I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11055

The Post-Reform Effectiveness of the New German Start-Up Subsidy for the Unemployed

Lutz Bellmann
Marco Caliendo
Stefan Tübbicke

SEPTEMBER 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11055

The Post-Reform Effectiveness of the New German Start-Up Subsidy for the Unemployed

Lutz Bellmann
IAB, University of Erlangen-Nuremberg and IZA

Marco Caliendo
University of Potsdam, IZA, DIW and IAB

Stefan Tübbicke
University of Potsdam

SEPTEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Post-Reform Effectiveness of the New German Start-Up Subsidy for the Unemployed*

Start-up subsidies for the unemployed have long been an important active labor market policy strategy in Germany. The current subsidy program underwent a major reform in 2011 that changed its key parameters: support was lowered, eligibility criteria were tightened and entitlement was abandoned by granting caseworkers the right to reject applications. Ex-ante predictions on the post-reform effectiveness of the program are ambiguous, and knowledge about the importance of institutional details of such a program is very limited, making a new evaluation necessary. In our descriptive analysis, we compare personal and business characteristics of participants before and after the reform and we find significant differences in terms of gender composition, educational attainment and industry-specific experience. Post-reform participants also perform better in terms of subsequent labor market integration and show signs of higher commitment. These findings give us some indication for interpreting our estimates of causal effects of the post-reform program. We find that for both men and women, employment and income effects of the post-reform program are positive, sizable, and larger than what was estimated for the pre-reform program. All in all, the programs effectiveness seems to have improved through the reform. Potential reasons for this are discussed and include better screening of participants by caseworkers, higher rates of commitment and changes in macroeconomic conditions.

JEL Classification: J68, H43, C14, L26

Keywords: start-up subsidies, policy reform, matching, effect heterogeneity

Corresponding author:
Marco Caliendo
University of Potsdam
Chair of Empirical Economics
August-Bebel-Str. 89
14482 Potsdam
Germany

E-mail: caliendo@uni-potsdam.de

* The authors would like to thank Martin Weißenberger, Kristin Kleinjans, and participants at the IWH Workshop on Entrepreneurship and the Labor market, the OFS workshop on Self-Employment/Entrepreneurship and Public Policy and the 5th IAB Workshop on Passive and Active Labour Market Policies for valuable comments and Katalin Evers for help with the data preparation. The authors would also like to thank the Institute for Employment Research (IAB) for cooperation and institutional support within the research project no. 1505.

1 Introduction

Start-up subsidies (SUS) for the unemployed have been an integral part of German active labor market policy (ALMP) for many decades. Introduced in 1986, their usage grew and eventually peaked during the major labor market reforms in 2003/2004 (see Caliendo and Hogenacker, 2012; Rinne and Zimmermann, 2012) when they became a major part of the toolbox of German ALMP. With a yearly budget of up to €3.5 billion, start-up subsidy programs have been able to recruit many participants: From 2002 to 2011, between 120,000 and 250,000 new people signed up, with 2004 seeing as many as 350,000. According to official statistics, between 300,000 and 450,000 start-ups were observed every year (see Piorkowsky *et al.*, 2013, for details on self-employment and start-up activity in Germany). Thus, subsidized start-ups out of unemployment were a significant share of yearly start-up activities. At the same time, between three and five million people were registered as unemployed, meaning that between 3 percent and – at the peak of the program – almost 10 percent of all unemployed individuals used such a program each year. Along with the widespread usage, the "Hartz reforms" also made output evaluations mandatory such that interesting data became available, allowing a detailed examination with respect to effect heterogeneity, long-term effects and the comparison to regular business founders (see Caliendo and Künn, 2011, 2014, 2015; Caliendo *et al.*, 2015). This is especially interesting since the international evidence on the effectiveness of start-up subsidies is still relatively scarce – compared to the abundant literature on other ALMP's such as vocational training and job creation schemes – even though it has been growing in recent years.[1]

The main goal of start-up subsidies as an ALMP is to reintegrate unemployed individuals into the first labor market while at the same time maintaining their livelihood. This is done by subsidizing their start-ups out of unemployment for a limited duration in order to help them overcome their specific entry-barriers into the labor market, such as lack of formal education, experience or financial means. Policymakers also usually hope for a "double dividend," i.e., additional job creation by subsidized businesses started out of unemployment. However, it is often feared that these types of programs have large deadweight effects and that money could be spent more effectively by subsidizing only a

[1] International evidence on causal effects for developed countries is provided, e.g., by Tokila (2009) for Finland, Duhautois *et al.* (2015) for France, O'Leary (1999) for Hungary and Poland, Perry (2006) for New Zealand, Rodríguez-Planas and Jacob (2010) for Romania and Behrenz *et al.* (2016) for Sweden; Caliendo (2016) provides a summary of international evidence on the effectiveness and institutional set-up of start-up subsidies.

few but very promising businesses instead (see Shane, 2009, for this type of argument). The empirical evidence on the effects of SUS as a labour market policy is quite strong: effects on employment probabilities and earned income are positive and relatively large in magnitude for most countries (Caliendo, 2016). Research has also shown that effects of these types of programs are heterogenous with respect to certain characteristics like age, education, qualification and migration status. Effects tend to be larger for individuals having trouble finding jobs in the first labor market by themselves, such as low-skilled or migrant workers. While accumulated evidence on positive effects of start-up subsidies can be regarded as convincing, little is known about the importance of the institutional details of these types of programs for future impacts. Germany provides an interesting case study with respect to this, as there have been two major reforms of start-up subsidy programs since the peak in participation in 2004.

In our study, we focus on the current German subsidy program (called "Gründungszuschuss," dubbed New Start-Up Subsidy), which was reformed at the end of 2011. The reform – in an attempt to reduce spending on start-up subsidies – changed the key parameters of the program: entry requirements were tightened, support was reduced and full discretion to reject applications was given to caseworkers at the local employment agencies. From an ex-ante point of view, predictions on whether one should expect labor market effects of the post-reform program to be smaller or larger compared to the effects of the pre-reform program are ambiguous.

Based on a sample of 1,729 pre-reform participants, 1,922 post-reform participants and 2,091 comparison individuals for the post-reform program, our empirical analysis therefore provides evidence on three research questions. First, do personal and business related-characteristics differ significantly between the pre- and post-reform participants? Second, do participants perform differently in terms of labor market outcomes after the reform, and are there any hints on harmful effects of lowering monetary support? Third, what is the causal effect of participating in the post-reform program on subsequent employment probabilities and net monthly earned income and how do these effects compare to previously estimated effects of the pre-reform program?

We find that the share of women participating in the post-reform program is substantially larger. Also, participants after the reform are better educated and a larger proportion of participants resides in Eastern Germany. Evidence on business quality in

terms of start-up capital and equity as well as previous self-employment experience is inconclusive. Participants after the reform show signs of less self-employment experience and industry-specific knowledge. On the other hand, average start-up capital is larger for female participants after the reform, and the share of founders that finance the investment of their start-up entirely through equity grew. Subsequent employment rates are significantly larger, and exit rates out of self-employment react less sensitively to the end of monetary support among post-reform participants, indicating somewhat higher commitment. Our estimates of program effects are much larger for participants of the post-reform program on net earned income and employment probabilities up to 18 months after entering the program compared to what Caliendo *et al.* (2016) estimated for pre-reform participants. In line with previous research, we find that effects for women are larger when considering employment probabilities but smaller with respect to earned income. The exact reason as to why the effects are larger after the program is unclear, but better screening of participants, more commitment to their start-up and lower GDP growth in the post-reform era potentially play a role in explaining our findings.

This paper is organized as follows. Section 2 provides a detailed overview of the institutional details of the New Start-Up Subsidy program and its reform at the end of 2011. It discusses potential reasons for different selection patterns of participants, incentive effects and program effects after the reform. Section 3 describes our dataset and presents our descriptive analysis. Section 4 discusses identifying assumptions, empirical strategy and the results of our causal analysis. Section 5 discusses our findings and concludes the analysis.

2 The New Start-Up Subsidy

2.1 Institutional Details Before and After the Reform

The New Start-up Subsidy (NSUS) was introduced in 2006 and replaced the two previous German start-up subsidy programs (which were called "Bridging Allowance" and "Start-Up Subsidy"). From 2007 onwards, the NSUS has been the only start-up subsidy program available to recipients of unemployment benefits I in Germany.[2] At the end of 2011, the NSUS was reformed, mainly to achieve budgetary goals.

[2]Unemployment benefits I are an insurance benefit for which employees acquire legal claims by paying a part of their salary into the public unemployment insurance system. Recipients of the tax-financed UB II can apply for another start-up subisidy ("Einstiegsgeld", see Wolff *et al.*, 2016, for an evaluation).

In order to be eligible for the subsidy before the reform, individuals had to be entitled to at least another 90 days of unemployment benefits, and applicants' business plans needed to be approved by an independent institution like the chamber of commerce. Conditional on meeting these criteria, applicants were guaranteed to receive the subsidy by law. Participants of the NSUS before the reform received a monthly payment equivalent to their unemployment benefits, which depend on previous labor earnings, plus a lump-sum of €300 for the first 9 months after entering into the program. Participants could also apply for a second benefit period which only provided monthly payment of the lump-sum for an additional 6 months. In contrast to the first benefit period however, there was no entitlement feature in the second period. Only if applicants provided evidence that their business was still running full-time and that the subsidy was indeed needed, caseworkers could chose whether or not to grant the extension of subsidy receipt. About 62% of participants received the second benefit period in the pre-reform era. Average total support was €13,100 for participants.[3] As can be seen in Table 1, the number of entries into the program between 2007 and 2011 ranged from about 120,000 to 150,000 individuals.

[Insert Table 1 about here]

The reform of the New Start-up Subsidy took effect on December 28th 2011, and was accompanied by program budget cuts. The reform of the program comprised of three main elements: First, there was an eligibility reform. The required minimum remaining benefit entitlement (RBE) was increased from 90 to 150 days, meaning individuals have to have at least 150 days of unemployment benefit receipt remaining in order to be eligible for the subsidy ever since the reform took place. Second, there was an entitlement reform. In contrast to before, there is no longer an entitlement to the subsidy, and thus caseworkers now have the discretionary power to reject applicants even if all eligibility criteria are met. Third, there was a reform of the subsidy scheme. While the duration of the first benefit period, during which participants receive their unemployment benefits and the lump-sum of €300, was shortened from 9 to 6 months, the second benefit period was extended from 6 to 9 months, leaving the total length of the program unchanged. However, conditional on the level of unemployment benefits and the share of participants receiving the second benefit period, this resulted in a significant reduction in monetary support.

[3]Data on the fraction of participants receiving the second benefit period and average total support is taken from our random samples of participants before and after the reform.

After the reform only about 57% of participants received the second benefit period, further reducing monetary support. All in all, the lower attractiveness of the program, lower support and more difficult access led to a sharp decline in the number of entries into the program in the year after the reform: while about 130,000 joined the NSUS program in 2011, only about 20,000 did so in the year after the reform. However, the reform did not only lead to a reduction in spending via the extensive margin, but total support per participant also decreased by about €2,800 to €10,350. As a result, public expenditure on the NSUS declined by about €700m from 2011 to 2012, and even further afterwards. The fact that spending declined less rapidly than entries into the program is due to the stock of participants, which drive the direct cost of the program, decreasing at a slower rate since there were still plenty of participants from the pre-reform era.

2.2 Ex-Ante Predictions

Policymakers' main goals associated with the reform of the New Start-up Subsidy in 2011 were to reduce deadweight effects and spending. As can be seen in Table 1, both spending on the New Start-Up Subsidy and the number of entries greatly decreased after the reform. Evers and Schleinkofer (2015) find that the drop in the number of entries into the program was disproportionately seen in regions with previously high shares of entries into the New Start-Up Subsidy program relative to total unemployment. They also report that deadweight effects, as measured analogously to Caliendo *et al.* (2015), have risen slightly. But due to the large absolute decrease in the number of participants, the importance of these deadweight effects has decreased. While policymakers' goals were largely achieved, the reform could potentially have other (possibly unintended) consequences for selection patterns, incentives and the program's effectiveness.

Selection Patterns The reform of the New Start-up Subsidy could result in different selection patterns through a variety of channels. The increase in the required remaining benefit entitlement probably leads to a selection of older individuals with better labor market history into the program, because benefit entitlement is a non-linear but increasing function of both age and past employment.[4] Therefore, increasing the required remaining

[4] Aside from certain exceptions, the minimum benefit entitlement (BE) a person can have amounts to 180 days if the person was employed for at least one year out of the past 5 years, irrespective of age. The BE increases in a step-wise manner in past employment. Currently, a BE of over 450 days can only be obtained if the person is at least 50 years of age and worked for at least 2.5 years in the past 5 years.

benefit entitlement makes access to the program for younger individuals and people with worse employment history more difficult. Shortening the first benefit period and increasing the duration of the second benefit period, while leaving the total length of the program unchanged, significantly reduces monetary support to participants, conditional on the level of unemployment benefits. This might lead individuals with higher previous earnings to select into treatment in order to offset the shortening of the first benefit period. At the same time, reducing support requires business ideas to provide higher returns to participants than before the reform, thus potentially leading to an increase in business quality. The biggest factor of uncertainty regarding selection patterns, however, is the behavior of caseworkers. Caseworkers might both positively or negatively select on previous labor market history and entrepreneurial skills. Selecting individuals with better labor market experience and higher entrepreneurial ability might reflect well on the caseworker and the local employment agency. On the other hand, caseworkers might also choose to grant the subsidy to people with bad labor market history and possibly low entrepreneurial skill as they will be the ones who definitely are in need of the subsidy. Qualitative evidence on the implementation of the reform and caseworkers' decision making processes are provided by Bernhard and Grüttner (2015). In their interviews with stakeholders from different local employment agencies as well as applicants, they find that most often applicants were denied access to the program if their specific labor market provided a sufficient number of vacancies for which to apply as judged by the individual caseworker. This is in line with the "placement priority," meaning active labor market programs are only meant to be considered if a successful and timely reintegration of unemployed individuals into the labor market is not feasible.[5] With respect to the quality of business plans, some applicants were turned down if their business idea was either assessed to be too good (i.e., the subsidy was unnecessary) or deemed to fail (again, see Bernhard and Grüttner, 2015, for details). All in all, ex-ante predictions on selection patterns after the reform are ambiguous.

Incentives and the Program's Effectiveness If selection patterns emerge that lead to higher entrepreneurial skills among participants, treatment effect estimates will be larger ceteris paribus after the reform. Reducing monetary support might influence the incentive structure, resulting in participants showing more commitment to their start-up and lower levels of moral hazard, which would also show up as larger treatment effects

[5]Placement priority is defined in paragraph 4, section 2 of the Social Code Book (SGB) III.

after the reform. However, reducing the duration of the first benefit period might also make the program less effective as an ALMP as the duration might be too short for participants to overcome their specific entry barriers into the labor market (e.g., lack of labor market experience). From previous research on effect heterogeneity, we know that individuals who face disadvantages in the labor market also tend to show larger effects of participating in active labor market programs (see Rodriguez-Planas, 2010; Caliendo and Künn, 2011; Wolff *et al.*, 2016, for examples in the case of start-up subsidies). Thus, if there is negative selection into treatment after the reform, we would expect to see larger effects on employment probabilities and income. Lastly, effects of the reformed program might be different simply because macroeconomic conditions have changed. Table 1 also provides some data on the German annual GDP growth rate and the unemployment rate. While the unemployment rate is trending down over time, it does so at a lower rate in the post-reform era than before the reform. This goes hand in hand with a lower GDP growth rate from 2012 onwards, probably reducing the speed at which unemployed individuals find a new job and therefore leading to larger treatment effects after the reform.

3 Data and Descriptive Analysis

3.1 Data

For our analysis we combine two datasets. For the causal part of our analysis, we employ a new comprehensive dataset that was compiled at the Institute for Employment Research to evaluate the reformed program. We use a sample of post-reform participants and a comparison group, both drawn from the Integrated Labor Market Biographies (IEB) of the Federal Employment Agency (FEA). Our sample of post-reform participants is a random sample of the entire population of previously unemployed individuals who joined the program between February and June 2012. Our comparison group consists of individuals who were unemployed for at least one day, eligible for the program but did not apply for it in this period. Comparison individuals were pre-selected based on a nearest-neighbor matching procedure using basic information on socio-demographics, education and short-term labor market history as control variables. Each person in the comparison group was assigned a fictitious entry month so that individuals in the treated and non-treated groups have, on average, the same time span from the beginning of their unemployment spell to

their (fictitious) entry into the program.[6] Our final dataset includes register data from the IEB which contain detailed information on individuals' labor market history, previous earnings, education and treatment history. These data are augmented by survey data collected via computer-assisted telephone interviews that include information on participants' start-up characteristics, intergenerational information and labor market outcomes up to about 20 months after start-up. The dataset contains 1,922 post-reform participants, 837 (43.5%) of which are women. Our comparison group used to estimate the counterfactual for the post-reform participants is comprised of 1,045 men and 1,046 women. For the sake of our descriptive analysis, we also use data on pre-reform participants that were collected in a similar manner as described above (see Caliendo *et al.*, 2016, for a description of the dataset, IAB project 1143). The pre-reform sample we use consists of 1,729 participants, 34.5% (597) of which are women. Surveys for both cohorts were very similar, with many questions being largely identical, making the data readily comparable across cohorts.

3.2 Descriptive Analysis

In our descriptive analysis we look at selection patterns and outcomes of participants before and after the reform. Since previous research has shown that men and women display different start-up characteristics (e.g., see Verheul *et al.*, 2012; Wagner, 2007; Kelley *et al.*, 2013) and treatment effects (see Caliendo and Künn, 2015; Wolff and Nivorozhkin, 2012), we conduct our analysis, both descriptive and causal, separated by gender. Table 2 provides summary statistics on socio-demographics, human capital and labor market history for the pre- and the post-reform cohort; Table 3 gives an overview of business-related related characteristics for participants before and after the reform. For a more complete overview of descriptive statistics on covariates for the post-reform sample, see Table A.1 in the Appendix.

[Insert Table 2, Table 3 and Table 4 about here]

Selection Patterns When comparing participants from before and after the reform of the New Start-Up Subsidy, we find significant differences in terms of individual and busi-

[6] For each participant who entered the program in month m, 20 non-applicants were randomly drawn from the unemployed population and assigned month m as their month of fictitious entry. Nearest-neighbor pre-matching on the timing of entry into unemployment (among other variables) assured balance on the time-elapsed between entry into unemployment and fictitious entry and decided which units were contacted for the survey.

ness related characteristics of participants. First, the gender composition of participants changed substantially: the share of women participating in the program increased from 34.5% to 43.5% from 2009 to 2012. For both men and women it holds true that after the reform participants are more likely to reside in Eastern Germany, which might be an indication of caseworkers granting the subsidy to applicants living in regions with worse local labor market conditions. Post-reform participants are also better educated, i.e., they are more likely to hold a higher secondary school degree or even a university degree. Mean previous earnings are slightly larger and short-term labor market history is worse, on average, among post-reform participants. However, there are also some differences across cohorts with respect to gender: male participants after the reform are significantly older, on average, and more likely to be married than the pre-reform founders. Female participants after the reform on the other hand spent significantly less time in unemployment in the past 10 years prior to entering the program.

When comparing business-related characteristics of the two cohorts, there is conflicting evidence on business quality and entrepreneurial skills. On the one hand, among both men and women, a smaller fraction of post-reform participants entered unemployment from self-employment, possibly indicating lower entrepreneurial experience among the post-reform cohort. In addition, the share of participants without any industry-specific knowledge has increased, but roughly the same fraction of participants across cohorts has some industry-specific knowledge from self- or regular employment. On the other hand, the amount of capital invested at start-up stayed constant for men at around €19,300 and even increased substantially from €11,000 to €16,000 among female participants. Also, the share of founders financing their start-up capital entirely through equity is larger after the reform, which we take as evidence for the more entrepreneurial commitment of post-reform founders. Moreover, there has also been a sectoral shift away from construction, manufacturing and logistics towards general services. In addition, Evers and Schleinkofer (2015) find that post-reform participants are significantly more likely to be influenced by pull motives, such as perceived market opportunities or the desire to be their own boss (see Amit and Muller, 1995; Caliendo and Kritikos, 2009, for a discussion of push and pull motives in entrepreneurship).

Incentives and Outcomes Looking at outcome statistics in Table 4, one can see that both participant cohorts show high self-employment rates 18 months after start-up, which corresponds to at least three months after the last subsidy payment was received.[7] However, post-reform participants are significantly more often self-employed at 18 months after start-up compared to pre-reform participants. While 86.7% of male pre-reform participants were self-employed 18 months after entry into the program, 93.7% of male post-reform participants were self-employed at that point in time. The self-employment rates for women in our sample are slightly smaller, but also significantly different across cohorts. Since policy-makers usually hope for additional job creation by subsidized founders, Table 4 also includes some information on subsequent job creation of the new businesses. For men, around 34 percent of start-ups have at least one employee, whereas female start-ups are less likely to hire (24%). These findings are essentially the same for both the pre- and post-reform participants. With respect to the number of full-time equivalent employees (FTE) we see a significant drop for men (from 2.6 to 1.6 FTE), but no statistically significant change for women (from 2.2 to 1.8 FTE).

For our causal analysis later on, we focus on net monthly earned income and an overall employment indicator. This indicator equals one if the person is either self-employed or regular employed and thus subject to the social security system. Both men and women of the post-reform cohort display larger employment rates than pre-reform participants. However, the difference in overall employment rates for women is only significant at the 10% level. For both men and women it holds that there are no significant differences in net earned income between pre- and post-reform participants. Since our comparison group for the post-reform participants has been pre-matched, the sample is not representative of any underlying population. However, we can see that the comparison group displays weaker labor market history and performs worse in terms of subsequent employment probability and income compared to the treatment group.

[Insert Figure 1 about here]

The previous analysis suggest that post-reform participants were more successful in terms of labor market integration. To gain further insights, Figure 1 provides some more information on self-employment and/or regular employment rates as well as exit rates from

[7] Again, 18 months corresponds to at least three months after the program has ended as only a fraction of participants actually received the second benefit period. While 62% of participants before the reform received the second benefit period, 57% of participants did so after the reform.

self-employment for both participant cohorts. Statistically significant differences between pre- and post-reform cohorts at the 5% level are marked with a cross.[8] One can see that up until about 9 months after start-up, which coincides with the end of the first subsidy period under the pre-reform program, both self-employment as well as overall employment rates are nearly identical and not significantly different. Only after that period is there a significant, and for men even widening, gap in (self-) employment rates. The gap in overall employment remains approximately constant and is only significant for a small number of time periods among women. When looking at exit rates out of self-employment, one can see interesting patterns: First, exit rates are quite small overall for both genders and cohorts. Second, for both men and women among the pre-reform participants, there is a relatively large spike in exit rates at 10 months, just after their first benefit period ended. This spike is even slightly larger for women in our sample. For the post-reform cohort, there is more of a permanent increase in exit rates after the end of the first benefit period (six months). This behavior cannot be observed when looking at the end of the second benefit period (15 months) as the lump-sum support of €300 probably is too low for a large fraction of participants to be at the margin of dropping out of self-employment at that point. While these findings could potentially be confounded by different macroeconomic conditions, the consistent pattern suggests that post-reform participants' survival in self-employment is less sensitive to the end of the first benefit period. We take this as further evidence of higher commitment to their business and potentially lower moral hazard among post-reform founders.

4 Caual Analysis

4.1 Identification and Estimation of Treatment Effects

The goal of our causal analysis is to estimate the treatment effects of the reformed NSUS program on participants' labor market outcomes. We rely on the well known potential outcomes framework (Roy, 1951; Rubin, 1974) and focus on the average treatment effect on the treated (ATT):

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

[8]The conclusions we draw from this descriptive analysis do not change in any important way if we additionally condition on observable characteristics.

where Y^1 and Y^0 are potential outcomes with and without treatment and D is a treatment indicator (= 1 if individual received a start-up subsidy). Since participants and non-participants are selected groups, estimating the counterfactual in (1) with the sample mean of non-participants would lead to selection bias. To correct for that we use propensity score matching estimators relying on the conditional independence assumption (CIA), which implies that conditional on the propensity score $P(X) = Pr(D = 1 \mid X)$ the counterfactual outcome is independent of treatment, where X corresponds to a vector of observable characteristics. In addition to the CIA, we also assume overlap which implies that there are no perfect predictors which determine participation, i.e., $P(X) < 1$ for all X, and that there are no general equilibrium effects. Under these assumptions the ATT can be estimated from sample moments (see, e.g., Imbens and Wooldridge, 2009, for more details).

The applicability of matching estimators crucially depends on the plausibility of the conditional independence assumption, which cannot be directly tested. As mentioned by Lechner and Wunsch (2013), for the matching procedure to give an unbiased estimate of the ATT, the propensity score estimation must include all such variables that both simultaneously affect selection into treatment and outcomes of participants. Hence, if there is selection on unobservables and these unobservables also significantly affect outcomes, the CIA will fail, and our estimates will be biased as the treated population will exhibit a different distribution of the outcome even in the untreated state. In our application this would be the case if, for example, individuals in the treated sample were more motivated and would thus search more intensely for a new job to find their way out of unemployment. Then, even in the absence of treatment, participants would be more likely to find a job, and our estimates would be biased upwards.

Evidence shows, however, that matching estimators in the context of program evaluation using administrative data are less sensitive to usually unobserved confounders than previously thought. Caliendo *et al.* (2014) look at the effects of training measures and wage subsidies. Although they find that usually unobserved variables like personality traits, job search intensity and subjective employment probabilities are predictive of receiving treatment, they find no significant differences in estimated effects when including these factors in their causal analysis. Even more relevant for our analysis is the investigation conducted by Caliendo *et al.* (2016). They analyze the consequences of omitting potentially impor-

tant confounders in the context of evaluating start-up subsidies. To be more specific, they investigate whether excluding the big 5 personality traits, locus of control and risk aversion from the propensity score estimation significantly affects estimated treatment effects. In line with findings of Caliendo *et al.* (2014), they conclude that these variables do predict treatment, but excluding them from the matching procedure does not have a significant impact on estimated program effects. In related work, Lechner and Wunsch (2013) also use administrative data to evaluate the bias of treatment effect estimates due to omitting variables relative to their baseline specification using a variety of variables from their very detailed dataset. They find that omitting the timing of entry into unemployment and active labor market programs, along with labor market and earnings histories, health status as well as regional information, results in the largest bias in estimates relative to their baseline results. Since our exhaustive dataset consists of both administrative and survey data, we are able to condition on a large number of socio-demographic, human capital, labor market history, earnings history, regional and intergenerational variables in a detailed and flexible fashion. We therefore argue that the CIA holds in our application.

4.2 Estimation Procedure

Our analysis is based on propensity score matching. That means we use a two-step estimation procedure. The first step consists of estimating a probit for the probability of participation in treatment. Our probit specification includes a large vector of variables with very detailed information on socio-demographics, human capital, labor market history, intergenerational transmission of education and self-employment as well as regional labor market types (see Dauth *et al.*, 2008, for details). The exact specification and estimated coefficients can be found in Table A.2. The distributions of the propensity scores are depicted in Fig 2a for men and Fig 2b for women. The distributions are skewed towards one for the treated and towards zero for the comparison group, indicating covariate imbalances before matching.

[Insert Figure 2 about here]

Matching Next we match the treatment and the comparison group with an Epanechnikov kernel matching algorithm using an optimal bandwidth chosen via weighted cross

validation.[9] Galdo *et al.* (2008) find in their Monte Carlo analysis that this combination performs well in terms of mean squared error. More recent work using real data combined with a Monte Carlo analysis by Huber *et al.* (2013) finds that radius matching with post-matching regressions for bias adjustment performs best overall. In our application, however, we found little difference between our results and estimates obtained by using bias-adjusted radius matching. We impose common support by dropping treated observations from the analysis if they exceed the minimum or the maximum value of the propensity score of the control group in order to avoid bad matches.[10]

[Insert Table 5 about here]

Matching Quality Since propensity score matching does not match on covariates directly, it is necessary to inspect whether the matching procedure sufficiently balances covariate distributions across treated and non-treated samples. Several indicators on the matching quality are displayed in Table 5. First, for both men and women, the number of significantly different means of control variables declines drastically through the matching procedure. In our sample of men, there only remain two variables significantly different at the 5% level. For women, there remains only one variable significantly different at the 10% level between treatment and control group after matching. Second, we use the (mean) absolute standardized bias (MSB) as described by Rosenbaum and Rubin (1983). The table gives an overview of the distribution of standardized biases before and after matching. The matching procedure greatly reduces the MSB from initially above 10% to an acceptable level of about 2.5% for men and 1.9% for women (Caliendo and Kopeinig, 2008). Re-estimating the propensity score on the matched sample should yield a low pseudo-R^2 and a high p-value of joint-significance (see Sianesi, 2004, for more details). This is exactly what we observe: the matching procedure reduces the pseudo-R^2 to 1% and the p-value of joint-significance increases to one. Other measures include Rubin's B and Rubin's R

[9]The weighted cross-validation alghorithm makes use of only nearest neighbors (with replacement) among the non-treated sample for leave-one-out cross-validation as to better approximate the propensity score distribution of the treated sample. As mentioned by Galdo *et al.* (2008), the optimal bandwidth chosen via weighted cross validation tends to be larger than using traditional cross-validation techniques. However, this does not affect our results in a significant manner.

[10]Table A.3 in the Appendix shows the results of a sensitivity analysis of our estimates with respect to matching algorithm and common support definition. We find no economically meaningful differences compared to our baseline estimates presented in the text.

as described in Rubin (2001).[11] Rubin's R is within the suggested bounds to indicate covariate balancing, while Rubin's B only slightly exceeds the suggested threshold of 25% for the male sample. Overall, we can conclude that the matching was highly successful in balancing the distribution of covariates across the two groups.

Inference Bodory *et al.* (2016) conduct an empirical Monte Carlo analysis of the performance of different inference methods. They show that empirically, bootstrapping often outperforms asymptotic variance approximations in terms of size and has comparable power even in the case of non-smooth estimators like nearest-neighbor matching, where bootstrapping is predicted to perform poorly (see Abadie and Imbens, 2008). Since kernel matching estimators are much smoother, we are confident that bootstrapping will yield reliable results in our application. Therefore, we follow Huber *et al.* (2014) and obtain p-values by bootstrapping the t-statistic. In addition, we provide 95% confidence intervals based on the percentile method. As suggested by MacKinnon (2006), we use 999 bootstrap replications for our estimates.

4.3 Estimated Treatment Effects

For the evaluation of the New Start-Up Subsidy, our outcome variables of interest are "self- or regular employed" as a binary indicator for labor market integration, cumulated months spent in self-employment or regular employment to gain some insight on the overall effect independent of timing and net monthly earned income.

[Insert Table 6 and Figure 3 about here]

Table 6 shows the estimated treatment effects for both men and women of the post-reform participant cohort. Our estimates show that participation in the reformed New Start-Up Subsidy significantly increased the likelihood of being either in self-employment or in regular employment by 25.8 percentage points for men and 30.4 percentage points for women 18 months after entering the program. While point estimates suggest that effects are larger for women, confidence intervals of the effects overlap in this case. The size of the effect is much larger than the estimated employment effects provided by Caliendo *et al.* (2016) for the pre-reform program. They find effects of around 14 percentage points

[11] Rubin's B is the standardized difference in means of the linear index of the propensity score and Rubin's R corresponds to the ratio of variances in the treated and control group of the propensity score index.

for men and 19 percentage points for women 21 months after entering into the program. Although this is a somewhat longer period of observation, their analysis also reveals that effects are approximately constant between 18 and 21 months. Thus estimated employment effects are about 11 percentage points larger than under the pre-reform program.

Cumulating the effects over time, we see that male participants of the post-reform program spend 6.9 months more in either self- or regular employment than their matched group of non-participants. For women, this effect is even larger: they spend, on average, 7.9 months more in employment than individuals in the control group. When looking at the effects on net monthly earned income, we see that men gain around €950 from participating in the program.[12] Effects on earned income are smaller for women: they gain about €530 through taking part in the subsidy program. For both the cumulated effects and the effect on net earned income, estimates are significantly different for men and women as indicated by the confidence intervals. Relative to the effects under the pre-reform program as estimated by Caliendo *et al.* (2016), post-reform effects again seem to be larger. The pattern of larger employment effects for women but smaller effects on income are consistent patterns that have emerged in a number of evaluations of other start-up subsidy programs (see, e.g., Caliendo and Künn, 2011, 2015, for similar evidence on an earlier program).

In Figure 3 we see the estimated effects on the probability of being in self- or regular employment over time. Effects for both men and women are statistically significantly different from zero for the whole period of observation. For both genders, the initial impact of taking part in the subsidy program is very large. This represents a positive lock-in effect, working in the opposite direction of a lock-in effect for participants of in-class training, for example, meaning participants of the subsidy program are unlikely to drop out of (self-) employment at least for the duration of the first benefit period. This lock-in effect subsides over time, and employment effect estimates decline relatively rapidly until the end of the first benefit period at six months. Up until the end of the subsidy program at 15 months, there is still a slight downward trend in estimated effects, but for the remaining observation period the effects remain stable.

[12]Note that the analysis of effects on net monthly earned income is unconditional, i.e., estimates are a combination of effects at the extensive and the intensive margin.

5 Discussion and Conclusion

The New Start-Up Subsidy was introduced in 2006 and underwent a major reform at the end of 2011. Shortening the first benefit period, reducing the size of the eligible population via tightening entry requirements and providing more discretionary power to caseworkers gave rise to relatively substantial differences in terms of selection patterns and a huge drop in the number of entries into the program. Not only did total direct expenditure of the program decrease after the reform, but average direct cost per participants also decreased substantially.

Our analysis can be summarized as follows: First, after the reform a significantly larger fraction of participants was female, and individuals participating in the program after the reform were also better educated. In addition, they are more likely to live in Eastern Germany, which might be indicative of caseworkers granting the subsidy to applicants in weaker labor markets. Furthermore, evidence on business quality is inconclusive: On the one hand, average capital invested at start-up is at least as large for post-reform participants, and more businesses are financed entirely through equity than before the reform. However, post-reform participants also have less industry-specific knowledge and show signs of less self-employment experience. Second, participants after the reform are to be less responsive towards the end the first benefit period, possibly indicating higher commitment or lower moral hazard induced by the reduction in monetary support. Third, our estimates of program effects on employment probabilities and net monthly earned income show large and positive effects on both. Our findings also indicate heterogenous effects with respect to gender: Women display larger estimated effects on employment but lower effects on net earned income. Estimated effects are also much larger in magnitude compared to what was estimated for the pre-reform program. Thus, the reform appears to have been very successful from a narrow point of view: Direct cost per participant decreased while program effects seem larger.

These larger treatment effects after the reform are most likely the result of a combination of factors. Potential channels include better screening of participants by caseworkers at the local employment agencies, higher commitment among post-reform participants and possibly lower GDP growth, making the transition from unemployment to employment for the comparison group harder. While the extent to which each factor contributes to our findings is uncertain, our analysis suggests that two cautious conclusions can be drawn

with respect to policy design: First, there is no indication that shortening the first benefit period had any negative effects on the labor market outcomes of participants. Therefore, reasonable monetary support for a relatively short duration is likely to be sufficient to ensure a successful transition from unemployment to self-employment beyond the length of the subsidy program. Second, giving caseworkers at local employment offices more discretionary power to grant/reject the subsidy does appear to significantly affect selection into the program. The geographical shift in participation patterns may be a sign that caseworkers tend to grant the subsidy to applicants from weaker labor markets, thereby increasing the effectiveness of the program through the channel of regional effect heterogeneity.

One limitation of this study is that we cannot identify the overall effect on welfare. Although we have provided evidence that the program has not become less effective for participants, one has to bear in mind the drastic drop in the number of participants after the reform. Potentially successful businesses might have never been started due to the reform. It is therefore possible that, overall, welfare has decreased through the reform, which we are not able to analyze with our microeconometric approach. Furthermore, we are only able to investigate short-term effects of the subsidy on participants' labor market prospects. The question whether the reformed program is also effective in the long term remains to be investigated.

References

Abadie, A. and Imbens, G. (2008). On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, **76** (6), 1537–1557.

Amit, R. and Muller, E. (1995). Push and Pull Entrepreneurship. *Journal of Small Business & Entrepreneurship*, **12** (4), 64–80.

Behrenz, L., Delander, L. and Månsson, J. (2016). Is Starting a Business a Sustainable way out of Unemployment? Treatment Effects of the Swedish Start-up Subsidy. *Journal of Labor Research*, **37** (4), 389–411.

Bernhard, S. and Grüttner, M. (2015). *Der Gründungszuschuss nach der Reform: Eine qualitative Implementationsstudie zur Umsetzung der Reform in den Agenturen*. IAB-Forschungsbericht 4/2015, IAB Nürnberg.

Bodory, H., Camponovo, L., Huber, M. and Lechner, M. (2016). *The Finite Sample Performance of Inference Methods for Propensity Score Matching and Weighting Estimators*. IZA Discussion Papers 9706, Institute for the Study of Labor (IZA).

Caliendo, M. (2016). Start-up subsidies for the unemployed: Opportunities and limitations. *IZA World of Labor*, **200**, 1–11.

— and Hogenacker, J. (2012). The German Labor Market after the Great Recession: Successful Reforms and Future Challenges. *IZA Journal of European Labor Studies*, **1** (1), 1–24.

—, —, Künn, S. and Wiessner, F. (2015). Subsidized start-ups out of unemployment: a comparison to regular business start-ups. *Small Business Economics*, **45** (1), 165–190.

— and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, **22** (1), 31–72.

— and Kritikos, A. (2009). *"I Want to, But I Also Need to": Start-Ups Resulting from Opportunity and Necessity*. IZA Discussion Paper No. 4661, Institute for the Study of Labor (IZA).

— and Künn, S. (2011). Start-Up Subsidies for the Unemployed: Long-Term Evidence and Effect Heterogeneity. *Journal of Public Economics*, **95** (3-4), 311–331.

— and — (2014). Regional Effect Heterogeneity of Start-Up Subsidies for the Unemployed. *Regional Studies*, **48 (6)**, 1108–1134.

— and — (2015). Getting back into the labor market: the effects of start-up subsidies for unemployed females. *Journal of Population Economics*, **28** (4), 1005–1043.

—, — and Weissenberger, M. (2016). Personality traits and the evaluation of start-up subsidies. *European Economic Review*, **86**, 87 – 108.

—, Mahlstedt, R. and Mitnik, O. A. (2014). *Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies*. IZA Discussion Papers 8337, Institute for the Study of Labor (IZA).

Crump, R., Hotz, V. J., Imbens, G. W. and Mitnik, O. A. (2009). Dealing with Limited Overlap in Estimation of Average Treatment Effects. *Biometrika*, **96** (1), 187–199.

Dauth, W., F., H. and Rüb, F. (2008). *Vergleichstypen 2008: Neufassung der SGB-III-Typisierung*. IAB-Forschungsbericht 8/2008, IAB Nürnberg.

Duhautois, R., Redor, D. and Desiage, L. (2015). Long Term Effect of Public Subsidies on Start-up Survival and Economic Performance: An Empirical Study with French Data. *Revue d'économie industrielle*, **149** (1), 11–41.

Evers, K. and Schleinkofer, M. (2015). *Der Gründungszuschuss vor und nach der Reform. Same, but different: Ein Vergleich der Teilnehmerstrukturen.* IAB-Forschungsbericht 05/2015, IAB Nürnberg.

Galdo, J., Smith, J. A. and Black, D. (2008). Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data. *Annales d'Économie et de Statistique*, **91/92**, 189–216.

Huber, M., Lechner, M. and Steinmayr, A. (2014). Radius Matching on the Propensity Score with Bias Adjustment: Tuning Parameters and Finite Sample Behavior. *Empirical Economics*, **49** (1), 1–13.

—, — and Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, **175** (1), 1–21.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47** (1), 5–86.

Kelley, D. J., Brush, C. G., Greene, P. G. and Litovsky, Y. (2013). *Global Entrepreneurship Monitor: 2012 Women's Report.* Report, Global Entrepreneurship Research Association.

Lechner, M. and Wunsch, C. (2013). Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Economics*, **21**, 111–121.

MacKinnon, J. G. (2006). Bootstrap Methods in Econometrics. *Economic Record*, **82**, 2–18.

O'Leary, C. J. (1999). *Promoting Self Employment Among the Unemployed in Hungary and Poland.* Working Paper 99-55, W.E. Upjohn Institute for Employment Research.

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

Piorkowsky, M., Buddensiek, M. and Eschweiler, R. H. (2013). *Selbstständige in Deutschland 1992-2011.* Report, University Bonn.

Rinne, U. and Zimmermann, K. F. (2012). Another economic miracle? The German labor market and the Great Recession. *IZA Journal of Labor Policy*, **1** (1), 1–21.

Rodriguez-Planas, N. (2010). Channels through which public employment services and small-business assistance programs work. *Oxford Bulletin of Economics and Statistics*, **72** (4), 458–485.

Rodríguez-Planas, N. and Jacob, B. (2010). Evaluating active labor market programs in romania. *Empirical Economics*, **38** (1), 65–84.

Rosenbaum, P. and Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70** (1), 41–55.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, **3** (2), 135–146.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomised and Nonrandomised Studies. *Journal of Educational Psychology*, **66** (5), 688–701.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, **2** (3), 169–188.

Shane, S. (2009). Why Encouraging More People to Become Entrepreneurs is Bad Public Policy. *Small Business Economics*, **33**, 141–149.

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.

Tokila, A. (2009). *Start-up grants and self-employment duration*. Working paper, School of Business and Economics, University of Jyväskylä.

Verheul, I., Thurik, R., Grilo, I. and van der Zwan, P. (2012). Explaining preferences and actual involvement in self-employment: Gender and the entrepreneurial personality. *Journal of Economic Psychology*, **33** (2), 325 – 341.

Wagner, J. (2007). What a Difference a Y makes - Female and Male Nascent Entrepreneurs in Germany. *Small Business Economics*, **28**, 1–21.

Wolff, J. and Nivorozhkin, A. (2012). Start me up: The effectiveness of a self-employment programme for needy unemployed people in Germany. *Journal of Small Business & Entrepreneurship*, **25** (4), 499–518.

—, — and Bernhard, S. (2016). You can go your own way! The long-term effectiveness of a self-employment programme for welfare recipients in Germany. *International Journal of Social Welfare*, **25** (2), 136–148.

Tables and Figures

Table 1: Official Statistics

	Pre-reform period					Post-reform period			
	2007	2008	2009	2010	2011	2012	2013	2014	2015
Data on the New Start-Up Subsidy									
Number of entries (thousands)	126	119	137	147	134	20	27	31	30
Total expenditure (bn euros)	1.13	1.41	1.48	1.80	1.68	0.89	0.23	0.32	0.32
Macroeconomic conditions									
Unemployment rate in %	9.0	7.8	8.1	7.7	7.1	6.8	6.9	6.7	6.4
Real GDP growth rate in %	3.9	1.2	-6.1	4.3	3.6	0.5	0.3	1.5	1.5

Note: Data is based on publications from the Federal Employment Agency (FEA) of Germany and Statistical Federal Office Germany. Own calculations. Data on total expenditure on the New Start-Up Subsidy program has been inflation-adjusted to 2012 price levels.

Table 2: Descriptives – Personal Characteristics

	Pre-reform	Post-reform			
	Part.	Part.	Non-part.	p-val.	p-val.
	(1)	(2)	(3)	(2)v(3)	(1)v(2)
A. Men					
Number of men in each sample	1132	1085	1045		
Share of men in each sample	0.655	0.565	0.500		
Socio-demographics					
Average age at (hypothetical) entry (years)	40.4	42.7	43.9	0.004	0.000
Eastern Germany	0.221	0.342	0.369	0.186	0.000
Married	0.579	0.614	0.561	0.013	0.091
Human capital					
Educational and vocational training (VT)					
Lower/middle secondary school, no VT	0.048	0.018	0.033	0.019	0.000
Lower/middle secondary school, with VT	0.509	0.441	0.637	0.000	0.001
Higher secondary school, no VT	0.034	0.053	0.020	0.000	0.028
Higher secondary school, with VT	0.084	0.138	0.088	0.000	0.000
University degree	0.240	0.336	0.209	0.000	0.000
Other/no training	0.086	0.015	0.012	0.646	0.000
Labor market history					
Fraction of last 10 years in unemployment	0.107	0.102	0.194	0.000	0.283
Fraction of last 10 years in ALMP	0.030	0.029	0.051	0.000	0.799
Months in employment one year before entry	9.692	8.171	7.184	0.000	0.000
Daily income from last employment (euros)	86.88	91.50	69.91	0.000	0.057
Entered unemployment from self-employment	0.141	0.053	0.011	0.000	0.000
B. Women					
Number of women in each sample	597	837	1046		
Share of women in each sample	0.345	0.435	0.500		
Socio-demographics					
Average age at (hypothetical) entry (years)	40.8	40.9	42.2	0.003	0.893
Eastern Germany	0.251	0.333	0.365	0.150	0.001
Married	0.553	0.545	0.554	0.705	0.765
Human capital					
Educational and vocational training (VT)					
Lower/middle secondary school, no VT	0.037	0.007	0.030	0.000	0.000
Lower/middle secondary school, with VT	0.471	0.344	0.511	0.000	0.000
Higher secondary school, no VT	0.039	0.072	0.034	0.000	0.008
Higher secondary school, with VT	0.099	0.160	0.134	0.109	0.001
University degree	0.268	0.393	0.274	0.000	0.000
Other/no training	0.087	0.024	0.016	0.235	0.000
Labor market history					
Fraction of last 10 years in unemployment	0.121	0.108	0.164	0.000	0.052
Fraction of last 10 years in ALMP	0.038	0.033	0.044	0.003	0.227
Months in employment one year before entry	9.002	7.167	6.499	0.000	0.000
Daily income from last employment (euros)	54.21	60.77	50.07	0.000	0.017
Entered unemployment from self-employment	0.114	0.065	0.009	0.000	0.001

Note: Reported are sample means if not indicated otherwise. p-values are based on t-tests of equal means. All variables measured in euros have been inflation-adjusted for the comparison over time.

Table 3: Descriptives – Business Characteristics

	Pre-reform Part.	Post-reform Part.	p-val.
A. Men			
Industry-specific knowledge			
None	0.097	0.132	0.010
Yes, from dependent employment	0.404	0.397	0.756
Yes, from self-employment	0.103	0.091	0.337
Investment at start-up			
Start-up capital (euros)	19,280	19,334	0.978
Capital is entirely equity	0.522	0.588	0.002
Sector			
Retail or wholesale	0.137	0.114	0.108
Construction	0.125	0.135	0.523
Manufacturing	0.154	0.036	0.000
Logistics	0.041	0.024	0.027
Services	0.358	0.514	0.000
Agriculture	0.010	0.018	0.081
Other	0.176	0.159	0.276
B. Women			
Industry-specific knowledge			
None	0.109	0.201	0.000
Yes, from dependent employment	0.315	0.326	0.653
Yes, from self-employment	0.103	0.091	0.337
Investment at start-up			
Start-up capital (euros)	11,158	16,414	0.022
Capital is entirely equity	0.472	0.570	0.000
Sector			
Retail or wholesale	0.151	0.117	0.063
Construction	0.025	0.026	0.892
Manufacturing	0.077	0.018	0.000
Logistics	0.013	0.000	0.000
Services	0.441	0.545	0.000
Agriculture	0.007	0.008	0.722
Other	0.286	0.286	0.971

Note: Reported are sample means if not indicated otherwise. p-values are based on t-tests of equal means. All variables measured in euros have been inflation-adjusted for the comparison over time.

Table 4: Descriptives – Outcomes

	Pre-reform	Post-reform			
	Part.	Part.	Non-part.	p-val.	p-val.
	(1)	(2)	(3)	(2)v(3)	(1)v(2)
A. Men					
Labor market outcomes 18 months after entry					
Self-employed	0.867	0.937	0.054	0.000	0.000
Self- or regular employed	0.939	0.972	0.676	0.000	0.000
Unemployed	0.058	0.013	0.185	0.000	0.000
Other[b]	0.003	0.015	0.14	0.000	0.055
Net monthly earned income (euros)[a,c]	2,354	2,197	920.3	0.000	0.117
Job Creation					
Start-up hired employees	0.349	0.333			0.455
Full-time equivalent employees					
per start-up	0.892	0.541			0.000
per hiring business	2.570	1.634			0.000
B. Women					
Labor market outcomes 18 months after entry					
Self-employed	0.881	0.926	0.051	0.000	0.004
Self- or regular employed	0.938	0.958	0.605	0.000	0.085
Unemployed	0.044	0.016	0.162	0.000	0.001
Other[b]	0.018	0.026	0.233	0.000	0.424
Net monthly earned income (euros)[a,c]	1,374	1,332	652.1	0.000	0.622
Job Creation					
Start-up hired employees	0.268	0.239			0.264
Full-time equivalent employees					
per start-up	0.581	0.438			0.142
per hiring business	2.183	1.840			0.299

Note: Reported are sample means if not indicated otherwise. p-values are based on t-tests of equal means. All variables measured in euros have been inflation-adjusted for the comparison over time. Sample averages for job-creation do not condition on survival.

Table 5: Matching Quality

	Men		Women	
	Before Matching	After Matching	Before Matching	After Matching
Number of variables with significant differences in means[a]				
at 1%-level	37	0	28	0
at 5%-level	46	2	34	0
at 10%-level	48	2	38	1
Number of variables with absolute standardized bias[b]				
0%≤ SB <1%	2	19	7	24
1%≤ SB <3%	7	31	13	39
3%≤ SB <5%	8	19	10	8
5%≤ SB <10%	17	5	14	4
SB ≥10%	41	1	31	0
Mean absolute standardized bias in %	15.249	2.524	11.174	1.912
(Re-)Estimation of the propensity score[c]				
Pseudo-R^2	.195	.0123	.1624	.0102
p-Value of joint-significance test	0.000	1.000	0.000	1.000
Other measures				
Rubin's B[d] in %	111.095	26.248	100.465	23.842
Rubin's R[e]	.902	1.734	1.086	1.64
Number of variables	75	75	75	75
Number of participants off support		91		19

Note: Shown are different indicators for covariate balancing before and after Epanechnikov-kernel matching using the optimally chosen bandwidths via the nearest-neighbor approach of weighted cross validation (see Galdo *et al.*, 2008, for details). The bandwidths used are 0.22 for men and 0.17 for women.
[a]: Number of variables with significantly different means is based on a t-test of equality of means.
[b]: The standardized absolute bias of a variable is the mean difference between treatment and control group as a percentage of the square-root of the mean of pre-matched variances of both groups.
[c]: Following Sianesi (2004) Pseudo-R^2 and p-value of joint significance from a probit estimation on the unmatched and the matched sample are also calculated.
[d]: Rubin's B is the standardized mean difference of the linear index of the propensity score of treatment and control group.
[e]: Rubin's R is the variance ratio of the propensity score index of the treated to control sample.

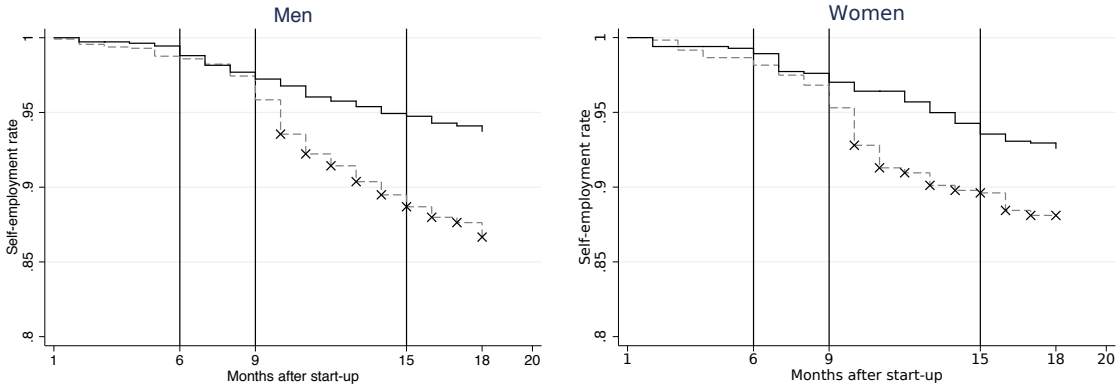
Table 6: Estimation Results

	$\hat{\tau}_{ATT}$	95% Confidence Interval
A. Men		
Self- or regular employed		
18 months after entry	$.2579^{***}$	[.2260; .2968]
Cumulated effect ($\sum_{t=1}^{18}$)	6.896^{***}	[6.340; 7.384]
Net monthly earned income (euros)	954.7^{***}	[762.4; 1105.4]
A. Women		
Self- or regular employed		
18 months after entry	$.3035^{***}$	[.2678; .3472]
Cumulated effect ($\sum_{t=1}^{18}$)	7.977^{***}	[7.460; 8.641]
Net monthly earned income (euros)	530.6^{***}	[424.5; 662.8]

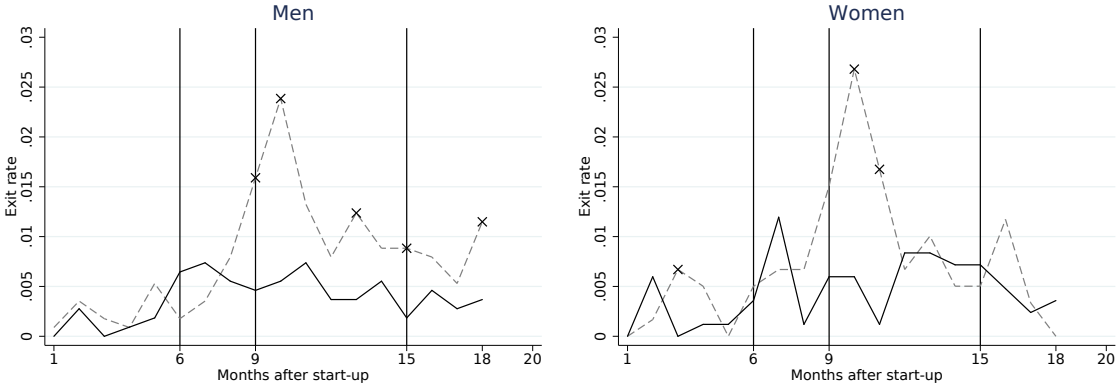
Note: Reported are estimates of ATT using Epanechnikov-kernel matching with optimally chosen bandwidths via the nearest-neighbor approach of weighted cross validation. Following Huber *et al.* (2014) and MacKinnon (2006), confidence intervals are obtained by bootstrapping the t-statistic and using the percentile-method with 999 replications. $^{***}/^{**}/^{*}$ denote significance at the 1/5/10 % level.

Figure 1: (Self-) Employment and Exit Rates

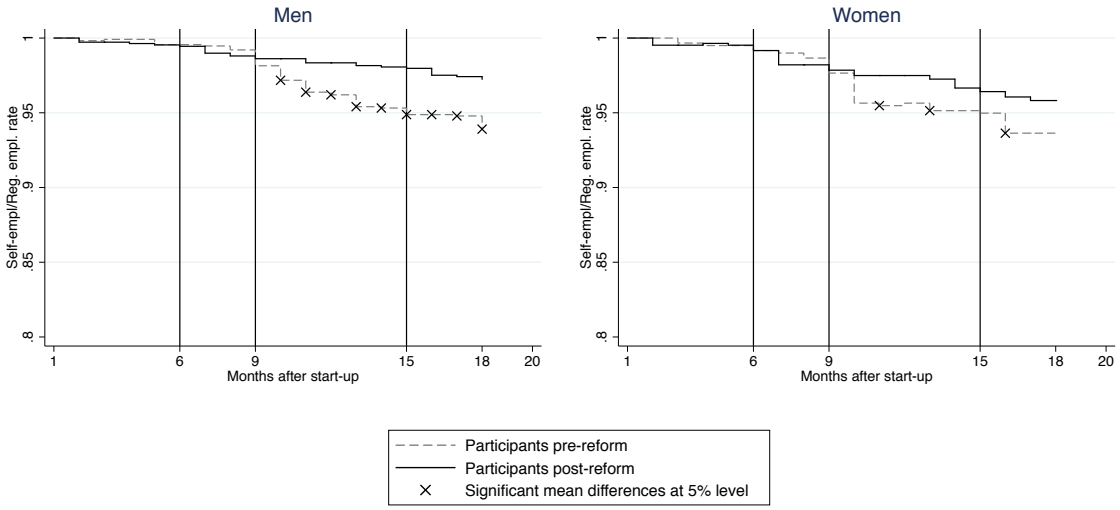
a. Self-employment rate



b. Exit rates out of self-employment

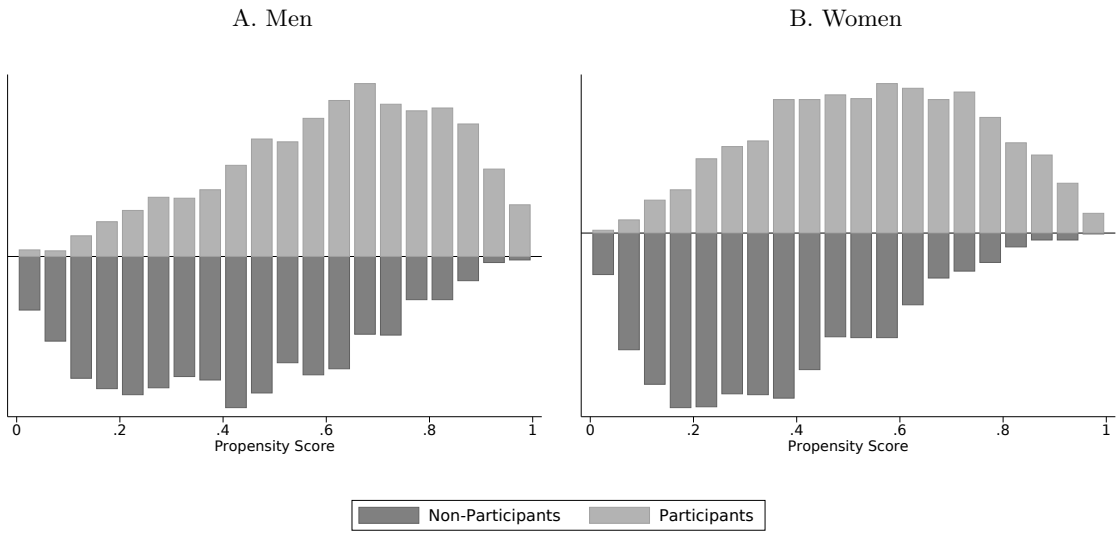


c. Self-employment or regular employment rate



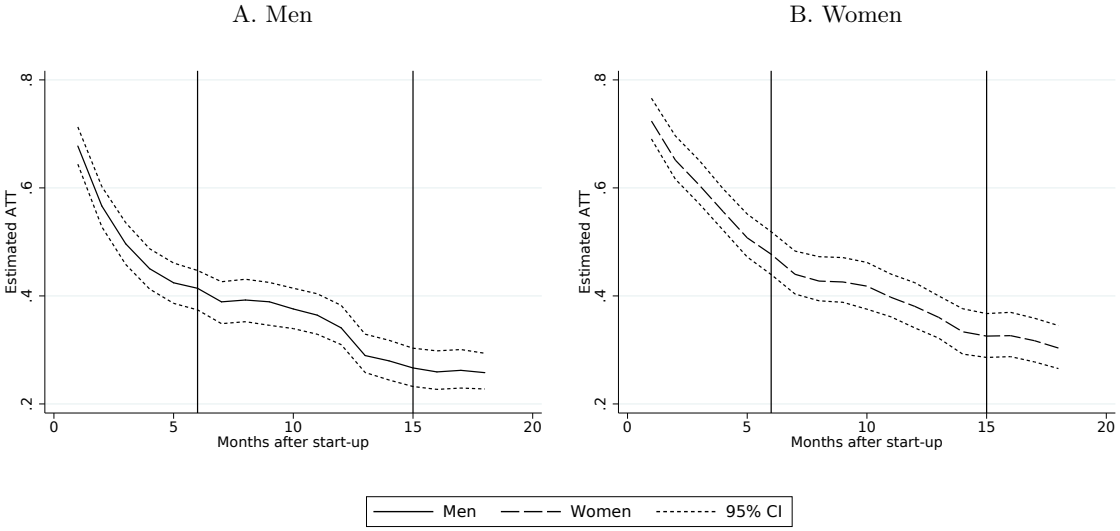
Note: Shown are self-employment, exit rates out of self-employment and rates of overall employment (that is, either self-employment or regular employment) for men and women separately. The vertical line at 6 months represents the end of the first benefit period (UB+€300) under the post-reform program, the line at 9 months shows the end of said benefit period under the pre-reform program. The line at 15 months depicts the end of the second benefit period (€300) and, therefore, the end of both subsidy programs.

Figure 2: Propensity Score Distributions



Note: These are the propensity score distributions for post-reform participants and non-participants based on probit estimations in Table A.2. Our specification includes controls for socio-demographics, human capital, labor market history, intergenerational transmission and regional labor market types.

Figure 3: Estimated Effect on Self- or Regular Employment



Note: Depicted are the estimated treatment effects on the probability of being in self- or regular employment up to 18 months after entry for both male and female participants. Treatment effects are estimated as the mean difference between participants and a group of matched non-participants. In the first 6 months of the program, participants received monetary support equivalent to their unemployment benefits plus a lump-sum of €300. Participants that are granted the second subsidy period receive only the lump-sum until 15 months after entry into the program. About 57% of participants received this second subsidy period.

A Supplementary Tables and Figures

Table A.1: Descriptive Statistics for Control Variables

	Men			Women		
	Part.	Non-part.	*p*-val.	Part.	Non-part.	*p*-val.
N	1085	1045		837	1046	
Socio-demographics						
Age at (hypothetical) entry						
average (in years)	42.655	43.877	0.004	40.881	42.169	0.003
younger than 25 years	0.020	0.018	0.725	0.020	0.011	0.122
25 to less than 35 years	0.203	0.190	0.474	0.249	0.250	0.922
35 to less than 45 years	0.334	0.297	0.066	0.378	0.332	0.039
45 to less than 56 years	0.193	0.192	0.987	0.173	0.181	0.674
56 years and older	0.251	0.302	0.008	0.180	0.226	0.016
Eastern Germany	0.342	0.369	0.186	0.333	0.365	0.150
Not German citizen	0.045	0.029	0.045	0.038	0.021	0.026
Health restrictions	0.042	0.082	0.000	0.031	0.047	0.082
Married	0.614	0.561	0.013	0.545	0.554	0.705
Number of children						
no child	0.626	0.580	0.030	0.497	0.451	0.048
one child	0.187	0.212	0.144	0.258	0.299	0.048
two children and above	0.187	0.208	0.233	0.245	0.250	0.818
Children under 10 present	0.212	0.171	0.017	0.293	0.272	0.332
Single parent	0.010	0.006	0.254	0.109	0.101	0.603
Highest schooling degree						
lower secondary school	0.146	0.314	0.000	0.069	0.164	0.000
middle secondary school	0.311	0.361	0.014	0.277	0.419	0.000
upper secondary school (specialized)	0.185	0.110	0.000	0.149	0.101	0.002
upper secondary school (general)	0.341	0.195	0.000	0.489	0.293	0.000
no schooling degree	0.018	0.020	0.661	0.016	0.022	0.310
Educational and vocational training (VT)						
lower/middle secondary school, no VT	0.018	0.033	0.019	0.007	0.030	0.000
lower/middle secondary school, with VT	0.441	0.637	0.000	0.344	0.511	0.000
higher secondary school, no VT	0.053	0.020	0.000	0.072	0.034	0.000
higher secondary school, with VT	0.138	0.088	0.000	0.160	0.134	0.109
university degree	0.336	0.209	0.000	0.393	0.274	0.000
other/no training	0.015	0.012	0.646	0.024	0.016	0.235
Professional qualification						
unskilled workers	0.588	0.473	0.000	0.562	0.497	0.005
skilled workers	0.271	0.430	0.000	0.270	0.376	0.000
skilled workers with techn. college educ.	0.041	0.033	0.333	0.016	0.020	0.462
top management	0.100	0.064	0.003	0.153	0.107	0.003
Labor market history						
Fraction of time in unemployment in the						
last 10 years average (in %)	0.102	0.194	0.000	0.108	0.164	0.000
less than 10%	0.660	0.393	0.000	0.657	0.496	0.000
10 to less than 20 %	0.190	0.226	0.041	0.191	0.212	0.259
20 to less than 40 %	0.119	0.256	0.000	0.108	0.185	0.000
40 to less than 60 %	0.021	0.083	0.000	0.033	0.076	0.000
more than 60 %	0.010	0.042	0.000	0.011	0.032	0.002
Five years before (hypothetical) entry						
months employed	10.142	9.383	0.000	8.516	8.651	0.594
months in labor market program	0.333	0.668	0.000	0.376	0.531	0.061
Four years before (hypothetical) entry						
months employed	10.594	9.681	0.000	9.160	9.139	0.927
months in labor market program	0.304	0.559	0.001	0.425	0.439	0.855
Three years before (hypothetical) entry						
months employed	10.932	9.729	0.000	9.835	9.515	0.150
months in labor market program	0.298	0.673	0.000	0.331	0.417	0.257

(Table continued on next page)

(Table A.1 continued)

	Men			Women		
	Part.	Non-part.	p-val.	Part.	Non-part.	p-val.
Two years before (hypothetical) entry						
months employed	11.429	10.122	0.000	10.134	9.729	0.048
months in labor market program	0.179	0.503	0.000	0.232	0.358	0.066
One year before (hypothetical) entry						
months employed	8.171	7.184	0.000	7.167	6.499	0.000
months in labor market program	0.393	0.414	0.702	0.424	0.400	0.678
Employment status before entering unemp.						
dependent employment	0.699	0.539	0.000	0.608	0.478	0.000
self-employment/family worker	0.053	0.011	0.000	0.065	0.009	0.000
school/apprenticeship	0.020	0.029	0.208	0.020	0.017	0.621
disable to work/unemployable	0.030	0.118	0.000	0.050	0.141	0.000
other	0.000	0.000		0.000	0.000	
Occupational group before entering unemp.						
manufacturing	0.030	0.033	0.686	0.023	0.015	0.237
forestry, fishing, animal breeding	0.206	0.364	0.000	0.049	0.079	0.008
technical profession	0.103	0.062	0.001	0.043	0.035	0.394
services	0.659	0.538	0.000	0.879	0.868	0.466
other	0.001	0.003	0.299	0.006	0.002	0.150
Daily income from last employment (euros)	91.496	69.906	0.000	60.771	50.070	0.000
Duration of last unemployment spell						
average (in months)	3.547	4.232	0.001	4.129	4.599	0.116
less than 1 month	0.114	0.025	0.000	0.108	0.033	0.000
1 to less than 3 months	0.399	0.404	0.823	0.363	0.363	0.997
3 to less than 6 months	0.280	0.334	0.007	0.270	0.315	0.032
6 to less than 12 months	0.175	0.190	0.361	0.216	0.242	0.190
12 to less than 24 months	0.026	0.035	0.198	0.029	0.036	0.355
24 months and above	0.006	0.011	0.134	0.014	0.011	0.454
Monthly unemployment benefit						
average (in euros)	1185.320	1015.856	0.000	865.251	773.549	0.000
less than 300 euros	0.085	0.047	0.000	0.104	0.053	0.000
300 to less than 600 euros	0.072	0.087	0.195	0.216	0.321	0.000
600 to less than 900 euros	0.182	0.341	0.000	0.254	0.317	0.003
900 to less than 1200 euros	0.190	0.245	0.002	0.194	0.182	0.511
1200 to less than 1500 euros	0.169	0.137	0.041	0.111	0.072	0.003
1500 euros and above	0.302	0.144	0.000	0.121	0.055	0.000
Intergenerational information						
Father and/or mother was born abroad	0.169	0.203	0.042	0.182	0.188	0.709
Father and/or mother is/was self-employed	0.341	0.223	0.000	0.378	0.242	0.000
Father employed when respondent 15 years old	0.902	0.834	0.000	0.913	0.846	0.000
Professional education of father						
vocational training	0.446	0.494	0.027	0.380	0.493	0.000
professional/vocational academy	0.153	0.130	0.131	0.174	0.110	0.000
technical college/university degree	0.244	0.141	0.000	0.293	0.178	0.000
other/no training	0.087	0.091	0.729	0.075	0.073	0.830
Regional labor market						
Regional cluster[a]						
type Ia	0.086	0.083	0.838	0.100	0.100	0.999
type Ib	0.080	0.061	0.089	0.063	0.045	0.077
type IIa	0.076	0.081	0.679	0.104	0.102	0.907
type IIb	0.107	0.118	0.430	0.114	0.128	0.335
type IIc	0.075	0.070	0.669	0.073	0.067	0.614
type IIIa	0.140	0.133	0.635	0.125	0.128	0.863
type IIIb	0.060	0.051	0.354	0.055	0.059	0.689
type IVa	0.055	0.065	0.343	0.047	0.045	0.864
type IVb	0.035	0.042	0.396	0.044	0.050	0.576
type IVc	0.011	0.017	0.228	0.012	0.016	0.435
type Va	0.079	0.075	0.689	0.078	0.081	0.774
type Vb	0.120	0.113	0.620	0.121	0.106	0.321
type Vc	0.076	0.091	0.200	0.065	0.072	0.540

Note: Reported are sample means. p-values are based on t-tests of equal means. [a] Dauth *et al.* (2008) cluster labor markets according to urbanization, unemployment, seasonality, employment-to-population ratio and the share of tertiary value-added of GDP.

Table A.2: Propensity Score Probit Estimation

	Men	Women
Socio-demographics		
Age		
(ref.: less than 25 years)		
25 to less than 35 years	-.185	-.414
35 to less than 45 years	-.268	-.264
45 to less than 56 years	-.262	-.283
56 years and older	-.322	-.290
Eastern Germany	-.195	-.267**
Female		
Not German citizen	0.626***	0.421**
Health restrictions	0.029	0.157
Married	0.093	-.023
Number of children		
(ref: no children)		
one child	-.190**	-.070
two children and above	-.219**	-.017
Children under 10 present	0.174*	0.035
Single parent	0.268	0.157
Highest schooling degree		
(ref.: no schooling degree)		
lower secondary school	0.119	-.167
middle secondary school	0.409*	0.053
upper secondary school (specialized)	0.621**	0.504**
upper secondary school (general)	0.564**	0.494**
Educational and vocational training		
(ref. other/no training)		
lower/middle secondary school, no vocational training	-.451	-.778**
lower/middle secondary school, with vocational training	-.188	-.216
higher secondary school, no vocational training	0.257	0.134
higher secondary school, with vocational training	-.018	-.249
university degree	-.333	-.402
Professional qualification		
(ref.: unskilled workers)		
skilled workers	-.139*	-.085
skilled workers with technical college education	0.034	-.323
top management	0.059	-.037
Labor market history		
Fraction of time in unemployment in the last 10 years		
(ref.: less than 10%)		
10 to less than 20 %	-.129	-.200**
20 to less than 40 %	-.438***	-.371***
40 to less than 60 %	-.769***	-.655***
more than 60 %	-.646**	-.865***
Five years before (hypothetical) entry		
months employed	-.013	-.021**
months in labor market program	-.019	-.017
Four years before (hypothetical) entry		
months employed	0.018	0.005
months in labor market program	0.05**	0.042
Three years before (hypothetical) entry		
months employed	-.007	0.013
months in labor market program	-.024	0.035
Two years before (hypothetical) entry		
months employed	0.026**	-.007
months in labor market program	-.053*	-.058*
One year before (hypothetical) entry		
months employed	0.018	0.01
months in labor market program	0.105***	0.05*

(Table continued on next page)

(Table A.2 continued)

	Men	Women
Employment status before entering unemployment		
(ref.: other)		
dependent employment	0.333***	0.261***
self-employment/family worker	1.398***	1.371***
school/apprenticeship	-.138	0.186
disable to work/unemployable	-.444***	-.329***
Occupational group before entering unemployment		
(ref.: other)		
forestry, fishing, animal breeding	0.444	-.386
manufacturing	0.175	-.798
technical profession	0.345	-.746
services	0.29	-.717
Daily income from last employment (euros)	-.0007	-.001
Duration of last unemployment spell		
(ref.: less than 1 month)		
1 to less than 3 months	-.600***	-.452***
3 to less than 6 months	-.597***	-.528***
6 to less than 12 months	-.570***	-.469***
12 to less than 24 months	-.585**	-.550**
24 months and above	-.493	0.331
Monthly unemployment benefit		
(ref.: less than 300 euros)		
300 to less than 600 euros	-.149	-.530***
600 to less than 900 euros	-.326**	-.501***
900 to less than 1200 euros	-.184	-.349**
1200 to less than 1500 euros	-.017	-.188
1500 euros and above	0.185	0.047
Intergenerational information		
Father and/or mother was born abroad	-.141	-.037
Father employed when respondent 15 years old	0.066	0.165
Father and/or mother is/was self-employed	.291***	.310***
Professional education of father		
vocational training	0.142*	-.073
professional/vocational academy	0.167	0.19
technical college/university degree	0.293***	0.138
Regional labor market		
Regional cluster[a]		
(ref.: type Ia)		
type Ib	0.272*	0.324*
type IIa	-.137	0.068
type IIb	-.003	0.205
type IIc	0.19	0.276*
type IIIa	0.063	0.26*
type IIIb	0.263	0.203
type IVa	-.00003	0.074
type IVb	0.047	0.241
type IVc	0.03	0.129
type Va	0.482**	0.618***
type Vb	0.469**	0.685***
type Vc	0.34*	0.612***
Const.	-.323	1.314*
Obs.	2130	1883
Pseudo-R^2	0.195	0.162
log-Likelihood	-1188.15	-1083.496

Note: Reported are probit coefficients. ***/**/* denote significance at the 1/5/10% level.

[a] Regional clusters group together similar regional labor markets based on urbanization, unemployment rate, seasonality, employment-to-population ratio and the share of tertiary value-added of GDP. For details see Dauth *et al.* (2008).

Table A.3: Sensitivity Analysis

	Epanechnikov-Kernel Matching			Radius Matching + Bias-Adj.		
	CS1[a]	CS2[b]	CS3[c]	CS1[a]	CS2[b]	CS3[c]
A. Men						
Self- or regular employed						
18 months after entry	.2579***	.2884***	.2998***	.2638***	.2696***	.2533***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Cumulated effect ($\sum_{t=1}^{18}$)	6.897***	7.321***	7.549***	7.100***	7.166***	6.693***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Net monthly earned income (euros)	954.7***	1191.6***	1149.0***	957.5***	978.2***	909.7***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
B. Women						
Self- or regular employed						
18 months after entry	.3035***	.3438***	.3604***	.2861***	.2889***	.3039***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cumulated effect ($\sum_{t=1}^{18}$)	7.977***	8.619***	8.890***	7.519***	7.621***	8.077***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Net monthly earned income (euros)	530.6***	665.5***	536.2***	462.6***	470.0***	363.9***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)

Note: Reported are estimated average treatment effects on the treated using Epanechnikov-kernel matching with optimally chosen bandwidths via the nearest-neighbor approach of weighted cross validation as well as radius-matching with post-matching bias adjustment as described by Huber *et al.* (2014) for different common support definitions. p-values are estimated by bootstrapping the t-statistic. Following MacKinnon (2006) 999 replications are used for the bootstrap procedure. p-values are shown in parentheses underneath the point-estimate. ***/**/* denote significance at the 1/5/10% level.

[a]: CS1 imposes common support by dropping observations of treated individuals from the estimation-sample that are above the maximum or below the minimum pscore in the untreated sample.

[b]: CS2 imposes common support using the rule-of-thumb of only using individuals in the $[0.1, 0.9]$-interval as provided by Crump *et al.* (2009).

[c]: CS3 imposes common support by setting a minimum density in the propensity score distribution among the untreated. We drop treated individuals from the analysis if there are less than 2% of all untreated in a corresponding 5%-interval.