular employed". Standard errors are in parentheses and are based on *bootstrapping* with 200 replications.
[a] SUS: $\sum_{t=37}^{56}$, BA: $\sum_{t=7}^{56}$.
[b] See Table 4 for definition of equivalent income.

for which we cumulate the effects only over the period after financial support ended. For the case of SUS, we find that participants are still on average 5.5 months longer self-employed or regular employed than non-participants which actually depicts 20% of the post-program period of 20 months. For BA participants, we find a partly cumulated effect of 10.8 months, which is also 20% of the remaining period (of 50 months in this case).

To shed light on the question of income gains for participants, we provide the causal effects for income differences at the end of the observation period at the bottom of Table 6. We use three income-related outcome variables: The most relevant one is monthly net income from self-employment or paid employment (working income). However, since it is often argued that differences between (low) labor income and unemployment benefits are especially low in Germany, we will also look at the total personal income of individuals, that is, including transfer payments such as unemployment and child benefits. Finally, in order to take the household size into account we additionally calculate the effects on the equivalent income. The results unambiguously show that participants earn significantly more than non-participants. Participants in SUS (BA) have on average a net working income which is €435 (€618) higher per month than the one of non-participants at the end of our observation period. If we look at the total income participants still have a higher income than non-participants (€270 for SUS and €485 for BA). Finally, looking at the equivalent income also shows that participants in SUS (BA) earn on average €248 (€546) more than non-participants.

In summary, our results suggest that supporting unemployed individuals by SUS or BA has been a success in terms of both employment prospects as well as income measures compared to non-participation. The employment effects at the end of our observation period and cumulated over time are substantial and so are the income effects. Relating the working income effects to the average monthly net

[17] We define individuals who are neither registered as unemployed nor in a program of active labor market policy (except the two start-up subsidies) as being "not unemployed". Moreover, individuals who are either employed subject to social security contributions or self-employed are treated as "self-employed or regular employed".

working income of non-participants (compare Table 4) shows that these are economically very significant gains of around 28% to 39%.

4.2. Effects for subgroups: effect heterogeneity

In the following, we take a closer look on effect heterogeneity. This is in particularly insightful when determining the type of individuals who benefit most from participation. Disadvantaged groups in the labor market, such as low educated or young individuals, are likely to face limited job offers and the opportunity of becoming self-employed depicts a chance to escape unemployment. Additionally, self-employment might also be an alternative for individuals who are potentially discriminated in dependent employment, for example if their work is not valued high enough (see Clark and Drinkwater, 2000, for some evidence regarding ethnic minorities in the UK). We also have to take into account, that more educated unemployed individuals with past working experience have a relatively high probability of finding dependent employment again. Therefore, the distance between participants and matched non-participants in terms of labor market perspectives should be rather small. Taken together, this leads us to expect that the net effects of start-up programs (when compared to non-participation) are highest for disadvantaged individuals.

To answer the question of who benefits most, we conduct the complete estimation procedure, that is propensity score estimation and *kernel*-matching, for different subgroups of our sample with respect to educational attainment, professional qualification, age and nationality. The results are summarized in Table A.4 in the Appendix, in which the upper part depicts the effects for the whole sample.

First of all, consider the results stratified by educational attainment. We split the sample into high (completed upper secondary school) and low (no degree, lower or middle secondary school) educated individuals. It can be seen that low educated participants perform better in both programs in terms of employment prospects; the total cumulated effect is about 5 months larger than for high educated individuals. This is mainly driven by the fact that the control group of the highly educated have a higher probability of being employed at all times than the respective low educated comparison group. We illustrate that in Fig. A.2 by showing the levels for the outcome variable "self-employed or regular employed" among participants and non-participants within the matched sample; the difference between the respective solid and dashed line corresponds to the ATT presented in Table A.4. This confirms our expectation that the low educated control group performs relatively worse and consequently the effects are bigger for that group. Hence, offering individuals with bad labor market prospects the opportunity to turn unemployment into self-employment can be considered an effective strategy. The income effects in Table A.4 do not reveal such obvious patterns. In the case of "SUS vs. NP" the low educated participants yield much higher income effects compared to non-participation than the highly educated do. For the comparison "BA vs. NP" it is the reverse, that is the highly educated are better off than their low educated counterparts. This suggests that highly educated BA recipients who survived in self-employment are also very successful in terms of income. Furthermore, we conduct a separate analysis for different levels of professional qualification. Here we define all individuals with tertiary or technical college education as highly qualified; whilst skilled or unskilled workers are low qualified. As we can see in Table A.4 the effect pattern is very similar to the one of educational attainment (because professional qualification and educational attainment are highly correlated).

We also conduct the analysis separately for individuals aged 30 or younger as well as for individuals above the age of 30. Here, the employment effects of the two programs go in opposing directions. The results suggest that SUS tends to be more effective for participants above the age of 30; whereas BA seems to be more effective for younger participants. Fig. A.2 reveals that this is again mainly due to different labor market performance of the respective control groups. For both programs, there is hardly any difference between the program participants, that is the solid lines almost overlap. However, in the case of SUS controls, a considerable higher share of young controls is employed or self-employed and the reverse applies for BA. Probably more experienced (>30 years) BA controls are more likely to be employed or self-employed which seems reasonable given that BA attracts rather highly educated individuals with higher earnings in the past (see Section 3.1). Apparently, for these individuals experience is important in order to find a job in the labor market and therefore older BA control individuals perform better in the labor market. On the other hand, low educated individuals with bad labor market performance in the past (mainly attracted by SUS) have fewer opportunities in the labor market the older they are. The income effects are consistently higher for younger individuals. What has to be kept in mind here is that the matching quality for the younger cohorts (summarized by the mean standardized bias and the *pseudo*-R^2 after matching) is less satisfying and the same is true for SUS participants with high qualification. These groups are quite small making it harder to find suitable comparison individuals. Hence, the results have to be interpreted with caution.

Finally, we stratify the analysis with respect to German or non-German citizenship and find higher employment effects for natives. Fig. A.2 shows that the higher effects for natives are driven by the success of the participants. It can be seen that control groups do not really differ for both groups. This in turn suggests that SUS and BA seem to be even more effective for German participants. Additionally, natives achieve higher income effects even though they are not significant for the SUS case.

Fig. 2 exemplifies our findings with respect to effect heterogeneity and depicts the effects of program participation conditional on labor market perspectives without program participation. Therefore, we contrast cumulated average treatment effects for the outcome variable "self-employed or regular employed" (horizontal axis) to the average months spent in "self-employment or regular employment" among matched non-participants (vertical axis), which is supposed to reflect the labor market perspectives in case of non-participation. The scatter plot clearly indicates a negative relationship, underscoring the finding that groups with bad labor market perspectives benefit most. For instance, for individuals with high education/high qualification the estimated effects (horizontal axis) of the programs are rather small, however, for the opposite case—low education/low qualification—the effects are large. This suggests that SUS and BA are most effective for particular disadvantaged groups who face limited options in dependent employment. As previously mentioned, such groups are at high risk of becoming long-term unemployed; and therefore, these ALMP programs potentially contribute to the reduction of long-term unemployment amongst disadvantaged unemployed.

5. Sensitivity analysis

After having presented strong positive effects for both programs, we now need to check the robustness of our results with respect to deviations from the identifying assumption. If participants and non-participants differ in terms of unobserved characteristics, the CIA is violated and therefore our results would be biased. Since it is not possible to test the CIA directly with non-experimental data, we assess the sensitivity of our results in four different directions. First, we extend the set of variables in the propensity score estimation in order to see whether this has an impact on the causal estimates. Second, we allow for time-invariant unobserved differences between participants and non-participants and re-estimate the effects on employment and income. Third, we examine how strong an unobserved component would need to be in order to undermine the results from our analysis. Fourth, we estimate the effects for different sub-sets of the population where participants and non-participants are most comparable.

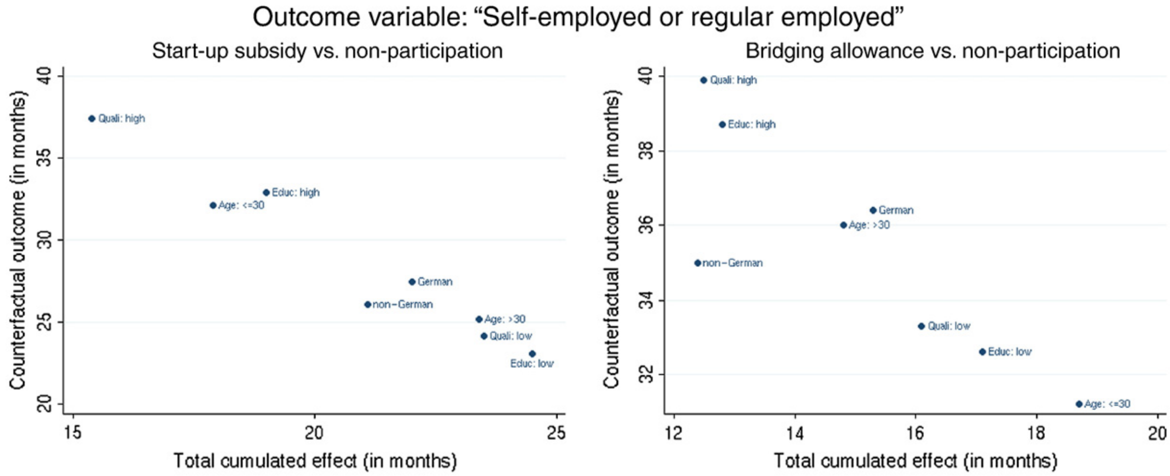

Fig. 2. Effect heterogeneity conditional on labor market perspectives among matched non-participants.
Note: Depicted on the horizontal axis are the cumulated average treatment effects on the treated consistent to Table A.4 for the outcome variable "self-employment or regular employment". On the vertical axis we provide the average months spent in "self-employment or regular employment" within the observation period of 56 months for the matched non-participants.

5.1. Extending the set of variables in the propensity score

Previous research has shown that entrepreneurs differ in various aspects from the general population. They are more likely to be male, higher educated and have self-employed parents. Clearly, this can also be true for our treatment groups and that is why we control for such characteristics in the propensity score estimation. However, there might still be personality traits which are not captured by the set of variables we control for. "Animal spirits" in the Schumpeterian sense will probably be more pronounced within the treatment group, even after controlling for observed characteristics and previous labor market experience.[18] One often cited and used proxy for such spirits are attitudes towards risk. The influence of risk aversion on the decision to become self-employed is a much discussed topic in the entrepreneurial literature. Conventional wisdom asserts that the role model of an entrepreneur requires to make risky decisions in uncertain environments and hence that more risk-averse individuals are less likely to become an entrepreneur. Caliendo et al. (2009) use experimentally-validated measures of risk attitudes in the most recent waves of the German Socio-Economic Panel (SOEP) to examine whether the decision of starting a business is influenced by objectively measurable risk attitudes at the time when this decision is made. The authors show that in general individuals with lower risk aversion are more likely to become self-employed.

In the second interview wave (28 months after start-up) of our data risk attitudes of participants and non-participants were elicited in a similar way as in the SOEP. Respondents were asked for attitudes towards risk in general and could indicate their willingness to take risks on an eleven-point scale ranging from zero (complete unwillingness) to ten (complete willingness). Table A.1 in the Appendix shows, that there are clear differences in the risk attitudes between participants and non-participants. Whereas participants have an average of 5.8, non-participants have an average of 5.5. Furthermore, 42% of the participants answer "7 or more" whereas this is only true for 33% of the non-participants.

Including this variable in the propensity score estimation is not without critique, since it was elicited 28 months after the decision to join the program and start a business. Hence, reverse causality might be an issue here, where the experience in the 28 months between starting the business and the interview taking place might have an influence on the attitudes towards risk. This is why we do not include risk attitudes in the final propensity score estimation in the previous section. However, most of the recent research (see, e.g., Dohmen et al., 2007) claims that risk attitudes are stable over time such that this might be less problematic. For the sensitivity analysis we have therefore included this variable in the propensity score estimation and replicated the full analysis. The variable is highly significant in the score estimation and we present the additional matching results in Panel A of Table 7 (employment effects) and Table 8 (income effects).[19]

Comparing the new results with the baseline results from before (compare with Table 6, also summarized at top of Table 7) we see that inclusion of the new variable "risk attitudes" lowers the effects slightly. For example, the effect on the outcome variable "self-employed or regular employed" after 56 months falls from 22.1% to 21.1% for SUS participants and the total cumulated effect goes down from 23.5 months to 23.4 months. For the BA participants the change is even smaller and slightly positive. Overall, we can conclude that adding this essential new variable "risk attitudes" does not change our results.

5.2. Conditional difference-in-differences

As already outlined in Section 3.3 we also test 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 (see Eq. (4)). For the outcome variables "not unemployed" and "self-employed or regular employed" we choose three different time periods for the comparison. In the first approach (CDID1) we use the time period from July 1998 to June 2003, that is, the five-year employment history before entering the program. 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 2.5 years (CDID2, January 2001–June 2003) as well as the earliest 2.5 years (CDID3, July 1998 to December 2000). For the CDID procedure with the income variables we use two reference levels: First, the average monthly income from regular employment in 2002 for the working income comparison and second, the total average monthly income in 2002 for the total income comparison.

Panel B in Tables 7 and 8 provides the cumulated employment effects and income effects for the conditional DID estimator. As we can see the results hardly differ from the matching estimates. For instance,

[18] We are grateful to one anonymous referee who mentioned this point.

[19] Full propensity score estimation results (and distributions) are available in the supplementary appendix.

Table 7
Sensitivity analysis — causal effects of start-up subsidy and bridging allowance: employment effects.

	Start-up subsidy vs. non-participation		Bridging allowance vs. non-participation	
	Outcome variables		Outcome variables	
	"Not UE"	"SE or RE"	"Not UE"	"SE or RE"
Main results (see Table 6)				
Effect after 56 months (in %-points)	15.6 (2.9)	22.1 (3.4)	10.6 (1.8)	14.5 (1.9)
Total cumulated effect ($\sum_{t=1}^{56}$)	18.7 (1.3)	23.5 (1.3)	12.2 (0.8)	14.6 (0.9)
Partly cumulated effect[a]	3.9 (0.6)	5.5 (0.6)	8.5 (0.7)	10.8 (0.9)
A) Alternative specification of the propensity score estimation				
Extended specification including risk attitudes				
Effect after 56 months (in %-points)	14.5 (3.2)	21.1 (3.4)	10.6 (1.8)	14.8 (2.1)
Total cumulated effect ($\sum_{t=1}^{56}$)	18.4 (1.2)	23.4 (1.3)	12.2 (0.8)	14.9 (0.9)
Partly cumulated effect[a]	3.7 (0.5)	5.3 (0.7)	8.5 (0.8)	11.0 (0.8)
B) Conditional difference-in-difference				
Total cumulated effect ($\sum_{t=1}^{56}$)				
CDID1	16.9 (1.5)	21.7 (1.4)	11.7 (0.7)	14.1 (0.9)
CDID2	17.7 (1.2)	22.6 (1.3)	12.2 (0.8)	14.6 (0.9)
CDID3	17.9 (1.3)	22.7 (1.4)	11.7 (0.7)	14.1 (0.9)
Partly cumulated effect[a]				
CDID1	2.1 (1.2)	3.7 (1.0)	8.0 (0.7)	10.2 (0.8)
CDID2	2.9 (0.7)	4.5 (0.8)	8.5 (0.7)	10.8 (0.8)
CDID3	3.1 (0.8)	4.6 (0.7)	8.0 (0.7)	10.2 (0.9)
C) Common support condition				
Thick support 1 — $0.33<\hat{P}(W)<0.67$				
Effect after 56 months (in %-points)	18.0 (4.0)	22.0 (4.4)	13.5 (2.1)	17.9 (2.7)
Total cumulated effect ($\sum_{t=1}^{56}$)	19.1 (1.5)	23.6 (1.6)	12.9 (0.9)	16.1 (1.1)
Partly cumulated effect[a]	4.0 (0.7)	5.5 (0.7)	9.2 (0.9)	12.2 (1.0)
Thick support 2 — $F(\hat{P}(W)>5\%)$				
Effect after 56 months (in %-points)	17.7 (2.7)	21.3 (3.3)	13.8 (1.7)	18.4 (2.1)
Total cumulated effect ($\sum_{t=1}^{56}$)	18.9 (1.1)	22.0 (1.1)	13.4 (0.8)	16.5 (1.0)
Partly cumulated effect[a]	4.0 (0.5)	4.9 (0.6)	9.8 (0.8)	12.6 (0.9)
Optimal subpopulation				
Effect after 56 months (in %-points)	15.0 (3.1)	21.1 (3.5)	11.1 (1.6)	15.3 (1.9)
Total cumulated effect ($\sum_{t=1}^{56}$)	17.9 (1.2)	22.9 (1.4)	12.4 (0.8)	14.9 (0.9)
Partly cumulated effect[a]	3.7 (0.6)	5.2 (0.6)	8.7 (0.7)	11.1 (0.9)

Note: Depicted are average treatment effects on the treated as the difference in outcome variables between participants and non-participants. Thereby the outcome variable "not unemployed" is depicted by "Not UE" and "self-employed or regular employed" by "SE or RE". All results are differences in months unless otherwise stated. Standard errors are in parentheses and are based on *bootstrapping* with 200 replications.
Alternative specification of the propensity score estimation: The extended specification contains risk attitudes in addition to the final specification.
Conditional difference-in-difference: The reference levels for the pre-treatment period are defined as follows: CDID1: July 1998–June 2003; CDID2: January 2001–June 2003; CDID3: July 1998–Dec. 2000.
Common support condition — thick support: We estimate the effects *(1)* in a region defined by $0.33<\hat{P}(W)<0.67$. Moreover, we divide the propensity score distribution into ten deciles and estimate the effects *(2)* only in regions where we have a density of at least 5% ($F(\hat{P}(W)>5\%)$) in both groups (participants and non-participants) respectively. A detailed Table with the distribution of participants and non-participants along the propensity score distribution is available in the supplementary appendix.
Common support condition — optimal subpopulation: The analysis is restricted to a subset of the original sample by dropping individuals with covariate values that are outside the optimal common support range (see Crump et al., 2009).
[a] SUS: $\sum_{t=37}^{56}$, BA: $\sum_{t=7}^{56}$.

Table 8
Sensitivity analysis — causal effects of start-up subsidy and bridging allowance: income effects.

	Start-up subsidy vs. non-participation		Bridging allowance vs. non-participation	
	Outcome variables: "Income 56 months after start-up"			
	Working income	Total income	Working income	Total income
Main results (see Table 6)	435 (135)	270 (121)	618 (110)	485 (110)
A) Alternative specification of the propensity score estimation Extended specification including risk attitudes	385 (153)	225 (149)	595 (117)	464 (118)
B) Conditional difference-in-difference	475 (130)	288 (139)	656 (128)	480 (128)
C) Common support condition				
Thick support 1 — $0.33<\hat{P}(W)<0.67$	226 (186)	114 (188)	588 (150)	468 (127)
Thick support 2 — $F(\hat{P}(W)>5\%)$	307 (179)	168 (151)	583 (129)	461 (123)
Optimal subpopulation	410 (137)	257 (153)	613 (118)	480 (105)

Note: Depicted are average treatment effects on the treated as the difference in outcome variables between participants and non-participants. All results are differences in €/month. Standard errors are in parentheses and are based on *bootstrapping* with 200 replications.
Alternative specification of the propensity score estimation: The extended specification contains risk attitudes in addition to the final specification.
Conditional difference-in-difference: The reference levels for the pre-treatment period are defined as follows: The working income is measured as the average monthly income from employment in 2002 and the total income as the average monthly total income in 2002.
Common support condition — thick support: We estimate the effects *(1)* in a region defined by $0.33<\hat{P}(W)<0.67$. Moreover, we divide the propensity score distribution into ten deciles and estimate the effects *(2)* only in regions where we have a density of at least 5% ($F(\hat{P}(W)>5\%)$) in both groups (participants and non-participants) respectively. A detailed Table with the distribution of participants and non-participants along the propensity score distribution is available in the supplementary appendix.
Common support condition — optimal subpopulation: The analysis is restricted to a subset of the original sample by dropping individuals with covariate values that are outside the optimal common support range (see Crump et al., 2009).

for the case of "BA vs. NP" we find participants being on average 14.6 months longer in employment or self-employment than non-participants using the total cumulated effect (cf. Table 6). Using conditional DID, the results vary from 14.1 to 14.6. The income effects are also very close to the matching results. This evidence indicates that controlling for time-invariant unobserved heterogeneity does not add essential information and consequently suggests that the CIA seems to be a reasonable assumption for our analysis.

5.3. Bounding and simulation analysis

Since it is not possible to test the CIA directly with non-experimental data; we now use a *bounding approach* initially suggested by Rosenbaum (2002). This approach consists of simulating an unobserved component and testing to which degree of unobserved heterogeneity results are robust. It should be clear that this approach does not answer the question whether or not the CIA is fulfilled but conveys information on the robustness of the results with respect to unobserved heterogeneity. The main idea is that in the presence of unobserved factors, identical individuals with respect to observable characteristics (W_i) have different probabilities of receiving treatment. Therefore, an artificial factor Γ is introduced to simulate an unobserved term. The underlying test statistic then tests up to which extent this unobserved factor Γ will influence the significance of the results (see Becker and Caliendo, 2007, for more details on the implementation of the test procedure and the STATA module *mhbounds.ado*).

We find strong positive effects for both programs and therefore we are only interested in the test-statistic for the upper bound under the assumption that we have overestimated the treatment effect. In other words, if unobserved factors lead to positive selection, i.e., those who participate always have a higher employment probability even in the absence of treatment, the test statistic Q^+ will become insignificant for a certain value of Γ. To ease the interpretation we also provide respective p-values (p^+).

Table A.5 summarizes test statistics separately for the outcome variables "not unemployed" and "self-employed or regular employed" and for "SUS vs. NP" and "BA vs. NP". We consider the outcome variables after 56 months since start-up.[20] Below the detailed test-statistics and respective p-values we provide the exact values of Γ at which results turn insignificant. First of all, in the case of the absence of unobserved heterogeneity, that is $\Gamma = 1.0$, we can see that the test statistic for the upper bounds are significant throughout, indicated by $p^+ < 0.05$. Starting from that point, we stepwise increase the value of Γ. As mentioned above, this actually simulates an ascending influence of unobserved factors. For the comparison "BA vs. NP" results are very robust against strong unobserved selection bias; up to $\Gamma = 3$ results remain significant. This implies that unobserved factors would need to have twice the influence (on selection and outcomes) as W_i in order to undermine the results. For the comparison "SUS vs. NP" on the other hand, results are more sensitive with critical values of 1.25/1.30 at the 1%-level and 1.40/1.45 at the 5%-level after 56 months. While this does not mean that there is unobserved heterogeneity influencing our results, this does call for a cautious interpretation of the results for SUS.

Since these critical values are rather abstract, we implement in addition a *simulation approach* as suggested by Ichino et al. (2008) to further investigate the influence of potential unobserved heterogeneity. The basic idea is to simulate an unobserved variable (or confounder) by adapting the distribution of an observable variable. Since we exactly know the influence of the observable characteristics on outcomes and selection we have a direct linkage to the potential unobserved leverage for the interpretation. The results are shown in Table A.6 in the Appendix where we concentrate on the effects on the outcome variable "self-employed or regular employed" after 56 months since start-up.[21] The first two columns show the effect of each confounder on the untreated outcome and on the selection into treatment. Thereby, a value below (above) one indicates a negative (positive) impact. The last column shows the resulting ATT given the existence of a confounder with a certain distribution. For instance, consider the effects for "SUS vs. NP" which are presented in the upper panel. In the absence of unobserved heterogeneity the impact on outcome and selection is zero and the ATT is 22 percentage points which is our baseline estimate from Table 6. If now an unobserved term is introduced which has the identical distribution as the age dummy "25–29 years", the influence on outcome (2.24) and on selection (1.52) would be positive. This means that such an unobserved term would have a positive effect on being "self-employed or regular employed" 56 months after start-up in case of no treatment and also on being treated at all. Including this simulated unobserved confounder leads to an ATT of 22 percentage points which is identical to the ATT in the absence of unobserved heterogeneity. We tested other confounders such as "upper secondary school", "duration of previous unemployment" and "parental self-employment". Even for an unobserved term associated with a strong positive effect on selection into treatment such as parental self-employment, the ATT hardly changes (to 21 percentage points). The finding that the ATT is always almost identical to the baseline effects confirms the robustness of our results with respect to unobserved heterogeneity.

5.4. Thick and optimal common support

The combined evidence of the sensitivity analysis so far suggests that the results are robust, but there may still remain concerns about any lingering selection on unobservables. Black and Smith (2004) show that such a lingering selection on unobservables will have its largest effects on bias for values of the propensity score in the tails of the distribution. This can be shown analytically (based on normality assumption of the joint error terms of the selection and outcome equations) but the underlying intuition is quite simple: when the probability of being in the treatment group is high, unobservable factors on average play a larger role than for probabilities near 0.5. This might lead to considerable selection bias if matching estimators must rely on the right tail of the distribution of propensity score in the comparison group. To deal with this, Black and Smith (2004) estimate the effects in a "thick support" region defined by $0.33 < \hat{P}(W) < 0.67$. We adopt their approach; additionally we divide the propensity score distribution into ten deciles and estimate the effects only in regions where we have a density of at least 5% in both groups respectively.[22] The results of both approaches are available in Tables 7 and 8 (Panel C). Our estimates based only on the "thick support" region of propensity scores around 0.5 are only slightly larger than those constructed using the full sample. The difference is a bit more pronounced for participants in BA where, e.g., the total cumulated effect on being self-employed or regular employed rises from 14.6 to 16.1 months. This difference could arise either from genuinely larger impacts in this region or lingering selection on unobservables which plays a bigger role outside the thick support region than within it. However, since the differences are quite small, lingering selection on unobservables does not seem to play a major role here. Using the second approach, i.e., restricting the analysis to regions where the propensity score density is above 5% for participants and non-participants, reduces the sample to the region $0.1 < \hat{P}(W) < 0.6$ for the SUS effects and $0.2 < \hat{P}(W) < 0.7$ for the BA effects. The results are also presented in Panel B and are very similar to the ones before.

[20] We also conducted the test for different points in time but the results hardly differ.

[21] Additional results are available on request from the authors.

[22] A detailed Table with the distribution of participants and non-participants along the propensity score distribution is available in the supplementary appendix.

Using the concept of "thick support" in this way means to restrict the propensity score distribution either arbitrarily or following a rule of thumb. Crump et al. (2009) suggest to base the common support decision rather on an objective measure. Restricting the propensity score distribution and hence excluding observations yields two opposing consequences for the variance term: while the variance increases due to the smaller sample size, the variance also decreases as participants with covariate values outside the range of the non-participants are excluded. They argue that the optimal common support is defined by balancing these two opposing variance components. To do so, we follow their approach and estimate the *optimal subpopulation average treatment effects* where we restrict the analysis to a subset of the original sample and drop individuals with covariate values that are outside the optimal common support range.[23] We do not find any significant differences to our main results.

6. Conclusion and outlook

In this paper, we analyze the effects of two programs designed to turn unemployment into self-employment in West Germany. The programs differ in their design and attract different types of persons. Individuals participating in the bridging allowance are more educated and have higher earnings in the past; whereas participants in the start-up subsidy are on average less educated and have a relatively poor previous labor market performance. Using an unique data set consisting of administrative and survey data, we are able to add two substantial aspects to previous literature: First of all, we observe individuals for nearly five years following start-up, such that we are able to provide first evidence on the long-term effects of these programs (especially for industrialized countries). Second, we carefully consider effect heterogeneity in order to determine for which groups programs work best.

We base our analysis on propensity score matching methods to assess the effectiveness of SUS and BA against non-participation. The identifying assumption is that conditional on the very informative data at hand selection into the programs can be assumed to be random such that outcome differences between participants and non-participants can be interpreted as causal effects. Since it has often been argued that individuals who participate in start-up programs and become self-employed have characteristics (observed and/or unobserved) which make them different from other unemployed individuals we need to carefully assess the sensitivity of our results with respect to deviations from the identifying assumption and we do so in four directions. First, we extend the set of variables used in the propensity score estimation—by incorporating risk attitudes which play a major role in entrepreneurial research—and show that this does not have an impact on the causal estimates. Second, we use the difference-in-difference analysis to show that accounting for time-invariant unobservable differences between participants and non-participants does not have an impact on the causal estimates. Third, we run a simulation analysis in order to show how strong any remaining unobserved differences need to be in order to undermine the results from our analysis. It turns out that results are very robust. Even if there would be an unobserved component having the strength of "parental self-employment" (which has a strong influence on selection and the outcome equation) the estimated effects hardly change. Fourth and final, we re-estimate the effects for subgroups of the propensity score distribution where we have a "thick overlap" between participants and non-participants. This is the region where any remaining unobserved heterogeneity should have the smallest effects. We show that the results inside and outside this region hardly differ. Overall, this makes us confident that the results are robust and not driven by any remaining unobserved heterogeneity.

Overall, we find persistent positive long-run effects of SUS and BA on the employment situation of former unemployed individuals. In particular, we use the probability of being employed (either self-employed or as an employee) and personal income as outcome variables. The results show that both programs are effective with respect to employment probabilities. Participants in SUS (BA) spend significant amounts of time longer in employment or self-employment than non-participants. Our results also unambiguously show that participants earn significantly more than non-participants. Additionally, self-employed participants are also more satisfied with their self-employment compared to previous dependent employment. Regarding effect heterogeneity, we estimate causal effects for different subgroups stratified by educational attainment, professional qualification, age and nationality. The results suggest that both programs are especially effective for individuals who are at high risk of being excluded from the labor market and becoming long-term unemployed like low educated and low qualified individuals. Following the concept of Sen (1997), SUS and BA helped to abolish labor market barriers for disadvantaged groups and sustainably integrated those into the labor market. Potentially, both programs are appropriate for fighting long-term unemployment, social exclusion and therefore poverty.

However, we also need to point out some limitations of our study and outline further research needed. First of all, it needs to be emphasized that we do a partial-equilibrium analysis focussing on the effects for participating individuals. Any general equilibrium or macroeconomic impacts cannot be considered in this setting. This is especially true for substitution effects and crowding-out effects. Hence, our positive findings (on an individual level) need to be verified on a macroeconomic level in order to judge the scope of the programs to generate any positive macro effects. Second, our estimation approach does not allow us to identify deadweight losses. The definition and identification of a deadweight loss in the context of start-up subsidies is—compared to other labor market policies such as wage subsidies—not straightforward. If an employer hires an unemployed person whose wage is subsidized but would have hired this unemployed anyway, we talk about a deadweight loss. In the context of the start-up subsidies this translates into the question whether the individuals would have founded the business even without a subsidy, and whether their success (or failure) would have had the same probabilities with and without the subsidy. Even if we know that people would have started without the subsidy, we are not able to answer the question whether the businesses would have been equally successful. A possible solution would be to compare subsidized start-ups out of unemployment with other start-ups. To do so, we need information on "regular" start-ups (unsubsidized, out of employment). This is one area of further research as is the question which of the two programs performs better. To analyze the latter, participants in both programs need to be compared directly which would also allow to reveal the scope of additional job creation by subsidized firms. Finally, investigating the effects for groups neglected here—women and the East German labor market—will allow to give a more complete picture of the performance of these two programs.

Acknowledgments

The authors thank Thomas Piketty, Paul Anand and two anonymous referees for helpful comments and suggestions. The paper has also benefited from comments and suggestions received in seminars and conferences at IZA, University of Kiel, University of Cracow, the 2009 IZA/Worldbank conference (Bonn), the 2009 SOLE meeting (Boston), the 2009 ESPE conference (Sevilla) and the 2009 AIEL conference (Sassari). Financial support of the IAB (Nuremberg) under the research grant No. 1007 is gratefully acknowledged. A supplementary appendix to this paper is available on request from the authors and can also be downloaded from http:/www.caliendo.de/Papers/sus_longterm_supplement.pdf.

[23] Restricting the estimation sample in such a way lowers external validity of the estimate, but probably enhances internal validity (Imbens and Wooldridge, 2009).

Appendix A

Table A.1
Selected descriptive statistics.

	Start-up subsidy	Bridging allowance	Non-participants
Number of observations[a]	472	756	853
Age (in years)	38.86	40.17	39.75
	(9.78)	(8.66)	(8.88)
Age bracket			
18 to 24 years	0.068	0.026	0.049
25 to 29 years	0.131	0.095	0.095
30 to 34 years	0.174	0.126	0.130
35 to 39 years	0.153	0.242	0.212
40 to 44 years	0.176	0.200	0.210
45 to 49 years	0.127	0.160	0.165
50 to 64 years	0.172	0.151	0.138
Marital status (Ref.: Single)			
Married	0.472	0.648	0.594
Number of children in household			
No children	0.708	0.595	0.639
1 child	0.144	0.155	0.145
2 or more children	0.148	0.250	0.216
Health restriction that affect job placement (Ref.: No)			
Yes	0.078	0.033	0.057
Nationality (Ref.: German)			
Non-German	0.328	0.265	0.249
Desired working time (Ref.: Part-time)			
Full-time	0.977	0.992	0.984
School achievement			
None	0.028	0.007	0.014
Lower secondary school	0.405	0.290	0.370
Middle secondary school	0.250	0.233	0.223
Specialized upper secondary school	0.104	0.193	0.150
Upper secondary school	0.214	0.278	0.243
Occupational group			
Manufacturing	0.040	0.011	0.018
Agriculture	0.333	0.233	0.277
Technical occupations	0.038	0.160	0.108
Services	0.517	0.565	0.539
Others	0.072	0.032	0.059
Professional qualification			
Workers with tertiary education	0.123	0.259	0.200
Workers with technical college education	0.068	0.112	0.106
Skilled workers	0.559	0.515	0.549
Unskilled workers	0.250	0.114	0.145
Duration of previous unemployment			
<1 month	0.133	0.074	0.014
≥1 month–<3 months	0.150	0.222	0.223
≥3 months–<6 months	0.212	0.249	0.251
≥6 months–<1 year	0.288	0.316	0.339
≥1 year–< 2 years	0.155	0.124	0.150
≥2 years	0.061	0.015	0.023
Professional experience (Ref.: Without professional experience)			
With professional experience	0.824	0.860	0.877
Duration of last employment (in months)	32.394	54.041	41.963
	(40.987)	(54.358)	(49.076)
Number of placement offers	5.367	3.758	5.181
	(8.563)	(6.921)	(7.664)
Daily income from regular employment in the first half of 2003 (in Euros)	9.969	25.783	20.700
	(21.571)	(41.503)	(34.970)
Unemployment benefit level (in Euros)	24.363	40.405	33.167
	(11.436)	(15.275)	(14.322)
Remaining unemployment benefit entitlement (in months)	4.752	7.317	7.054
	(5.759)	(6.380)	(6.397)
Employment status before job-seeking			
Employment	0.591	0.782	0.769
Self-employed	0.053	0.024	0.036
School attendance/never employed before/apprenticeship	0.123	0.073	0.063
Unemployable	0.083	0.042	0.059
Other, but employed at least once before	0.131	0.070	0.066
Other	0.019	0.009	0.007
Regional cluster[b]			
II a	0.013	0.024	0.028
II b	0.153	0.159	0.135
III a	0.127	0.071	0.088
III b	0.083	0.091	0.110
III c	0.222	0.237	0.244

Table A.1 (*continued*)

	Start-up subsidy	Bridging allowance	Non-participants
Regional cluster[b]			
IV	0.127	0.144	0.117
V a	0.036	0.042	0.038
V b	0.165	0.148	0.176
V c	0.074	0.083	0.064
Intergenerational transmission			
Parents are/were self-employed	0.284	0.284	0.155
General willingness to take risk[c] (Scale: 0 = complete unwillingness; 10 = complete willingness)			
Mean	5.816	5.847	5.490
	(2.177)	(2.071)	(2.011)
Share with risk attitude ≥ 7	0.419	0.427	0.329

Note: Numbers are percentages unless otherwise stated. Measured in the third quarter 2003; standard deviation in parentheses.
[a] Differences to realized interviews in Table 3 are due to missing information in the administrative data for some individuals.
[b] The regional clusters categorize German labor office districts with comparable local labor market characteristics (see Blien et al., 2004). For instance, the category IIa contains urban districts with relatively high unemployment rates, IIIc primarily rural areas with below-average unemployment rates and few dynamic, while the category Vc captures districts characterized by favorable labor market conditions and high dynamic.
[c] Measured at the second interview, i.e., 28 months after start-up.

Table A.2
Propensity score estimation.

	Start-up subsidy vs. non-participation	Bridging allowance vs. non-participation
Age bracket (Ref.: 18 to 24 years)		
25 to 29 years	0.430**	0.354*
30 to 34 years	0.508**	0.254
35 to 39 years	0.266	0.291
40 to 44 years	0.361*	0.119
45 to 49 years	0.433**	0.196
50 to 64 years	0.863***	0.316
Marital status (Ref.: Single		
Married	−0.098	0.009
Number of children in household (Ref.: No children)		
One child	0.184	−0.105
Two or more children	0.089	−0.160
Health restriction that affects job placement (Ref.: No)		
Yes	−0.129	−0.090
Nationality (Ref.: German)		
Non-German	0.095	0.164**
Desired working time (Ref.: Part-time)		
Full-time	−0.037	0.135
School achievement (Ref.: None)		
Lower secondary school	−0.081	0.228
Middle secondary school	0.069	0.293
Specialized upper secondary school	−0.063	0.333
Upper secondary school	0.038	0.288
Occupational group (Ref.: Manufacturing)		
Agriculture	−0.250	0.100
Technical occupations	−0.705**	0.261
Services	−0.395	0.089
Others	−0.597**	−0.342
Professional qualification (Ref.: Workers with tertiary education)		
Workers with technical college education	0.126	−0.038
Skilled workers	0.071	0.042
Unskilled workers	0.198	0.066
Duration of previous unemployment (Ref.: <1 month)		
≥1 month–<3 months	−1.634***	−0.893***
≥3 months–<6 months	−1.488***	−0.943***
≥6 months–<1 year	−1.639***	−1.069***
≥1 year–<2 years	−1.765***	−1.118***
≥2 years	−1.316***	−1.145***
Professional experience (Ref.: without professional experience)		
with professional experience	−0.123	−0.251**
Duration of last employment (in months)	0.001	0.002**
Number of placement offers	−0.006	−0.010**
Remaining unemployment benefit entitlement (in months)	−0.028***	−0.024***
Unemployment benefit level (in Euros)	−0.029***	0.024***
Daily income from regular employment in the first half of 2003 (in Euros)	−0.002	−0.002*
Employment status before job-seeking (Ref.: Employment)		
Self-employed	0.290	−0.373*
School attendance/never employed before/apprenticeship	0.362**	0.225
Unemployable	0.197	−0.072
Other, but employed at least once before	0.458***	0.246*
Other	0.352	0.456

(*continued on next page*)

Table A.2 (*continued*)

	Start-up subsidy vs. non-participation	Bridging allowance vs. non-participation
Regional cluster (Ref.: II a)		
II b	0.730**	0.224
III a	0.744**	0.043
III b	0.545*	0.157
III c	0.609*	0.118
IV	0.911***	0.183
V a	0.636*	0.415
V b	0.707**	−0.041
V c	0.782**	0.262
Intergenerational transmission		
Parents are/were self-employed	0.476***	0.453***
Constant	1.240**	−0.607
Number of observations		
Participants	472	756
Non-participants	853	853
Hit-rate (share of correct predictions in %)	70.79	65.26
Pseudo-R^2	0.196	0.105
Log-likelihood	−693.612	−995.964

Note: * 10%, ** 5%, *** 1% significance level.

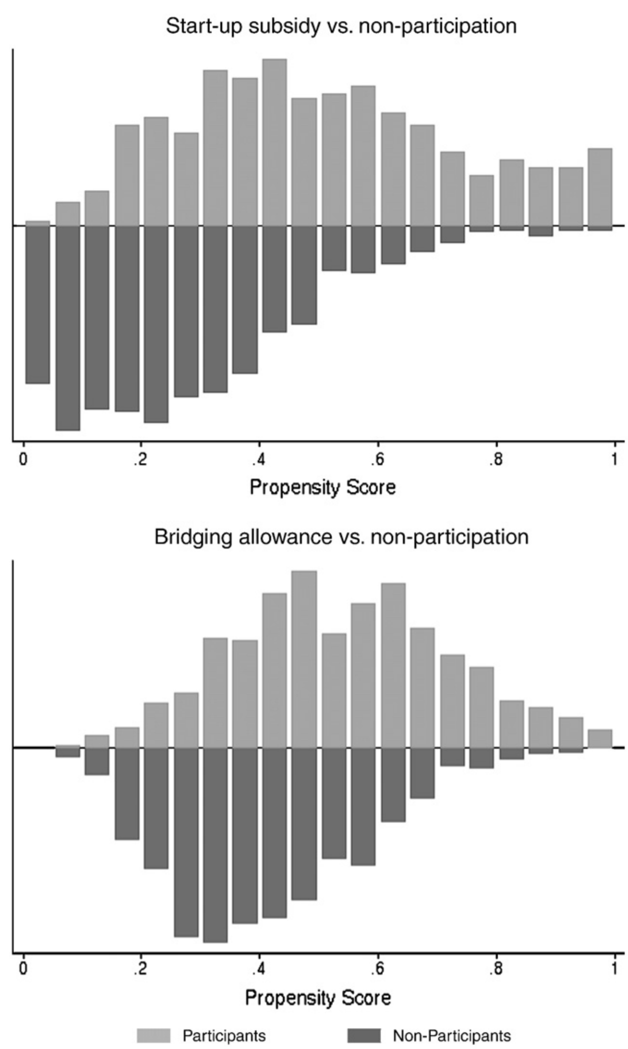

Fig. A.1. Propensity score distributions.
Note: These are propensity score distributions for participants and Non-participants based on estimations in Table A.2.

Table A.3
Matching quality.

	Start-up subsidy		Bridging allowance	
	Before matching	After matching	Before matching	After matching
t-test of equal means[a]				
1%-level	19	0	9	0
5%-level	28	0	15	0
10%-level	33	0	17	0
Standardized bias				
Mean standardized bias	14.550	3.539	8.565	2.194
Number of variables with standardized bias of a certain amount				
< 1%	2	12	3	22
1% until < 3%	4	14	11	18
3% until < 5%	4	14	6	7
5% until < 10%	15	15	21	9
≥ 10%	31	1	15	0
Pseudo-R^2	0.196	0.013	0.105	0.007

[a] Depicted is the number of variables which differ significantly between treated and controls. The decision is based on a simple t-test of equal means. There are 56 observable variables in total.

Table A.4
Effect heterogeneity: Causal effects of start-up subsidy and bridging allowance.

	Start-up subsidy vs. non-participation		Bridging allowance vs. non-participation	
Main results				
# Treated	472		756	
# Controls	853		853	
Outcome variable: "Self-employed or regular employed"				
After 36 months (in %-points)	29.4		15.3	
After 56 months (in %-points)	22.1		14.5	
Total cumulated effect ($\sum_{t=1}^{56}$, in months)	23.5		14.6	
Outcome variable: "Income 56 months after start-up" (in €/month)				
Working income	435		618	
Equivalent income[a]	(248)		546	
Matching quality[b]				
Mean standardized bias	3.539		2.194	
Pseudo-R^2	0.013		0.007	
	Educational level			
	Low	High	Low	High
# Treated	322	150	400	356
# Controls	518	335	518	335
Outcome variable: "Self-employed or regular employed"				
After 36 months (in %-points)	29.6	25.5	20.0	10.6
After 56 months (in %-points)	23.7	17.6	19.2	11.7
Total cumulated effect ($\sum_{t=1}^{56}$, in months)	24.5	19.0	17.1	12.8
Outcome variable: "Income 56 months after start-up" (in €/month)				
Working income	616	(−100)	416	768
Equivalent income[a]	(328)	(−23)	286	732
Matching quality[b]				
Mean standardized bias	3.753	7.393	2.244	3.375
Pseudo-R^2	0.015	0.059	0.007	0.013
	Professional qualification			
	Low	High	Low	High
# Treated	382	90	475	281
# Controls	592	261	592	261
"Outcome variable: Self-employed or regular employed"				
After 36 months (in %-points)	27.3	16.3	15.8	12.7
After 56 months (in %-points)	20.5	11.5	17.1	12.4
Total cumulated effect ($\sum_{t=1}^{56}$, in months)	23.5	15.4	16.1	12.5
Outcome variable: "Income 56 months after start-up" (in €/month)				
Working income	628	−464	486	865
Equivalent income[a]	353	(−189)	439	725
Matching quality[b]				
Mean standardized bias	4.145	14.048	2.822	4.166
Pseudo-R^2	0.019	0.000	0.008	0.020

(continued on next page)

Table A.4 (*continued*)

	Start-up subsidy vs. non-participation		Bridging allowance vs. non-participation	
	Age			
	≤30	>30	≤30	>30
# Treated	112	360	110	646
# Controls	141	712	141	712
Outcome variable: "Self-employed or regular employed"				
After 36 months (in %-points)	21.9	27.0	20.1	15.7
After 56 months (in %-points)	(8.7)	21.3	10.5	16.2
Total cumulated effect ($\sum_{t=1}^{56}$, in months)	17.9	23.4	18.7	14.8
Outcome variable: "Income 56 months after start-up" (in €/month)				
Working income	543	374	914	573
Equivalent income[a]	506	(242)	761	525
Matching quality[b]				
Mean standardized bias	9.968	3.740	14.308	2.492
Pseudo-R^2	0.000	0.017	0.000	0.008
	Nationality			
	Native	Non-German	Native	Non-German
# Treated	317	155	556	200
# Controls	641	212	641	261
Outcome variable: "Self-employed or regular employed"				
After 36 months (in %-points)	27.3	20.6	15.9	10.6
After 56 months (in %-points)	20.0	15.7	14.2	14.5
Total cumulated effect ($\sum_{t=1}^{56}$, in months)	22.0	21.1	15.3	12.4
Outcome variable: "Income 56 months after start-up" (in €/month)				
Working income	(305)	(249)	612	587
Equivalent income[a]	(147)	(339)	547	543
Matching quality[b]				
Mean standardized bias	3.424	11.871	2.197	5.202
Pseudo-R^2	0.016	0.000	0.007	0.032

Note: Depicted are average treatment effects on the treated as the difference in outcome variables between participants and non-participants. The educational level is decomposed into "high" education, capturing individuals who have successfully completed upper secondary school, and "low" education, including individuals who have either not completed school or have completed lower or middle secondary school. With respect to professional qualifications we define individuals with tertiary or technical college education as "highly" qualified, while skilled or unskilled workers are categorized as "low" qualified. Effects which are not significant different from zero at the 5%-level are in parentheses; standard errors are based on bootstrapping with 200 replications.

[a] See Table 4 for definition of equivalent income.

[b] The matching quality indicators are measured after matching. More detailed statistics are available in the supplementary appendix.

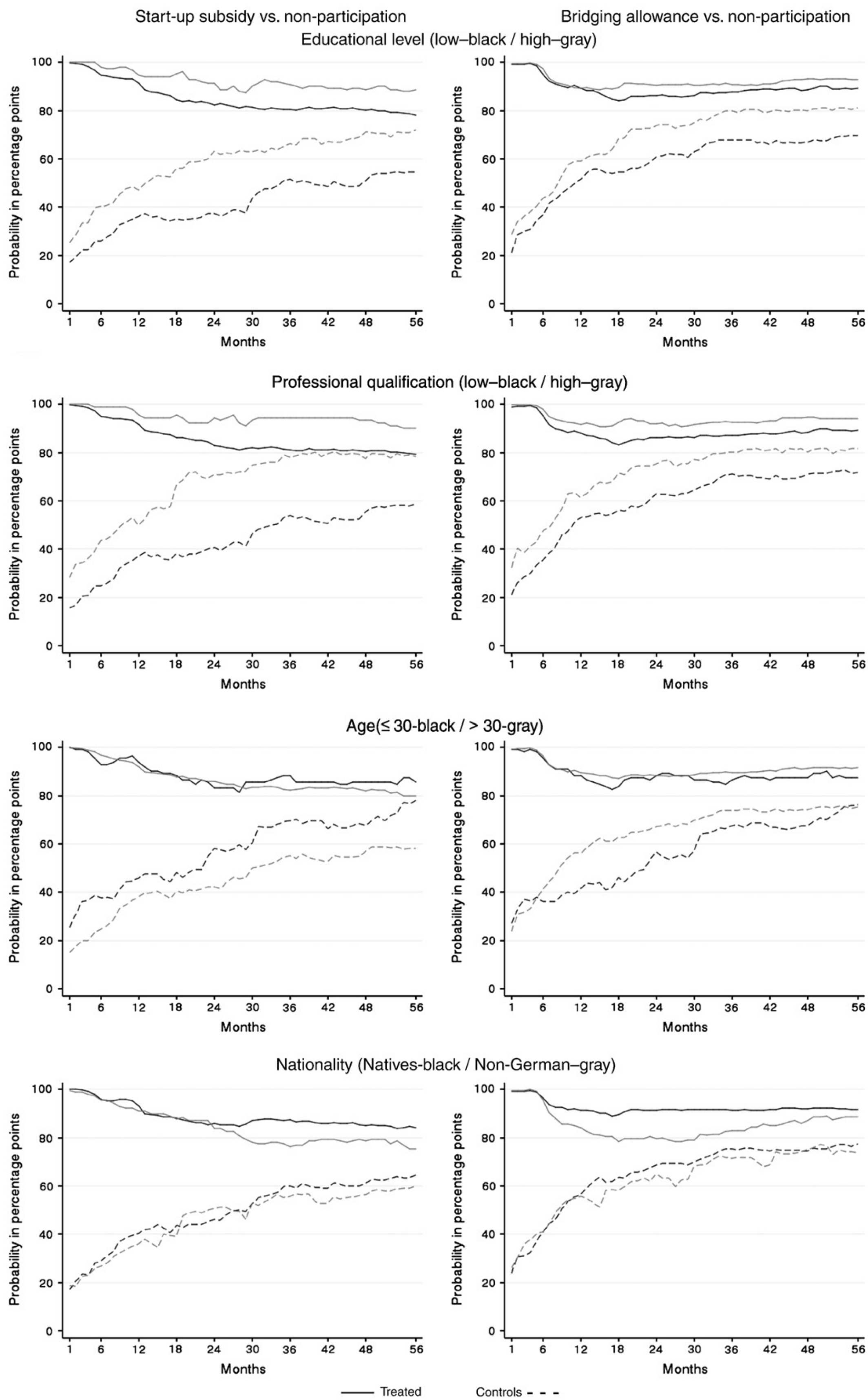

Fig. A.2. Effect heterogeneity: Probability levels among participants and matched non-participants.

Note: Depicted are probability levels for the outcome variable "self-employment or regular employment" among participants and non-participants within the matched sample, i.e., the difference between the solid and dashed line is the average treatment effect on the treated. For instance, consider the case of start-up subsidy vs. non-participation on the left panel. After 56 months 88.7% (72.1%) of the highly educated participants (matched non-participants) are in self-employment or regular employment; while only 78.0% (54.5%) of the low educated participants (matched non-participants) are either self-employed or regular employed.

Table A.5
Sensitivity to unobserved heterogeneity — bounding approach.

Γ	Outcome variable: not unemployed				Outcome variable: self-employed or regular employed			
	SUS vs. NP		BA vs. NP		SUS vs. NP		BA vs. NP	
	Q^+	p^+	Q^+	p^+	Q^+	p^+	Q^+	p^+
After 56 months since start-up								
1.00	3.721	0.000	8.596	0.000	4.473	0.000	10.332	0.000
1.25	2.355	0.009	7.163	0.000	2.862	0.002	8.616	0.000
1.50	1.252	0.105	6.034	0.000	1.558	0.060	7.254	0.000
1.75	0.325	0.372	5.107	0.000	0.460	0.323	6.128	0.000
2.00	0.307	0.379	4.321	0.000	0.346	0.364	5.169	0.000
Critical values								
1%	1.25–1.30		2.80–2.85		1.30–1.35		3.00–3.05	
5%	1.40–1.45		3.20–3.25		1.45–1.50		3.30–3.35	
10%	1.45–1.50		3.35–3.40		1.55–1.60		3.45–3.50	

Note: Reported results are achieved by using mhbounds.ado (see Becker and Caliendo, 2007). Critical values are related to the exact values of Γ at which results turn insignificant. BA — Bridging Allowance, SUS — Start-up Subsidy, NP — Non-Participation.

Table A.6
Sensitivity to unobserved heterogeneity — simulation approach.

Confounder	Influence of unobserved confounder on		ATT (S.E.)
	Outcome	Selection	
Start-up subsidy vs. non-participation			
No unobserved heterogeneity (see Table 6)	0.00	0.00	0.22(0.04)
Confounder with an influence like (see Table A.2)			
Age bracket (25–29 years)	2.24	1.52	0.22(0.01)
Upper secondary school	2.28	0.76	0.23(0.01)
Duration of previous unemployment (1 month–<3 months)	1.50	0.65	0.22(0.01)
Parents are/were self-employed	1.66	2.19	0.21(0.01)
Bridging allowance vs. non-participation			
No unobserved heterogeneity (see Table 6)	0.00	0.00	0.14(0.02)
Confounder with an influence like (see Table A.2)			
Age bracket (25–29 years)	2.18	1.01	0.15(0.00)
Upper secondary school	2.34	1.37	0.14(0.00)
Duration of previous unemployment (1 month–<3 months)	1.45	1.02	0.14(0.00)
Parents are/were self-employed	1.63	2.19	0.14(0.01)

Note: Reported results are achieved by using sensatt.ado (see Nannicini, 2007) and are related to the binary outcome variable "self-employment or regular employment" measured 56 months after start-up. The first two columns show the effect of an unobserved confounder distributed like particular observable confounders on the untreated outcome and on the selection into treatment. Thereby, a value below (above) one indicates a negative (positive) impact. In case of no unobserved heterogeneity, the unobserved term is excluded and both impacts are zero.

Appendix B. Supplementary results

Supplementary results to this article can be found online at doi:10.1016/j.jpubeco.2010.11.003.

References

Almeida, R., Galasso, E., 2007. Jump-Starting Self-Employment? Evidence Among Welfare Participants in Argentina. IZA. Discussion Paper 2902.
Baumgartner, H., Caliendo, M., 2008. Turning unemployment into self-employment: effectiveness of two start-up programmes. Oxford Bulletin of Economics and Statistics 70 (3), 347–373.
Becker, S., Caliendo, M., 2007. Sensitivity analysis for average treatment effects. Stata Journal 7 (1), 71–83.
Betcherman, G., Olivas, K., Dar, A., 2004. Impacts of active labor market programs: new evidence from evaluations with particular attention to developing and transition countries. Social Protection Discussion Paper Series No. 0402. The World Bank.
Black, D., Smith, J., 2004. How robust is the evidence on the effects of the college quality? Evidence from matching. Journal of Econometrics 121 (1), 99–124.
Blien, U., Hirschenauer, F., Arendt, M., Braun, H.J., Gunst, D.-M., Kilcioglu, S., Kleinschmidt, H., Musati, M., Roß, H., Vollkommer, D., Wein, J., 2004. Typisierung von Bezirken der Agenturen der Arbeit. Zeitschrift für Arbeitsmarktforschung 37 (2), 146–175.
Blundell, R., Dearden, L., 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 168 (3), 473–512.
Caliendo, M., 2009. Income Support Systems, Labor Market Policies and Labor Supply: The German Experience. IZA. Discussion Paper 4665.
Caliendo, M., Fossen, F., Kritikos, A., 2009. Risk attitudes of nascent entrepreneurs: new evidence from an experimentally-validated survey. Small Business Economics 32 (2), 153–167.
Caliendo, M., Hujer, R., 2006. The microeconometric estimation of treatment effects — an overview. Allgemeines Statistisches Archiv 90 (1), 197–212.
Caliendo, M., Hujer, R., Thomsen, S., 2008. The employment effects of job creation schemes in germany — a microeconometric evaluation. In: Millimet, D.L., Smith, J.A., Vytlacil, E. (Eds.), Modelling and Evaluating Treatment Effects in Econometrics.: Advances in Econometrics, vol.21. Elsevier, Amsterdam, pp. 381–428.
Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys 22 (1), 31–72.
Caliendo, M., Kritikos, A., 2009. Die reformierte Existenzgründungsförderung für Arbeitslose — Chancen und Risiken. Perspektiven der Wirtschaftspolitik 10 (2), 189–213.
Card, D., Kluve, J., Weber, A., 2010. Active labour market policy evaluations: a meta-analysis. The Economic Journal 120 (548), F452–F477.
Carling, K., Gustafson, L., 1999. Self-employment grants vs. subsidized employment: is there a difference in the re-unemployment risk? IFAU — Institute for Labour Market Policy Evaluation. Working Paper 1999:6.
Clark, K., Drinkwater, S., 2000. Pushed out or pulled in? Self-employment among ethnic minorities in England and Wales. Labour Economics 7, 603–628.
Cressy, R., 1996. Are business startups debt-rationed? The Economic Journal 106 (438), 1253–1270.
Crump, R., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. Biometrika 96 (1), 187–199.
Cueto, B., Mato, J., 2006. An analysis of self-employment subsidies with duration models. Applied Economics 38, 23–32.
Dar, A., Gill, I.S., 1998. Evaluating retraining programs in OECD countries: lessons learned. The World Bank Research Observer 13 (1), 79–101.
Dar, A., Tzannatos, Z., 1999. Active Labor Market Programs: A Review of the Evidence from Evaluations. The World Bank. SP Discussion Paper 9901.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2007. Risk as a Personality Trait: On the Stability of Risk Attitudes. Discussion paper, Berlin.
Dunn, T., Holtz-Eakin, D., 2000. Financial capital, human capital, and the transition to self-employment: evidence from intergenerational links. Journal of Labor Economics 18 (2), 282–305.
Fay, R., 1996. Enhancing the effectiveness of active labor market policies: evidence from programme evaluations in OECD countries. Labour Market and Social Policy Occasional Papers, 18. OECD.
Federal Employment Agency (various issues): *Arbeitsmarkt*. Nuremberg.
Fitzenberger, B., Osikominu, A., Völter, R., 2008. Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany. Annales d'Èconomie et de Statistique 91–92, 321–355.
Fritsch, M., 2008. How does new business development affect regional development? Introduction to the special issue. Small Business Economics 30, 1–14.
Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. Econometrica 66 (5), 1017–1098.
Heckman, J., LaLonde, R., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. IIIA. Elsevier, Amsterdam, pp. 1865–2097.
Ichino, A., Mealli, F., Nannicini, T., 2008. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity. Journal of Applied Econometrics 23, 305–327.
Imbens, G., 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. The Review of Economics and Statistics 86 (1), 4–29.
Imbens, G., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47 (1), 5–86.
Kelly, R., Lewis, P., Mulvey, C., Dalzell, B., 2002. A Study to Better Assess the Outcomes in the New Enterprise Incentive Scheme: Report Prepared for the Department of Employment and Work Place Relations. Centre for Labour Market Research, University of Western Australia.
Kluve, J., Schmidt, C.M., 2002. Can training and employment subsidies combat european unemploment? Economic Policy 17 (35), 409–448.
Lechner, M., Miquel, R., Wunsch, C., 2004. Long-Run Effects of Public Sector Sponsored Training in West Germany. IZA, Bonn. Discussion Paper No. 1443.
Lechner, M., Wunsch, C., 2008. What did all the money do? On the general ineffectiveness of recent West German labour market programmes. Kyklos 61 (1), 134–174.
Martin, P., Grubb, D., 2001. What works and for whom: a review of OECD countries experiences with active labour market policies. Swedish Economic Policy Review 8, 9–56.
Meager, M., Bates, P., Cowling, M., 2003. An evaluation of business start-up support for young people. National Institute Economic Review 186, 59–72.
Meager, N., 1993. Self-employment and Labour Market Policy in the European Community. WZB. Discussion Paper FS I 93–201.
Nannicini, T., 2007. A simulation-based sensitivity analysis for matching estimators. Stata Journal 7 (3), 334–350.
OECD, 2008. Growing unequal? Income distribution and poverty in OECD countries. OECD.
OECD, 2009. Employment Outlook: Tackling the Jobs Crisis. Paris.

O'Leary, C.J., 1999. Promoting Self Employment Among the Unemployed in Hungary and Poland. W.E. Upjohn Institute for Employment Research. Working Paper.
Perry, G., 2006. Are Business Start-Up Subsidies Effective for the Unemployed: Evaluation of Enterprise Allowance. Auckland University of Technology. Working paper.
Pfeiffer, F., Reize, F., 2000. Business start-ups by the unemployed — an econometric analysis based on firm data. Labour Economics 7, 629–663.
Rodriguez-Planas, N., 2008. Channels through which public employment services and small-business assistance programs work. Oxford Bulletin of Economics and Statistics 72 (4), 458–485.
Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–50.
Rosenbaum, P.R., 2002. Observational Studies. Springer, New York.
Roy, A., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers 3 (2), 135–145.
Rubin, D., 1974. Estimating causal effects to treatments in randomised and nonrandomised studies. Journal of Educational Psychology 66, 688–701.
Sen, A., 1997. Inequality, unemploment and contemporary Europe. International Labour Review 136 (2), 155–172.
Sianesi, B., 2004. An evaluation of the Swedish system of active labour market programmes in the 1990s. The Review of Economics and Statistics 86 (1), 133–155.
Smith, J., Todd, P., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125 (1–2), 305–353.
Stephan, G., 2008. The effects of active labor market programs in Germany: an investigation using different definitions of non-treatment. Journal of Economics and Statistics 228 (5 + 6), 586–611.
Stephan, G., Pahnke, A., 2008. A Pairwise Comparison of the Effectiveness of Selected Active Labour Market Programmes in Germany. IAB. Discussion Paper 29/2008.
Storey, D., 1994. Understanding the Small Business Sector. Routledge, London.
Whiteford, P., Adema, W., 2007. What works best in reducing child poverty: a benefit or work strategy? Social, Employment and Migration Working Papers, 51. OECD.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press.