New Evidence on Long-Term Effects of Start-Up Subsidies. Matching Estimates and their Robustness

Marco Caliendo

Stefan Tübbicke

CEPA

Center for Economic Policy Analysis

New Evidence on Long-Term Effects of Start-Up Subsidies: Matching Estimates and their Robustness*

Marco Caliendo
University of Potsdam, IZA Bonn, DIW Berlin, IAB Nuremberg

Stefan Tübbicke
University of Potsdam

ABSTRACT

The German start-up subsidy (SUS) program for the unemployed has recently undergone a major make-over, altering its institutional setup, adding an additional layer of selection and leading to ambiguous predictions of the program's effectiveness. Using propensity score matching (PSM) as our main empirical approach, we provide estimates of long-term effects of the post-reform subsidy on individual employment prospects and labor market earnings up to 40 months after entering the program. Our results suggest large and persistent long-term effects of the subsidy on employment probabilities and net earned income. These effects are larger than what was estimated for the pre-reform program. Extensive sensitivity analyses within the standard PSM framework reveal that the results are robust to different choices regarding the implementation of the weighting procedure and also with respect to deviations from the conditional independence assumption. As a further assessment of the results' sensitivity, we go beyond the standard selection-on-observables approach and employ an instrumental variable setup using regional variation in the likelihood of receiving treatment. Here, we exploit the fact that the reform increased the discretionary power of local employment agencies in allocating active labor market policy funds, allowing us to obtain a measure of local preferences for SUS as the program of choice. The results based on this approach give rise to similar estimates. Thus, our results indicating that SUS are still an effective active labor market program after the reform do not appear to be driven by "hidden bias".

Corresponding author:
Marco Caliendo
University of Potsdam
Department of Economics and Social Sciences
August-Bebel-Str. 89
14482 Potsdam
Germany
E-Mail: caliendo@uni-potsdam.de

1 Introduction

Start-up subsidies (SUS) for the unemployed are an unconventional active labor market program (ALMP). They help unemployed individuals to escape unemployment by incentivizing them to start their own business and securing their livelihood during the first uncertainty-ridden months of the start-up. The usage of SUS has recently been on the rise: according to official statistics by the OECD (2015), participation in this type of programs is high, whereby in Spain 8.7% of the stock of unemployed participated in a start-up incentive program, closely followed by France with 6.7% and Poland with 3.8%.[1] The empirical evidence on the effectiveness of SUS as an ALMP is more scarce compared to other programs such as training measures, although the body of evidence is growing. In general, almost all studies find positive and relatively large effects on individual labor market outcomes.[2] However, all of the mentioned studies rely on the conditional independence assumption (Lechner, 2001), also known as the selection-on-observables assumption (Heckman and Robb, 1985), whereby they assume that – conditional on a vector of observable characteristics – treatment is as good as randomly assigned. Thus, these estimates are susceptible to "hidden bias" if the researcher does not observe all relevant pre-treatment characteristics.

In this paper, we provide first evidence on long-term individual labor market effects of the German SUS program called "Gründungszuschuss" – which we dub New Start-Up Subsidy (NSUS) – after its reform in 2011. The reform altered the institutional setup of the program and was mainly intended to reduce spending on SUS (see Bernhard and Grüttner, 2015). The reduction in spending was achieved through abandoning entitlement to the program, thereby giving caseworkers at local employment agencies more discretionary power to reject applicants as well as by instituting large budget cuts of about €800m from 2011 to 2012. Additionally, monetary support to participants was reduced, leading to ambiguous predictions of the post-reform effectiveness of the program.[3] Furthermore, abolishing the entitlement to the program introduced an additional layer of selection, thus potentially reducing the credibility of making inference using methods relying on the conditional independence assumption. Therefore, studying the effects of SUS and their sensitivity to deviations from the identifying assumption under

---

[1] For an overview of the importance of SUS programs in OECD countries, see Figure 1.

[2] E.g., effect estimates are provided by Tokila (2009) for Finland, Duhautois *et al.* (2015) for France, Caliendo and Künn (2011) and Wolff *et al.* (2016) for Germany, 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. An in-depth review of estimated effects and the institutional setup is given by Caliendo (2016).

[3] For a detailed description of the program before and after the 2011 reform for the NSUS in Germany, estimated short-term program effects and a discussion of the importance of the institutional setup of the program, see Bellmann *et al.* (2018).

this post-reform setting provides an interesting case study to shed some light on the reliability of estimates under these circumstances. This is especially true because SUS programs in other countries operate with a similar selection mechanism, requiring joint decision-making by the unemployed individual and the caseworker (see, e.g. Behrenz *et al.*, 2016, on the current Swedish program). In addition, many countries' SUS programs are designed in a similar fashion where, support is granted by paying out a series of periodic transfers to recipients, mostly dependent on previous labor earnings (O'Leary, 1999).

Our main approach to estimating long-term effects of the German NSUS makes use of propensity score matching (PSM), as introduced by Rosenbaum and Rubin (1983). Within the matching framework, we assess the robustness of our estimates with respect to implementation-related issues and deviations from the conditional independence assumption (CIA). Going beyond the standard matching approach, we also provide estimates using an instrumental variable (IV) identification approach based on regional variation in the likelihood of receiving treatment. Here we exploit the fact that the reform increased the discretionary power of local employment agencies in allocating ALMP funds, allowing us to obtain a measure of local preferences for the NSUS as the ALMP of choice. As a proxy for these preferences, we use regional application approval rates for the NSUS, conditional on local labor market conditions. Using a sample of 1,248 participants and 1,204 non-participants, our matching results indicate persistent and positive long-term effects on individual employment probabilities and labor earnings up to 40 months after entering the program. Our sensitivity analysis within the matching framework shows that these findings are robust with respect to both issues related to the implementation of the matching estimators as well as deviations from the CIA. Finally, our estimates based on the IV strategy also give rise to similar estimates. Thus, our findings of large and positive effects of SUS for participants are unlikely to be driven by "hidden bias".

The remainder of this paper is organized as follows. Section 2 provides an overview of the institutional details of the NSUS program, gives details on the selection by caseworkers and discusses theoretical predictions on the post-reform effectiveness. Section 3 describes our dataset and presents some descriptives. Section 4 discusses the necessary identifying assumptions of our matching approach. Section 5 provides our main estimates and discusses effect heterogeneity. Section 6 performs our extensive sensitivity analyses and Section 7 concludes.

2 Institutional Setup of the New Start-Up Subsidy

The post-reform program In its current form, the NSUS has been in place since December 2011.[4] In order to be eligible for the program, unemployed individuals have to be entitled to at least another 150 days of unemployment benefits and obtain proof of sustainability for their business plan issued by an independent institution like the chamber of commerce. In contrast to previous programs, there is no legal entitlement to the subsidy under the reformed NSUS conditional on meeting the aforementioned eligibility criteria.[5] Thus, caseworkers at local employment agencies (LEAs) can deny access to the program to eligible applicants. Successful applicants receive a monthly payment equivalent to their unemployment benefits, which depends on previous labor earnings, plus a lump sum of € 300 for the first 6 months after entering into the program. Participants may also apply for a second benefit period that only provides monthly payment of the lump sum for an additional 9 months. Thus, in total, the program provides financial support to participants for a maximum of 15 months. In our sample, about 57% of participants received transfers for the second benefit period. The average total support was € 10,350 for participants.

Selection by caseworkers For the purpose of our analysis, it holds particular importance to understand the selection mechanism that determines participation and non-participation. Selection into different ALMPs is regulated by §7, social code book III, which states that caseworkers make an individual decision on the necessity of activation measures and the appropriateness of certain measures for the unemployed individual. When making this decision, the abilities of the unemployed individuals are to be taken into consideration. For the case of SUS, this means that the applicant needs to be considered as sufficiently entrepreneurial to run a business. Bernhard and Grüttner (2015) provide important qualitative evidence on caseworkers' behavior and the way in which they and their LEAs handled the transition to more discretionary power induced by the reform of the program. In their interviews with stakeholders from different LEAs, they find that the most commonly-cited reason why applications were rejected was a sufficiently large number of applicant-specific vacancies in the local labor market, as judged by the individual caseworker. This is consistent with the so-called "placement priority" as defined by §4, social code book III, which states that caseworkers are only meant to consider ALMPs as an option for unemployed individuals if they are necessary for the re-integration of the individual. Taken

---

[4]It is currently the only SUS program available to unemployment benefits I recipients. Unemployment benefits II recipients, which are mostly long-term unemployed or individuals with very sparse employment history, are eligible for a different program called "Einstiegsgeld", which is not the focus of this study.

[5]For a description and evaluations of the predecessor programs, see Caliendo and Künn (2011, 2014, 2015); Caliendo *et al.* (2016).

together, this information suggests that the most important confounders in our analysis are the individuals' re-employment probability in the absence of treatment and their entrepreneurial affinity. Arguably, the former can be controlled for relatively well using pre-treatment labor market outcomes, local labor market conditions and measures of human capital. However, the latter is generally unobservable and difficult to proxy for (see Caliendo *et al.*, 2016, for a detailed discussion of this issue) and thus at the center of our sensitivity analysis in Section 6.

Theoretical predictions By comparison, the pre-reform program required fewer days of unemployment benefits to be eligible and the first benefit period lasted 9 months, instead of 6. Shortening of the first benefit period might lead to larger effects through a reduction in moral hazard, although it may also reduce the effectiveness of the program due to lower financial support to help overcome capital constraints. Moreover, the additional layer of selection induced by the reform may potentially lead to larger effects due to the previously-mentioned "placement priority" by selecting individuals who benefit more from the program. Furthermore, effects may be different simply due to macroeconomic forces. Overall, these considerations lead to ambiguous predictions on the magnitude of effects after the reform relative to before.

3 Data and Descriptive Analysis

3.1 Data

For our analysis, we use a random sample of previously-unemployed participants who joined the program between February and June 2012. Data on participants from January was not used as most entrants still joined the program under pre-reform conditions, i.e. they applied before the reform was in place. 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. Both samples were drawn from the Integrated Labor Market Biographies (IEB) of the Federal Employment Agency (FEA). Our dataset combines extensive register data from the IEB with informative survey data collected via two computer-assisted telephone interviews around 20 and 40 months after entering the program. In order to reduce survey costs, non-participants to be interviewed were selected via a pre-matching strategy to avoid interviewing individuals with very dissimilar observed characteristics compared to actual participants. For this purpose, 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. A nearest neighbor matching was conducted based on basic variables such as age, gender, education, regional labor

market types and short-term labor market history as measured by the employment status at the end of 2011, the timing of entry into unemployment as well as the (hypothetical) entry month. Aside from ensuring the basic comparability of participants and non-participants, this yields a balanced duration of the unemployment spell from the date of entry into unemployment and the (hypothetical) month of entry into the program across the two groups.[6] Among non-participants, only nearest neighbors were contacted for the survey.

Due to the combination of register and survey data, our dataset contains extensive covariates on individuals' labor market history, previous earnings, socio-demographics, human capital, ALMP history, participants' start-up characteristics, intergenerational information as well as usually-unobserved personality traits. From the surveys, we are able to use labor market outcome data up to 40 months after (hypothetical) entry into the program. The final dataset contains 1,248 participants and 1,204 non-participants. Participants in our sample account for about 17% of all entrants into the SUS program during our sampling frame.[7]

3.2 Some Descriptives

In this part, we provide a brief descriptive overview of our sample of participants and comparison individuals. Table 1 provides summary statistics on socio-demographics, human capital, labor market history, intergenerational transmission, regional macroeconomic conditions and personality traits. For a more extensive overview of descriptive statistics on covariates, see Table A.1 in the Appendix. Outcome statistics can be found in Table 2.

[Insert Table 1 and Table 2 about here]

Pre-treatment characteristics Participants are on average about 42 years old and about one year younger than non-participants. In addition, participants are less likely to be female. While a sizable fraction of about 43% of participants have attained a general upper secondary school degree, which grants access to the German university system, only 28% of non-participants have such a degree. With respect to labor market history, participants spent on average 10% of the last 10 years in unemployment. On the other hand, non-participants were unemployed for 17% of the last 10 years. Short-term employment history shows that participants were employed for about 7.7 months in employment in the previous year before entering the program. On average, non-participants were employed for about one month less during this time period. The

---

[6]Participants spent on average 2.8 months in unemployment before entering the program. Our sample of non-participants was unemployed for 2.7 months on average prior to the assigned date of entry. The p-value of a t-test of equality of means is about 0.22.

[7]According to the FEA, about 7,400 individuals entered the program between February and June 2012.

majority of participants and non-participants (67% and 52%, respectively) were in dependent employment before entering unemployment. While 5.4% of participants were self-employed before entering unemployment, only 1.2% of non-participants had the same employment status. An economically-significant fraction of 35% of the treated and 25% of comparison individuals have at least one self-employed parent, which is described as one of the key drivers in the decision to become self-employed by the entrepreneurship literature (e.g. see Dunn and Holtz-Eakin, 2000; Lindquist *et al.*, 2016). Participants and non-participants also differ with respect to personality traits. For example, participants are on average more conscientious, extraverted, open to new experience and more risk-tolerant than comparison individuals.

This shows that although comparison individuals have been pre-matched and thus their sample is not representative of the underlying general unemployed population, there remain significant in-sample differences in key characteristics between the treated and non-treated.[8]

Labor market outcomes Table 2 provides mean labor market outcomes for participants and non-participants at 20 and 40 months after (hypothetical) entry. At the first interview – about 20 months after entry – 88.8% of participants and only 3.7% of non-participants are self-employed. Despite being smaller at the second interview, the gap remains substantial. For our causal analysis later, we will focus on an overall employment indicator, without discriminating between self-employment or regular employment, as both types of employment are seen as a successful integration into the labor market. At the first interview – 20 months after entry – 95.8% of participants and 61.3% of non-participants are in self- or regular employment. At the second interview – after 40 months – the overall employment rate is slightly lower for participants with 93.3% and higher for non-participants with 67.4%. At both interviews, there is a substantial raw gap in net monthly labor earnings in favor of the participants.

4 Main Empirical Approach

The goal of our causal analysis in this and the next section is to estimate the treatment effects of the SUS program on individuals' labor market outcomes in terms of overall employment and earned income. We rely on the well-known potential outcomes framework, mainly attributed to Roy (1951) and Rubin (1974). Our main focus is to estimate the average treatment effect on the treated (ATT)

---

[8]The significant gap between treated and comparison group characteristics is due to the fact that pre-matching was done in a very coarse way to ensure minimal overlap between the two groups.

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

where Y^1 and Y^0 are potential outcomes with and without treatment and D is a treatment indicator ($= 1$ if individual received a SUS). Since $E(Y^0 \mid D = 1)$ is generally unobservable, it has to be inferred from data on non-participants' outcomes. However, simply using the mean outcome of non-participants will lead to biased estimates in the absence of random assignment of treatment due to differential characteristics between the two groups.

Propensity Score Matching   PSM techniques aim to eliminate selection bias by balancing a rich set of observable characteristics X across the two groups. To give consistent estimates, the so-called CIA

$$Y^0 \perp\!\!\!\perp D \mid P(X) \tag{2}$$

needs to hold, where $P(X) = Pr(D = 1 \mid X)$ is the propensity score. In addition, one has to assume overlap $(P(X) < 1 \ \forall \ X)$ and rule out spill-over effects of treatment (*Stable Unit Treatment Value Assumption*). As noted by Imbens and Wooldridge (2009), if these three assumptions hold, we can estimate (1) as the simple mean difference between treated and comparison individuals on the *re-weighed* sample as

$$\hat{\tau}_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N_1} Y_i^1 - \sum_{j=1}^{N_0} \hat{w}_j Y_j^0, \tag{3}$$

where N_1 and N_0 are the number of treated and untreated observations and i and j are their respective indices. Estimated balancing weights $\hat{w}_j$ are obtained through matching, where the resulting weights satisfy $\sum_{j=1}^{N_0} \hat{w}_j = 1$ and $\hat{w}_j \geq 0$.

Inference   In order to account for the multi-step estimation procedure of PSM, we make use of re-sampling methods for hypothesis testing. In particular, we obtain p-values by bootstrapping the t-statistic with 999 replications, as this has been shown to have better properties than bootstrapping standard errors directly (see Huber *et al.*, 2015; MacKinnon, 2006, for details).

Risk of Hidden Bias   For the CIA to be a valid assumption, X must contain all such variables that simultaneously determine selection into treatment and the outcome of interest (Lechner and Wunsch, 2013). Consequently, if there is some unobserved characteristic U that has an impact

on treatment assignment and the outcome, the CIA will fail. Put formally, $\hat{\tau}_{ATT} \xrightarrow{p} \tau_{ATT} + b$ with $b \neq 0$ if the true treatment probability is given by

$$P_i := Pr(D = 1 \mid X = x_i, U = u_i) = F(x_i'\beta + \gamma u_i) \qquad (4)$$

where $\gamma \neq 0$ and $E[Y^0 \mid X = x, U = u] \neq E[Y^0 \mid X = x, U = u']$ for $u \neq u'$. The size of the inconsistency b depends on the selectivity parameter γ and the responsiveness of Y^0 with respect to U. Since the reform of the NSUS introduced an additional layer of selection, γ may be larger in magnitude and estimates more susceptible to "hidden bias" (see Rosenbaum, 2002, for more details on the problem of "hidden bias"). Thus, careful sensitivity analyses are necessary.

4.1 Specification and Estimation of the Propensity Score

Our extensive dataset allows us to control for a wide range of pre-treatment characteristics. Our baseline propensity score specification includes variables containing information on socio-demographics such as age, gender, health status, German citizenship, marital status and single parent status, number of children and the presence of young children. Human capital attainment is included using the highest schooling degree, professional education and qualification. In order to break the dependence between D and Y^0, it is arguably most important to include a detailed and sufficiently flexible specification of labor market and earnings history. We do this by adding information on short- and long-term unemployment history, short- to medium-term employment and treatment history, the employment status before unemployment, previous occupation, the size of unemployment benefits received as well as last labor earnings. We make the specification flexible by either including categorical dummies for important confounders or – as in the case of previous earnings – using data-driven selection of fractional polynomials (Sauerbrei and Royston, 1999). In addition, we also include a battery of regional characteristics to control for different local labor markets. For this purpose, we include regional dummies as well as explicit control for local macroeconomic conditions and self-employment activity. The baseline specification also includes a number of interaction terms, which were added iteratively to improve subsequent matching quality (see next Section 4.2).

As mentioned in Section 2, another potentially important confounder is the entrepreneurial affinity of individuals. We attempt to proxy for it using some variables that are seen to be important factors in determining the decision to become self-employed, such as previous self-employment status, intergenerational transmission of self-employment as observed through our survey data and the regional controls on start-up activity out of unemployment and the share of

self-employed in the general labor force. Aiming to strike a balance regarding the comparability of our results to most previous evidence for Germany as well as evidence for other countries, we make use of intergenerational information but abstain from including our measures of personality traits in our baseline specification. However, as part of our sensitivity analysis later on, we will extend this standard set of variables with some measures of personality traits or non-cognitive skills since they are likely to be correlated with entrepreneurial affinity. The baseline specification is estimated using a probit regression on the pooled sample, as a Chow test of different selection patterns into treatment for men and women could not be rejected. The details of the specification, estimated coefficients and results from the Chow test can be found in Table A.2. Figure 2 shows the resulting predicted values of the propensity score used to estimate balancing weights in the next step.

[Insert Figure 2 about here]

4.2 Matching to Improve Balance

In our baseline design, we use the estimated propensity score in combination with non-parametric kernel matching with an Epanechnikov kernel to estimate balancing weights.[9] The kernel bandwidth is chosen to maximize post-matching balance.[10] In order to avoid extrapolation, we impose common support by restricting the analysis to the subset of treated individuals who satisfy

$$\left\{ i \in S_1 : \hat{p}(x_i) < \max_j(\hat{p}(x_j) \mid D = 0) \right\}, \tag{5}$$

where S_1 denotes the set of all treated units. Since matching on the propensity score does not control for differences in covariates directly, the appropriateness of the propensity score specification has to be judged against the resulting balancing quality (Rosenbaum and Rubin, 1983). Table 3 provides several commonly-used indicators for the balance achieved.

[Insert Table 3 about here]

Kernel matching dramatically improves in-sample balance as measured by several indicators. There remain no significant mean differences in the matched sample at any traditional level using a t-test of equal means. This is supported by a reduction of the mean absolute standardized bias from 11.5% to 2.3% through matching. Moreover, inspecting the distribution of absolute

---

[9]The matching is performed using the *psmatch2* ado-package by Leuven and Sianesi (2003).

[10]In the spirit of Imai *et al.* (2008), a grid search is performed, choosing the bandwidth that maximizes balance by minimizing the pseudo-R^2 after matching. We found this to be the case for $h = 0.13$.

standardized biases reveals that the number of covariates with standardized biases with relatively large differences is drastically reduced. For example, the number of covariates with a standardized bias above 7% is reduced from 38 to zero. Similarly, the number of covariates with a bias of at least 5% but less than 7% decreased from 28 to just 8 in the matched sample. Following Sianesi (2004), pseudo-R^2 of the propensity score estimation decreases to 1.5% in the matched sample and the null hypothesis of all covariates having no predictive power regarding treatment status cannot be rejected at virtually any significance level. The balancing measures based on the propensity score due to Rubin (2001) also point towards a drastic increase in balancing quality. Rubin's B – defined as the standardized mean difference in the linear index $(x\hat{\beta})$ of the propensity score – decreases from over 100% to 29.3%, while the ratio of the propensity score's variance in the treated and untreated sample (Rubin's R) remains close to one. In addition, quantile-quantile plots for the important pre-treatment outcomes "fraction of time spent in unemployment in the last 10 years" and "last daily earnings" follow the 45-degree line quite closely, indicating successful balancing of the distribution for these important covariates after matching. Overall, balancing quality can be regarded as sufficient to proceed with the outcome analysis.

[Insert Figure 3 about here]

5 Estimates Based on Propensity Score Matching

5.1 Main Estimates

Panel A in Table 4 presents the estimated average treatment effects for participants using our baseline empirical approach as described in Section 4.

[Insert Table 4 about here]

Consistent with the existing literature on SUS for the unemployed, we find persistent and large effects on both the employment probability and monthly net earned income of program participants. Participants are about 28 percentage points more likely to be in self- or regular employment and they earn on average about € 760 more than the matched comparison group at the first interview 20 months after entering the program. Regarding long-term effects, estimates suggest that participants are on average 21.5 percentage points more likely to be self-employed or regular employed 40 months after entry. Effects on net monthly earned income are even greater at the second interview compared to the first one, whereby participants gain around

€ 980 by joining the program. These estimated effects are both statistically significant and economically substantial. The size of the effects is of similar magnitude to what Caliendo and Künn (2011) found for an older SUS program introduced by the "Hartz reforms" in 2003. Compared to estimates for the pre-reform program by Caliendo *et al.* (2016), our point estimates are around 11 percentage points larger with respect to employment effects and about € 250 larger in terms of effects on earned income. Thus, our empirical results may be cautiously interpreted as pointing towards a positive role of the institutional changes regarding the program's effectiveness despite their ambiguous theoretical impact, indicating room for improvement of SUS programs by changing entry conditions and support.

5.2 Effect Heterogeneity

In order to gain further insight into how effects vary with respect to certain pre-treatment characteristics and tease out potential channels through which the program works, we estimate ATTs for sub-groups according to age, education, local GDP per capita and gender.[11] The results are displayed in Figure 4.

[Insert Figure 4 about here]

In general, one can say that the groups that have a particularly low estimated counterfactual probability of being in self- or regular employment 40 months after entering the program display the largest estimated gains from participating. These are the lower educated without a (specialized) higher secondary schooling degree – which grants access to the university system – and workers who are at least 45 years old. Our findings support the view that low-skilled individuals benefit more from participating in SUS programs than high-skilled workers. Larger effects for older workers either point towards more entrepreneurial success among older founders or reflect more difficulties for older workers in finding dependent employment. The estimated effects for individuals residing in areas with relatively high GDP per capita are slightly larger and may be due to better business opportunities in these regions. Gender differences in estimated effects are small, with the long-term effects being marginally larger for men.

---

[11]The entire estimation procedure is repeated for each sub-sample. Balancing indicators and propensity score distributions for the sub-samples are available upon request from the authors. Generally, matching quality is somewhat worse due to smaller sample but still within the recommended range of 3-5% in terms of mean standardized bias as given by Caliendo and Kopeinig (2008).

6 Sensitivity Analysis

In this section, we extensively test the sensitivity of our main results with respect to the implementation of PSM and deviations from the CIA, both within the standard PSM framework and by using an instrumental variable strategy.

6.1 Sensitivity with Respect to Implementation

In order to test whether our baseline results are driven by peculiarities of our chosen matching approach, we check the robustness with respect to more technical details such as the link function $F(\cdot)$ used to estimate the propensity score, the imposition of common support and the matching or weighting algorithm. These choices have been shown to significantly affect the finite sample performance of estimators (e.g. see Huber *et al.*, 2015; Lechner and Strittmatter, 2017). Our findings of this analysis are shown in Panels B to D of Table 4 and can be summarized by stating that none of these discretionary choices regarding the implementation of PSM have any economically-significant effect on point estimates. The choice of the link function has very little effect on our estimates, even though the robit regression – which makes use of the t-distribution with optimally-chosen degrees of freedom ν – yields quite different predicted propensity scores, as can be seen in Figure A.1 (see Liu, 2005, for details on robit regression). The imposition of common support through defining a minimum density ($\hat{f}(\hat{p}) \geq c$) of comparison individuals as done by Heckman *et al.* (1997) or restricting the analysis to an optimal interval $[0, \alpha^*]$ as proposed by Crump *et al.* (2009) yields estimates very close to our baseline estimates.[12] Furthermore, different choices of matching or weighting algorithms also do not play a crucial role for our results. For comparison, we tried pair matching with replacement, radius matching with bias adjustment based on Lechner *et al.* (2011) and inverse probability weighting with weights rescaled to unity.[13] The latter two were chosen as they have been found to perform well in Monte-Carlo simulations on finite sample properties by Busso *et al.* (2014) and Huber *et al.* (2013). Point estimates are very similar across these estimators. While the results are robust to alterations in the design phase of our study, the conclusions drawn depend on the applicability of the CIA, which we aim to assess in the next sub-sections.

---

[12]The interval derived by Crump *et al.* (2009) is optimal in the sense that it minimizes the asymptotic variance of matching estimators. Choosing α involves a trade-off: larger α reduce imbalance and extrapolation leading to lower variance, while discarding information increases variance. As software implementation, we use their accompanying *optselect* package to obtain α.

[13]The radius matching with bias adjustment is implemented using the *radiusmatch* package of Huber *et al.* (2015).

6.2 Robustness to the Inclusion of Non-Cognitive Skills

Recent literature has found that measures of personality traits and non-cognitive skills significantly correlate with labor market outcomes and that non-cognitive skills are about as important in determining wages as cognitive abilities (Heckman *et al.*, 2006). Furthermore, Caliendo *et al.* (2014) show that personality traits are associated with the decision to become and remain self-employed. Thus, these types of usually-unobserved variables are potentially important but omitted confounders that help us to proxy for entrepreneurial affinity. Available through the survey data, we include measures of the individuals' characteristics like the Big Five personality traits, locus of control, risk attitudes, impulsiveness, patience and general self-efficacy in the estimation of the propensity score. Doing so increases the pseudo-R^2 of the probit estimation markedly from about 20% to 32%. Thus, differences in personality and non-cognitive skills explain a relatively large part of selection into treatment. For details on estimated probit coefficients, see Table A.2. The resulting propensity score distribution can be seen in Figure A.2. If incorporating these variables into the propensity score estimation significantly changes the resulting treatment effects estimates, this would hint towards a violation of the CIA for our baseline results. However, Panel C in Table 5 shows that the point estimates barely change compared to our baseline estimates and the estimated effects are still highly significant at all conventional levels. These findings are also consistent with those of Caliendo *et al.* (2016), who analyze the interplay of personality traits and the effects of SUS for a sample of participants of the pre-reform NSUS program with fewer layers of selection at play in much more detail.[14]

[Insert Table 5 about here]

6.3 Robustness to Time-Invariant Unobserved Heterogeneity

The longitudinal nature of our outcome data also allows us to control for time-constant unobserved confounders by means of conditional difference-in-differences (CDID) (see e.g. Blundell and Costa Dias, 2009). The CDID estimator combines the difference-in-differences approach with matching to control for observed characteristics. We choose symmetric differencing around the date of entry, following Chabé-Ferret (2015), who finds that CDID estimators perform the best under this setup.[15] Since the CDID approach requires a weaker form of the conditional

---

[14]Interestingly, their results suggest a lesser role of personality traits for selection into SUS, which may indeed indicate more severe selection into treatment through caseworkers after the reform.

[15]One additional finding of Chabé-Ferret (2015) is that it is advisable not to condition on pre-treatment characteristics in the matching process when using CDID. However, for our application, this does not make any significant difference.

independence assumption, this provides a test for the applicability of the original CIA defined in (2). For the CDID estimator to give consistent results, the individual time difference in Y^0 must be independent of treatment when conditioning on the propensity score. Formally, it is required that

$$(Y^0_{t+k} - Y^0_{t-k}) \perp\!\!\!\perp D \mid P(X), \tag{6}$$

where k is the number of months before or after (hypothetical) entry t into the program. Our results are shown in Panel C of Table 5. As becomes readily apparent, controlling for unobserved time-invariant heterogeneity via conditional differences barely affects our estimates. Thus, the results also do not seem to be very sensitive with respect to this kind of deviation from the CIA.

6.4 Assessing Sensitivity Using Bounding Analysis

In this part, we follow Rosenbaum (2002) and test the sensitivity of our inference with respect to the degree of departure from the CIA by using a bounding approach. Let Γ denote the ratio of the odds of receiving treatment for two observationally-identical individuals i and j, but different unobserved characteristics U, then

$$\frac{1}{\Gamma} \le \frac{P_i/(1-P_i)}{P_j/(1-P_j)} \le \Gamma, \tag{7}$$

where $P_i = F(x'_i\beta + \gamma u_i)$ and $\Gamma = e^\gamma$ for the case of a logistic $F(\cdot)$. The bounding exercise essentially varies γ and thus Γ and tests whether the estimated effect remains significant at that level of "hidden bias". In our application, we assume that we have over-estimated the true effect and gradually increase Γ until we obtain the value for which our estimates turn insignificant.[16] Thus, if this critical value is large, our estimates are relatively robust with respect to deviations from the CIA. Panel D of Table 5 gives the critical Γ^*s for our four outcome variables of interest. Generally, inference with respect to employment prospect is more robust than earnings. The smallest Γ^* that we obtain is for net monthly earnings after 40 months, which is equal to roughly 2.5, meaning that an unmeasured covariate would need to increase the odds of receiving treatment by the factor of 2.5 compared to someone without this characteristic to turn our conclusions invalid. Hence, our results indicating positive long-term effects on employment and income are very robust with respect to general unobserved confounders.

---

[16]For binary outcomes, we use the *mhbounds* package by Becker and Caliendo (2007) and for continuous outcomes, *rbounds* is employed as described by DiPrete and Gangl (2004).

6.5 Estimates Using an Instrumental Variable Approach

Should the CIA indeed fail in our application – despite the evidence presented so far – one can still estimate average treatment effects under the condition that we find exogenous variation in the treatment probability. In this section, we aim to estimate treatment effects using an instrumental variable strategy based on regional variation in the likelihood of receiving treatment, using both the standard two-stage least squares (2SLS) estimator and the semi-parametric approach by Frölich (2007). For a dummy instrument Z, both estimators can be displayed as

$$\tau_{IV}(x) = \frac{E[Y \mid X = x, Z = 1] - E[Y \mid X = x, Z = 0]}{E[D \mid X = x, Z = 1] - E[D \mid X = x, Z = 0]},$$ (8)

where 2SLS conditions on X through linear regression and the IV-matching approach by conditioning on the scalar propensity score $\tilde{P}(Z = 1 \mid X) = \tilde{P}(X)$. The latter estimator is used to test the sensitivity of the IV estimates with respect to the inherent linearity assumption.

For the construction of an instrument, we make use of the regional discretionary power of LEAs after the reform to allocate their allotted funds with respect to different ALMPs. If – conditional on local labor market conditions – an LEA makes stronger use of a certain ALMP, this can be regarded as having a stronger preference for this type of program. Our proxy for the regional preference for SUS is the ratio of approved applications to the total number of applicants in the same region during our sampling time frame, albeit in months other than the month of entry.[17] We call this the leave-one-month-out approval rate, or approval rate for short. Dropping the individual's month of (hypothetical) entry should purge the instrument from a direct relationship with the individual's characteristics. Figure 5 shows the spatial distribution of approval rates across the 178 LEAs in Germany, both unconditional and conditional on local labor market conditions.

[Insert Figure 5 about here]

A number of identifying assumptions need to be fulfilled for our IV approach to give consistent estimates. However, even if the assumptions are true, the IV estimates in general only yield a local average treatment effect for the part of the population that changes treatment status due

---

[17]There are several reasons why we are constrained to contemporaneous data for the instrument. First, data from the previous year corresponds to the pre-reform program and thus measures the preference for a non-existent program. Second, data from the month of January 2012 (the first month after the reform took place) cannot be used as the number of approved applications is contaminated by applicants from before the reform. Third, data after our sampling time frame cannot be used as there was a reform of LEA districts, which led to the disappearance of 22 LEAs. The data on applications for the program and actual entries are obtained from administrative data from the FEA.

to the instrument, called compliers (see Imbens and Angrist, 1994).

IV Identifying Assumptions   First, the instrument needs to be *relevant*, i.e. the instrument must have an impact on the likelihood of receiving treatment. It can be assumed that relevance is fulfilled if the instrument has an influence in the first stage, i.e. when the denominator of (8) is significantly different from zero. Second, conditional on X (or $\tilde{P}(X)$), the instrument Z must satisfy independence with respect to D and Y. Put formally, it is required that

$$Y(z), D(z) \perp\!\!\!\perp Z \mid X, \tag{9}$$

where $Y(z)$ and $D(z)$ denote the observed outcomes and treatment status, both dependent on the value of the instrument Z. This implies that – conditional on X – the instrument is as good as randomly assigned and it does not have an effect on Y that does not go through D (*exclusion restriction*). Third, there must be *no defiers*, implying that $D(z_1) \geq D(z_0)$ for values of the instrument $z_1 > z_0$.

Plausibility of Assumptions   Apart from the relevance condition, the other identifying assumption cannot be directly tested empirically and needs to be discussed. Assumption (9) is also called the exogeneity assumption in linear regression and it is often put in terms of correlations: once we condition on X, Z must not be correlated with relevant omitted factors U (e.g. entrepreneurial affinity). The credibility of this assumption thus depends on the richness of controls. We largely employ the same specification as described in Section 4.1.[18] Since the regional controls hold particular importance in this case, we control for the geographic location of the individual and local labor market conditions. The former are implemented using dummy variables for northern, eastern and southern Germany, while the latter include measures of the unemployment rate, the vacancy-to-unemployed ratio, GDP per capita and – probably most importantly – the start-up rate out of unemployment and the overall self-employment rate using flexible categorical dummies to soak up as much variation due to different local labor market conditions and the local tendency to become self-employed within geographic regions as possible. However, the assumption (9) would still fail if LEAs with large *conditional approval rates* tended to select individuals with lower entrepreneurial affinity into treatment. Of course, we cannot test for this. However, we can test for a conditional correlation of the instrument with the previously-

---

[18]The only difference is that we drop interaction terms as these were only included to further improve balance in X across treatment groups D. This choice does not affect our IV results in any significant manner. Results with the interaction terms included can be obtained from the authors on request.

mentioned personality traits, which are very predictive of receiving treatment. For this purpose, we run an auxiliary regression of the instrument on personal and regional controls as well as our measures of personality traits.[19] The null hypothesis of zero conditional correlation between personality traits and the instrument cannot be rejected at any traditional significance level, supporting the case for the validity of the independence assumption. The exclusion restriction also seems quite natural to us, as we would only expect an effect of regional participation in SUS on individuals' employment prospects and earnings if there are spill-over effects of treatment, in which case the local stock of participants would be a relevant regressor in the outcome equation, but not regional approval rates. Furthermore, the no defier assumption seems plausible since higher approval rates should weakly increase treatment receipt for everyone.

Estimates   Table 6 provides the results from our IV estimation. Panel A gives the results using 2SLS in combination with a continuous instrument. Panels B and C show results for a dummy instrument, coded as one if the person lives in an LEA with a leave-one-month-out approval rate larger than the median[20]. While panel B still makes use of 2SLS, Panel C uses the IV-matching estimator by Frölich (2007). The latter can be regarded as the ratio of two matching estimators, thus avoiding functional form restrictions regarding the impact of the instrument on treatment probability and outcomes.

[Insert Table 6 about here]

First-stage estimates suggest a significant impact of the instrument on treatment receipt. This is supported by the z-statistics as they are well above the weak instrument threshold of $\sqrt{10} \approx 3.17$ as given by Staiger and Stock (1997). Focusing on the 2SLS estimates based on the continuous instrument first, we find that a one percentage point increase in the approval rate increases the likelihood of receiving treatment by 0.3 percentage points. Similar to our baseline matching results, second-stage estimates suggest large and positive effects on both employment and earnings. The long-term effect on the probability of being in self- or regular employment is estimated to be about 29 percentage points and thus actually larger than our matching estimates. Effects on earnings are comparable to the matching results, albeit only being significant at the 10% level. Turning to estimates based on the dummy version of the instrument, we find a similar pattern, although the effects are somewhat smaller and less significant when using 2SLS. The first-stage coefficient gives us a direct measure of the size of the complier population, which

---

[19]Coefficients on personality traits and test results from the auxillary regression are shown in Table A.3.
[20]The median corresponds to roughly a 50% leave-one-month-out approval rate.

is estimated to be around 10% when using 2SLS and 8.6% using the IV-matching estimator. Comparing second-stage results, we find very similar results to the 2SLS estimates, albeit they are insignificant due to the higher variability. Hence, the linearity assumption of the 2SLS approach does not seem too restrictive in our application. Overall, our IV approach also suggests positive effects on employment probabilities and earnings for the sub-group of compliers. Making a comparison with our matching results is difficult as the IV approach identifies a different parameter, but matching estimates are included in IV confidence bands.[21]

7 Conclusion

In this paper, we provide the first long-term evidence on the causal effects of the new German SUS program for the unemployed after its reform in 2011. The reform significantly altered the institutional setup of the program, leading to ambiguous predictions on the post-reform effectiveness of the subsidy, e.g. due to uncertainty regarding the effects of shortening the first benefit period, during which the bulk of transfers are paid to participants. Our main results based on PSM techniques suggest that effects on employment probabilities and earned income (up to 40 months after entering the program) are positive, economically important and larger compared to the previous program. Thus, there seems to be room for improvement in terms of SUS effectiveness by altering design features of the programs such as the duration of support. The analysis of effect heterogeneity indicates that the program is especially beneficial for older and low-skilled workers.

Since the reform of the program introduced an additional layer of selection by increasing caseworkers' discretionary power, it is necessary to critically assess identification assumptions used to estimate the treatment effects. Hence, we assess the sensitivity of our conclusions with respect to deviations from the CIA within the matching framework and also by using an instrumental variable strategy that exploits regional variation in the likelihood of receiving treatment induced by the reform. Our results of the sensitivity analyses indicate that the matching estimates are very robust to deviations from the CIA and the IV results also point towards large positive effects. Since SUS programs in other countries operate in a similar manner in terms of both selection and support, our findings of robust positive effects may lend some credibility to other matching estimates in the literature.

While our microeconometric estimation approach provides evidence on individual-level effects of SUS for previously-unemployed participants, there may be spill-over or general equilibrium

---

[21]To see this, compare the square brackets in Table 6 with the baseline results from Table 4.

effects of such a program that our empirical strategy is unable to identify. On the one hand, SUS programs may have negative spill-overs; for example, by displacing other regular businesses due to a competetive advantage of subsidzed businesses. On the other hand, SUS may also lead to positive spill-overs, e.g. by also leading to job creation for other unemployed jobseekers. Future research should aim to identify these potentially important effects of SUS as this would allow a more thorough analysis of benefits and costs of SUS. Furthermore, it would be interesting to experimentally validate the individual-level results of the observational studies conducted thus far. Given the large number of applications that had to be rejected due to budgetary reasons, this seems a natural way to proceed and would potentially also allow testing different design features (e.g. duration of support). This could help to learn more about the optimal design of SUS for the unemployed.

Compliance with Ethical Standards: Marco Caliendo declares that he has no conflict of interest. Stefan Tübbicke declares that he has no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

References

BECKER, S. and CALIENDO, M. (2007). Sensitivity analysis for average treatment effects. *The Stata Journal*, **7** (1), 71–83.

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.

BELLMANN, L., CALIENDO, M. and TÜBBICKE, S. (2018). The Post-Reform Effectiveness of the New German Start-Up Subsidy for the Unemployed. *LABOUR*, **32** (3), 293–319.

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

BLUNDELL, R. and COSTA DIAS, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, **44** (3), 565–640.

BUSSO, M., DINARDO, J. and MCCRARY, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *The Review of Economics and Statistics*, **96** (5), 885–897.

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

—, FOSSEN, F. and KRITIKOS, A. (2014). Personality characteristics and the decisions to become and stay self-employed. *Small Business Economics*, **42** (4), 787–814.

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

— 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.

CHABÉ-FERRET, S. (2015). Analysis of the bias of Matching and Difference-in-Difference under alternative earnings and selection processes. *Journal of Econometrics*, **185** (1), 110–123.

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.

DIPRETE, T. and GANGL, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology*, **34**, 271–310.

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.

DUNN, T. and HOLTZ-EAKIN, D. (2000). Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links. *Journal of Labor Economics*, **18**, 282–305.

FRÖLICH, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics*, **139** (1), 35–75.

HECKMAN, J. J., ICHIMURA, H. and TODD, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, **64** (4), 605–654.

— and ROBB, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, **30** (1), 239–267.

—, STIXRUD, J. and URZUA, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, **24**, 411–482.

HUBER, M., LECHNER, M. and STEINMAYR, A. (2015). 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.

IMAI, K., KING, G. and STUART, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A*, **171** (2), 481–502.

IMBENS, G. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** (2), 467–75.

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

LECHNER, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica-Verlag HD, Heidelberg, pp. 43–58.

—, MIQUEL, R. and WUNSCH, C. (2011). Long-Run Effects of Public Sector Sponsored Training in West Germany. *Journal of the European Economic Association*, **9** (4), 742–784.

— and STRITTMATTER, A. (2017). *Practical Procedures to Deal with Common Support Problems in Matching Estimation*. Discussion Papers 10532, Institute for the Study of Labor (IZA), Bonn.

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

LEUVEN, E. and SIANESI, B. (2003). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*. Tech. rep., Statistical Software Components S432001, Boston College Department of Economics, revised 30.April 2004.

LINDQUIST, M., SOL, J., VAN PRAAG, M. and VLADASEL, T. (2016). *On the Origins of Entrepreneurship: Evidence from Sibling Correlations*. CEPR Discussion Papers 11562, C.E.P.R. Discussion Papers.

LIU, C. (2005). *Robit Regression: A Simple Robust Alternative to Logistic and Probit Regression*, John Wiley & Sons, Ltd, pp. 227–238.

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

OECD (2015). *Public Expenditure and Participant Stocks on LMP*. OECD.Stat Database.

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.

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.

ROSENBAUM, P. R. (2002). *Observational Studies*. New York: Springer, 2nd edn.

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.

SAUERBREI, W. and ROYSTON, P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **162** (1), 71–94.

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.

STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65** (3), pp. 557–586.

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

WOLFF, J., NIVOROZHKIN, A. 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: Selected Descriptives for Control Variables

	Part.	Non-Part.	p-value
Socio-demographics			
Age at (hypothetical) entry	42.77	43.94	0.002
Less than 25 years	0.012	0.006	0.103
25 to less than 35 years	0.206	0.192	0.383
35 to less than 45 years	0.336	0.322	0.478
45 to less than 56 years	0.195	0.191	0.817
56 years and older	0.252	0.289	0.037
Female	0.425	0.509	0.000
Human capital			
Highest schooling degree			
Lower secondary school	0.107	0.198	0.000
Middle secondary school	0.278	0.376	0.000
Upper secondary school (specialized)	0.170	0.128	0.004
Upper secondary school (general)	0.433	0.282	0.000
No schooling degree	0.013	0.016	0.537
Labor market history			
Fraction of time in unemployment in the last 10 years	0.101	0.170	0.000
One year before (hypothetical) entry			
Months employed	7.773	6.699	0.000
Months in labor market program	0.401	0.390	0.834
Employment status before entering unemployment			
Dependent employment	0.674	0.516	0.000
Self-employment/family worker	0.054	0.012	0.000
School/apprenticeship	0.017	0.020	0.567
Disable to work/unemployable	0.033	0.135	0.000
Other	0.081	0.263	0.000
Daily income from last employment (Euro)	81.39	61.60	0.000
Duration of last unemployment spell (months)	3.808	4.390	0.010
Intergenerational information			
Father and/or mother was born abroad	0.151	0.177	0.079
Father and/or mother is/was self-employed	0.349	0.252	0.000
Father employed when respondent 15 years old	0.913	0.855	0.000
Geographic region			
Northern Germany	0.196	0.167	0.059
Eastern Germany	0.333	0.374	0.033
Southern Germany	0.155	0.182	0.080
Local macroeconomic conditions			
Local unemployment rate in %	7.905	7.927	0.853
Ratio of vacancies to unemployed	16.92	17.24	0.439
GDP per capita in 2011	31.26	30.67	0.088
Local start-up rate out of unemployment	0.058	0.058	0.932
Local self-employment rate	0.112	0.114	0.002
Personality traits			
Big Five			
Conscientiousness	4.383	4.106	0.000
Extraversion	3.692	3.297	0.000
Agreeableness	3.401	3.368	0.300
Neuroticism	2.365	2.677	0.000
Openness	3.723	3.255	0.000
Other personality traits			
Readiness to take risks	5.777	5.316	0.000
Locus of control	4.260	3.785	0.000
Patience	6.370	6.380	0.911
Impulsiveness	5.388	5.409	0.809
General self-efficacy	4.391	4.062	0.000
Number of observations	1,248	1,204	

Note: Reported are sample shares for dummy variables and means for multi-valued variables. p-values are based on t-tests of equal means.

Table 2: Descriptives for Outcome Variables

	Part.	Non-Part.	p-value
Labor market outcomes after 20 months			
Main labor market status			
Self-employed	0.884	0.037	0.000
Self or regular employed	0.958	0.615	0.000
Unemployed	0.021	0.208	0.000
Earnings			
Net monthly earnings (Euro)	1,901.2	886.6	0.000
Labor market outcomes after 40 months			
Main labor market status			
Self-employed	0.804	0.038	0.000
Self or regular employed	0.938	0.676	0.000
Unemployed	0.022	0.110	0.000
Earnings			
Net monthly earnings (Euro)	2,264.4	1,046.3	0.000
Number of observations[a]	1,248	1,204	

Note: Reported are sample shares for labor market status variables and means for the earnings variables. p-values are based on t-tests of equal means. [a]: The number of observations for the earnings variables is slightly lower due to item non-response.

Table 3: Balancing Quality

	Before Matching	After Matching
Number of variables with significant differences in means[a]		
at 1%-level	38	0
at 5%-level	50	0
at 10%-level	58	0
Number of variables with absolute standardized bias[b]		
0 to less than 1%	9	29
1 to less than 3%	9	35
3 to less than 5%	12	24
5 to less than 7%	28	8
more than 7%	38	0
Mean absolute standardized bias in %	11.47	2.26
(Re-)Estimation of the propensity score[c]		
pseudo-R^2	.204	.015
p-value of joint-significance test	0.000	0.999
Other measures		
Rubin's B[d]	111.0	29.3
Rubin's R[e]	1.00	1.37
Number of variables	96	96
Number of participants off support		58

Note: Different indicators are shown for covariate balancing before and after Epanechnikov kernel matching using a bandwidth of 0.13.
[a]: The 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 4: Baseline Results and Sensitivity Analyses with Respect to Implementation

	Outcomes after 20 months		Outcomes after 40 months	
	Self- or regular employment	Net monthly earned income	Self- or regular employment	Net monthly earned income
A. Baseline results:				
	0.280***	764.2***	0.215***	980.2***
	(0.000)	(0.000)	(0.000)	(0.000)
B. Sensivitiy with respect to choice of CDF $F(\cdot)$:				
Logit	0.279***	762.6***	0.213***	977.3***
	(0.000)	(0.000)	(0.000)	(0.000)
Robit	0.278***	787.6***	0.208***	979.7***
	(0.000)	(0.000)	(0.000)	(0.000)
C. Sensitivity with respect to common support imposition:				
$f(\hat{p}) > 0.2$	0.283***	739.6***	0.218***	948.2***
	(0.000)	(0.000)	(0.000)	(0.000)
$[0; \alpha^*]$	0.281***	748.8***	0.216***	959.0***
	(0.000)	(0.000)	(0.000)	(0.000)
D. Sensitivity with respect matching/weighting algorithm:				
NN-Matching	0.285***	803.1***	0.239***	909.7***
	(0.000)	(0.000)	(0.000)	(0.000)
IPW	0.283***	794.5***	0.216***	990.1***
	(0.000)	(0.000)	(0.000)	(0.000)
LMW Radius	0.283***	784.2***	0.234 ***	933.1***
	(0.000)	(0.000)	(0.000)	(0.000)

Note: The table shows ATT estimates using different PSM approaches. *p*-values are shown in parentheses underneath the point estimate, obtained by bootstrapping the *t*-statistic using 999 replications (MacKinnon, 2006; Huber *et al.*, 2015). ***/**/* denote significance at the 1/5/10 % level.

Panel A shows our baseline results using kernel matching, common support imposition via min/max criterion using a probit regression. Panel B tests the sensitivity with respect to the parametric regression of the propensity score. Panel C tests whether the way in which we impose common support affects our results significantly, defining a minimum density $f(\hat{p})$ in the comparison group (Heckman *et al.*, 1997) or choosing an optimal common support interval $[0; \alpha^*]$ (Crump *et al.*, 2009). Finally, Panel D tests the robustness regarding different estimators: $k = 1$ nearest neighbor matching, inverse probability weighting with standardized weights and radius matching with bias adjustment due to Lechner, Miquel and Wunsch (2011).

Table 5: Sensitivity Regarding CIA – PSM

	Outcomes after 20 months		Outcomes after 40 months	
	Self- or regular employment	Net monthly earned income	Self- or regular employment	Net monthly earned income
A. Baseline results:				
	0.280***	764.2***	0.215***	980.2***
	(0.000)	(0.000)	(0.000)	(0.000)
B. Including personality traits:				
	0.271***	735.0***	0.201***	915.8***
	(0.000)	(0.000)	(0.000)	(0.000)
C. Conditional DiD:				
	0.287***	666.8***	0.255***	914.7***
	(0.000)	(0.000)	(0.000)	(0.000)
D. Critical Γs for Rosenbaum bounds:				
	12.63	2.73	6.19	2.49

Note: The table shows ATT estimates based on PSM and results from bounding analyses to assess the sensitivity of the estimates regarding unobserved confounders. *p*-values for ATT estimates are shown in parentheses underneath the point estimate, obtained by bootstrapping the *t*-statistic using 999 replications (MacKinnon, 2006; Huber *et al.*, 2015). ***/**/* denote significance at the 1/5/10 % level.

Panel A shows our baseline results. Panel B includes usually unobserved personality traits into the score specification. Panel C tests the sensitivity of our results with respect to time-constant unobserved confounders by means of conditional DiD. Finally, Panel D gives the factor Γ by which the odds of receiving treatment between observationally identical individuals must differ to overturn our inference by using Rosenbaum (2002) bounds.

Table 6: Sensitivity Regarding CIA – IV Approach

	First Stage	Second Stage			
	At entry	After 20 months		After 40 months	
	SUS receipt	Self- or regular employment	Net monthly earned income	Self- or regular employment	Net monthly earned income
A. 2SLS (continuous):					
	0.003***	0.413***	792.7*	0.291**	1,002.6*
	(5.937)	(3.114)	(1.956)	(2.213)	(1.905)
		[.195;.631]	[126.0; 1,459.3]	[.075;.507]	[134.5; 1,870.6]
B. 2SLS (dummy):					
	0.098***	0.342**	880.0*	0.290*	730.5
	(4.705)	(2.094)	(1.709)	(1.764)	(1.069)
		[.073;.611]	[33.0;1,727.0]	[.020;.560]	[-393.6; 1,854.6]
C. IV-Matching:					
	0.086***	0.326	849.5	0.338	861.4
	(3.728)	(1.476)	(1.162)	(1.466)	(0.885)
		[-.037; .689]	[-353.1; 2,052.1]	[-.041; .717]	[-739.7; 2,462.5]

Note: The table shows first-stage coefficients as well as LATE estimates using the IV approach. z-statistics are shown in parentheses underneath the point estimate. 2SLS standard errors are heteroscedasticity robust. Inference for IV matching is based on standard errors estimated by bootstrap using 999 replications (MacKinnon, 2006). ***/**/* denote significance at the 1/5/10 % level. 90% confidence bands are shown in brackets. The first-stage passes the weak-instrument test by Staiger and Stock (1997) if the respective z- statistic is larger than $\sqrt{10} \approx 3.17$.
Panel A shows the baseline IV estimates using 2SLS and the continuous leave-one-month-out (lomo) approval rate. Panel B also makes use of 2SLS but employs a dummy-instrument that equals one if the lomo approval rate is larger than the median of the distribution. Panel C uses the same dummy instrument but employs the semi-parametric IV matching estimator due to Frölich (2007).

Figure 1: Participants in SUS in OECD Countries 2015
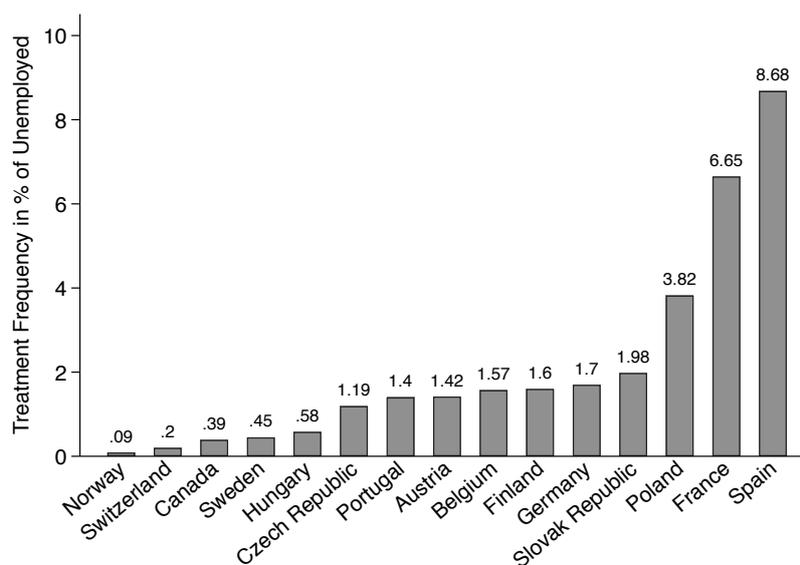
Source: OECD (2015), own calculations.

29

Figure 2: Propensity Score Distribution – Baseline Specification
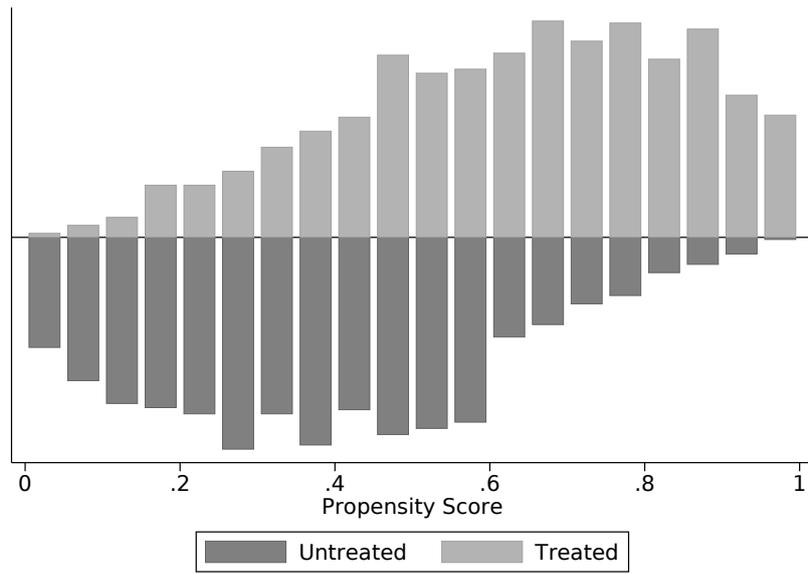
Note: This graph shows the distribution of estimated propensity scores for the treated and comparison group using a probit regression based on the baseline specification including information on socio-demographics, human capital, labor market history, intergenerational transmission, and regional controls for local labor market conditions and self-employment activity. For details on the specification and estimated coefficients, see Table A.2.

Figure 3: Graphical Analysis of Balancing Quality

QQ-Plot: Last Daily Earnings

QQ-Plot: Fraction in Unemployment in last 10 years

● Before matching   ◆ After matching
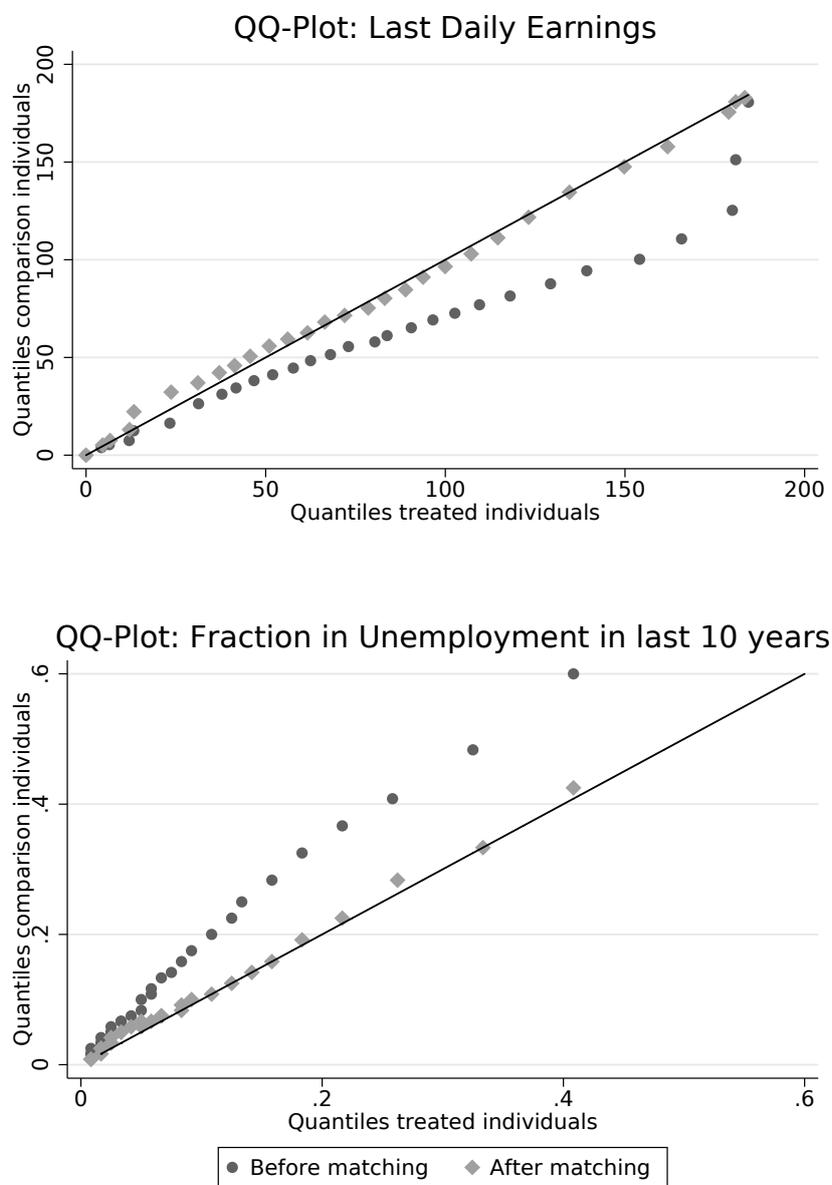
Note: This figure plots distribution quantiles for the treated against those of the untreated, both for the raw and the matched sample. Scatter dots following the 45 degree line indicate covariate balance for continuous variables.

Figure 4: Effect Heterogeneity Regarding Employment Effects
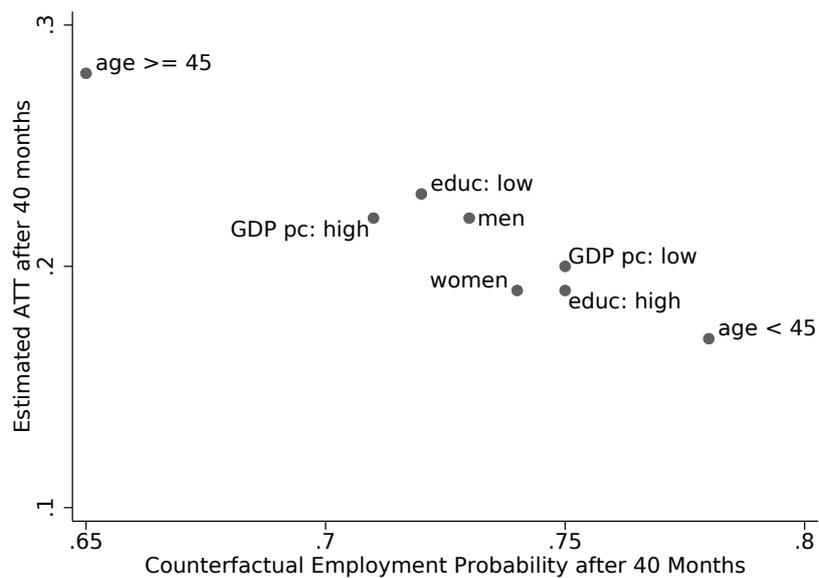
Note: This graph plots the estimated ATTs against the estimated counterfactual means $\hat{E}[Y^0 \mid D = 1]$ for sub-samples. Results are obtained by repeating the steps of the main analysis for each sub-sample seprately. Sample splits are performed based on binary indicators regarding age (age$\geq$ 45 or not), education (=high if individual has a (specialized) higher secondary school degree), GDP per capita (= high if the individual lives in a region with above median GDP per capita) and gender.

Figure 5: Regional Approval Rates

A. Approval Rates

B. Conditional Approval Rates

95th to 100th %tile
90th to 95th %tile
75th to 90th %tile
50th to 75th %tile
25th to 50th %tile
10th to 25th %tile
5th to 10th %tile
0 to 5th %tile
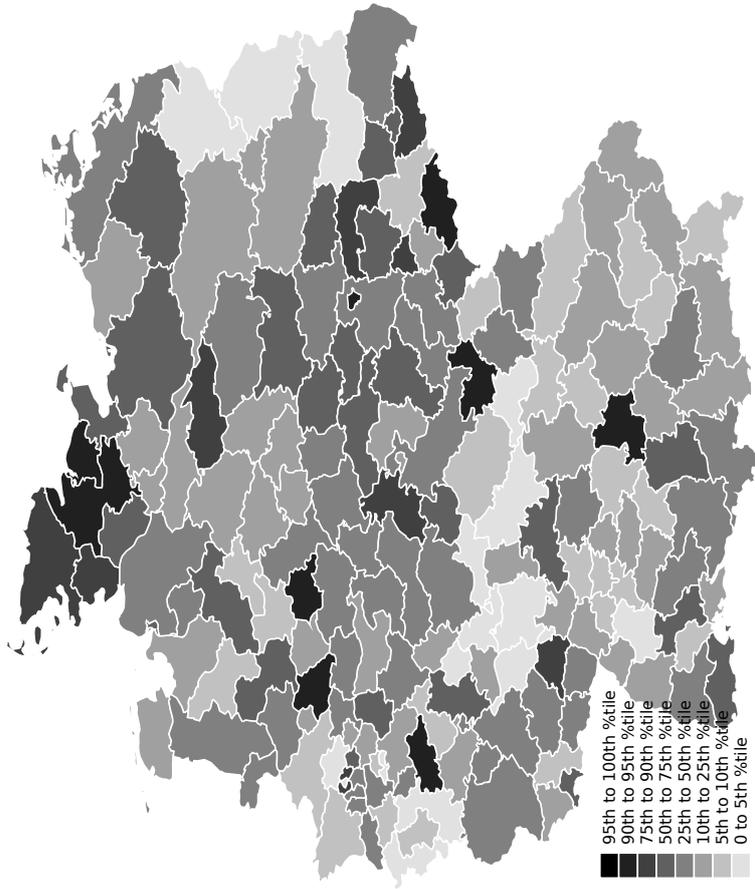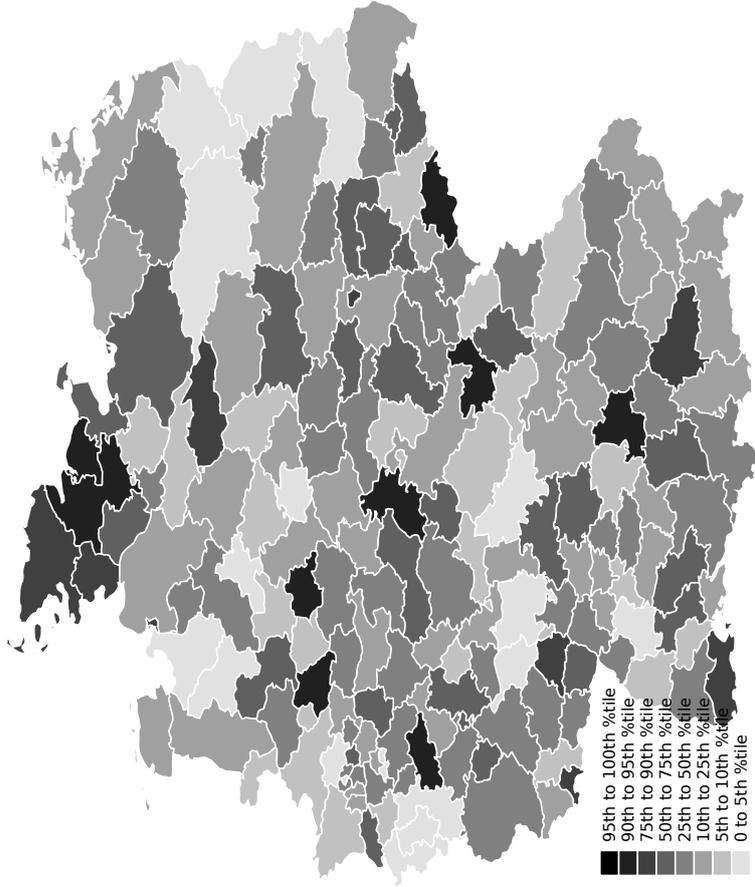
Note: This figure shows the geographical distribution of the regional of approval rates across the 178 Local Employment Agencies. Panel A depicts the regional distribution of the raw data and Panel B shows the distribution of the conditional approval rates, i.e. $E(Z \mid$ regional characteristics). Regional characteristics include dummies for northern, eastern and southern Germany and categorical dummy variables for the local unemployment rate, the ratio of vacancies to unemployed, the regional GDP per capita, regional self-employment rates and the start-up rate out of unemployment. The conditional approval rate is interpreted as a measure of the LEA's preference for start-up subsidies as the ALMP of choice.

33

A Supplementary Tables and Figures

Table A.1: Descriptive Statistics for Control Variables

	Participants	Non-Participants	p-value
Socio-demographics			
Age at (hypothetical) entry (years)	42.77	43.94	0.002
Less than 25 years	0.012	0.006	0.103
25 to less than 35 years	0.206	0.192	0.383
35 to less than 45 years	0.336	0.322	0.478
45 to less than 56 years	0.195	0.191	0.817
56 years and older	0.252	0.289	0.037
Female	0.425	0.509	0.000
Not German citizen	0.038	0.017	0.002
Health restrictions	0.038	0.058	0.017
Married	0.599	0.581	0.388
Number of children			
No child	0.571	0.515	0.006
One child	0.218	0.255	0.031
Two children and above	0.212	0.230	0.269
Children under 10 present	0.243	0.218	0.139
Single parent	0.054	0.049	0.600
Human capital			
Highest schooling degree			
Lower secondary school	0.107	0.198	0.000
Middle secondary school	0.278	0.376	0.000
Upper secondary school (specialized)	0.170	0.128	0.004
Upper secondary school (general)	0.433	0.282	0.000
No schooling degree	0.013	0.016	0.537
Professional education			
Vocational training	0.371	0.576	0.000
Professional/vocational academy	0.153	0.090	0.000
College/university degree	0.411	0.267	0.000
Other/no training	0.065	0.065	0.990
Professional qualification			
Unskilled workers	0.566	0.479	0.000
Skilled workers	0.258	0.381	0.000
Skilled workers with technical college education	0.035	0.031	0.531
Top management	0.141	0.109	0.016
Labor market history			
Fraction of time in unemployment in the last 10 years	0.101	0.170	0.000
Less than 10%	0.673	0.466	0.000
10 to less than 20 %	0.175	0.227	0.001
20 to less than 40 %	0.118	0.199	0.000
40 to less than 60 %	0.027	0.073	0.000
more than 60 %	0.007	0.035	0.000
Five years before (hypothetical) entry			
Months employed	8.841	8.626	0.266
Months in labor market program	0.326	0.534	0.002
Four years before (hypothetical) entry			
Months employed	9.252	8.921	0.069
Months in labor market program	0.368	0.518	0.030

	Participants	Non-Participants	*p*-value
Three years before (hypothetical) entry			
Months employed	9.825	9.158	0.000
Months in labor market program	0.261	0.500	0.000
Two years before (hypothetical) entry			
Months employed	10.36	9.452	0.000
Months in labor market program	0.191	0.403	0.000
One year before (hypothetical) entry			
Months employed	7.773	6.699	0.000
Months in labor market program	0.401	0.390	0.834
Employment status before entering unemployment			
Dependent employment	0.674	0.516	0.000
Self-employment/family worker	0.054	0.012	0.000
School/apprenticeship	0.017	0.020	0.567
Disable to work/unemployable	0.033	0.135	0.000
Other	0.081	0.263	0.000
Occupational group before entering unemployment			
Manufacturing	0.026	0.027	0.883
Forestry, fishing, animal breeding	0.127	0.204	0.000
Technical profession	0.080	0.056	0.020
Services	0.765	0.711	0.002
Other	0.002	0.001	0.585
Daily income from last employment (Euro)	81.39	61.60	0.000
Duration of last unemployment spell (months)	3.808	4.390	0.010
Less than 1 month	0.104	0.027	0.000
1 to less than 3 months	0.388	0.383	0.802
3 to less than 6 months	0.280	0.319	0.034
6 to less than 12 months	0.192	0.224	0.051
12 to less than 24 months	0.027	0.038	0.127
24 months and above	0.009	0.008	0.891
Monthly unemployment benefit (Euro)	1,086.6	927.0	0.000
Less than 300 Euros	0.090	0.043	0.000
300 to less than 600 Euro	0.127	0.196	0.000
600 to less than 900 Euro	0.201	0.321	0.000
900 to less than 1200 Euro	0.184	0.207	0.160
1200 to less than 1500 Euro	0.145	0.118	0.047
1500 Euro and above	0.253	0.115	0.000
Intergenerational information			
Father and/or mother was born abroad	0.151	0.177	0.079
Father and/or mother is/was self-employed	0.349	0.252	0.000
Father employed when respondent 15 years old	0.913	0.855	0.000
Professional education of father			
Vocational training	0.349	0.428	0.000
Professional/vocational academy	0.256	0.196	0.000
Technical college/university degree	0.288	0.192	0.000
Other/no training	0.022	0.034	0.082
Geographic region			
Northern Germany	0.196	0.167	0.059
Eastern Germany	0.333	0.374	0.033
Southern Germany	0.155	0.182	0.080
Western Germany	0.316	0.277	0.038

	Participants	Non-Participants	*p*-value
Local macroeconomic conditions			
Unemployment rate			
Less than 4%	0.095	0.101	0.619
4 to less than 6%	0.185	0.213	0.079
6 to less than 8%	0.241	0.189	0.002
8 to less than 10%	0.202	0.203	0.923
More than 8%	0.276	0.293	0.358
Vacancy to unemployment rate			
Less than 10%	0.244	0.276	0.077
10 to less than 15%	0.294	0.269	0.170
15 to less than 20%	0.165	0.140	0.089
20 to less than 30%	0.209	0.209	0.992
More than 15%	0.087	0.105	0.128
GDP per capita			
Less than 25,000 EUR	0.279	0.294	0.406
25,000 to less than 30,000 EUR	0.256	0.234	0.202
30,000 to less than 35,000 EUR	0.207	0.225	0.270
35,000 to less than 45,000 EUR	0.161	0.158	0.826
More than 45,000 EUR	0.097	0.089	0.491
Start-up rate out of unemployment			
Less than 4%	0.283	0.286	0.875
4 to less than 5%	0.247	0.267	0.262
5 to less than 6.5%	0.207	0.179	0.087
6.5 to less than 10.5%	0.154	0.160	0.661
More than 10.5%	0.110	0.108	0.886
Self-employment rate			
Less than 10%	0.224	0.199	0.117
10 to less than 11%	0.268	0.253	0.420
11 to less than 12%	0.266	0.232	0.050
12 to less than 13%	0.099	0.150	0.000
More than 13%	0.143	0.166	0.120
Personality traits			
Big Five			
Conscientiousness	4.383	4.106	0.000
Extraversion	3.692	3.297	0.000
Agreeableness	3.401	3.368	0.300
Neuroticism	2.365	2.677	0.000
Openness	3.723	3.255	0.000
Other personality traits			
Readiness to take risks	5.777	5.316	0.000
Locus of control	4.260	3.785	0.000
Patience	6.370	6.380	0.911
Impulsiveness	5.388	5.409	0.809
General self-efficacy	4.391	4.062	0.000
Number of observations	1,248	1,204	

Note: Reported are sample shares for dummy variables and means for multi-valued variables. *p*-values are based on *t*-tests of equal means.

Table A.2: Propensity Score Probit Estimation

	Baseline Specification	Extended Specification
Socio-demographics		
Age		
(Reference category: less than 25 years)		
25 to less than 35 years	-.458	-.437
35 to less than 45 years	-.413	-.318
45 to less than 56 years	-.301	-.164
56 years and older	-.367	-.225
Female	-.201***	-.361***
Not German citizen	0.659***	0.73***
Health restrictions	0.195	0.122
Married	0.014	-.034
Number of children		
(Reference category: no children)		
One child	-.229**	-.230**
Two children and above	-.226**	-.169
Children under 10 present	0.211**	0.135
Single parent	0.208	0.209
Human capital		
Highest schooling degree		
(Reference category: no schooling degree)		
Lower secondary school	0.194	0.439
Middle secondary school	0.307	0.559**
Upper secondary school (specialized)	0.415	0.627**
Upper secondary school (general)	1.188***	1.549***
Educational and vocational training		
(Reference category: other/no training)		
Lower/middle secondary school, no vocational training	0.366	0.581
Lower/middle secondary school, with vocational training	0.066	0.088
Higher secondary school, no vocational training	-.052	-.193
Higher secondary school, with vocational training	0.605***	0.753***
University degree	-.199**	-.226**
Professional qualification		
(Reference category: unskilled workers)		
Skilled workers	-.054	-.076
Skilled workers with technical college education	-.026	0.061
Top management	-.104	-.036
Labor market history		
Fraction of time in unemployment in the last 10 years		
(Reference category: less than 10%)		
10 to less than 20 %	-.202**	-.185**
20 to less than 40 %	-.326***	-.280***
40 to less than 60 %	-.568***	-.637***
more than 60 %	-1.032***	-.946***
Five years before (hypothetical) entry		
Months employed	-.014	-.014
Months in labor market program	-.017	-.0006
Four years before (hypothetical) entry		
Months employed	-.002	0.0007
Months in labor market program	0.038*	0.027
Three years before (hypothetical) entry		
Months employed	0.01	0.009
Months in labor market program	-.095***	-.098***
Two years before (hypothetical) entry		
Months employed	0.015	0.011
Months in labor market program	0.042	0.027

(Table continued on next page)

	Baseline Specification	Extended Specification
One year before (hypothetical) entry		
Months employed	0.022*	0.023*
Months in labor market program	0.18***	0.18***
Employment status before entering unemployment		
(Reference category: other)		
Dependent employment	0.332***	0.38***
Self-employment/family worker	0.974***	0.846**
School/apprenticeship	0.067	0.149
Disable to work/unemployable	-.574***	-.544***
Occupational group before entering unemployment		
(Reference category: other)		
Forestry, fishing, animal breeding	-.423	-1.005
Manufacturing	-.782	-1.630*
Technical profession	-.593	-1.270
Services	-.542	-1.304
Daily income from last employment (Euro)		
Root of income	-.197	-.350**
Root of last income x log of income	0.407***	0.433***
Duration of last unemployment spell		
(Reference category: less than 1 month)		
1 to less than 3 months	-.448***	-.409***
3 to less than 6 months	-.473***	-.385**
6 to less than 12 months	-.422***	-.358**
12 to less than 24 months	-.386*	-.272
24 months and above	0.291	0.177
Monthly unemployment benefit		
(Reference category.: less than 300 Euro)		
300 to less than 600 Euro	-.374***	-.299**
600 to less than 900 Euro	-.520***	-.533***
900 to less than 1200 Euro	-.292**	-.291**
1200 to less than 1500 Euro	-.191	-.139
1500 Euro and above	0.047	0.039
Intergenerational information		
Father and/or mother was born abroad	-.098	-.076
Father and/or mother is/was self-employed	0.259***	0.241***
Father employed when respondent 15 years old	0.099	0.066
Professional education of father		
(Reference category: other / no training)		
Vocational training	0.059	0.023
Professional/vocational academy	0.314***	0.256**
Technical college/university degree	0.333***	0.251**
Geographic region		
(Reference category: western Germany)		
Northern Germany	0.11	0.139
Eastern Germany	-.001	-.036
Southern Germany	0.091	0.103
Local macroeconomic conditions		
Unemployment rate		
(Reference category: less than 4%)		
4 to less than 6%	-.389***	-.439***
6 to less than 8%	0.046	-.083
8 to less than 10%	-.253	-.296
more than 8%	-.204	-.224
Vacancy to unemployment rate		
(Reference category: less than 10%)		
10 to less than 15%	0.039	0.128
15 to less than 20%	0.129	0.177
20 to less than 30%	0.018	0.087
more than 15%	-.166	-.185

	Baseline Specification	Extended Specification
Gross Domestic Product per capita		
(Reference category: less than 25,000 EUR)		
25,000 to less than 30,000 EUR	-.009	-.021
30,000 to less than 35,000 EUR	-.273*	-.254*
35,000 to less than 45,000 EUR	-.462***	-.460**
more than 45,000 EUR	-.491**	-.554**
Start-up rate out of unemployment		
(Reference category: less than 4%)		
4 to less than 5%	-.179*	-.197**
5 to less than 6.5%	-.130	-.208*
6.5 to less than 10.5%	-.181	-.232*
more than 10.5%	-.066	-.127
Self-employment rate		
(Reference category: less than 10%)		
10 to less than 11%	-.037	-.043
11 to less than 12%	-.151	-.165
12 to less than 13%	-.304**	-.319**
more than 13%	-.281**	-.277*
Interaction terms		
Self-employed before entering unemployment x		
Months employed two years before	0.048	0.079**
Months in labor market program 1 year before x		
Months in labor market program 2 years before	-.072***	-.075***
Months in labor market program 2 years before x		
Months in labor market program 3 years before	0.008	0.008
Last unemployment spell longer than 2 years x		
Months in labor market program 1 year before	0.604***	0.687***
Last unemployment spell longer than 2 years x		
Spent at least 60% in unemployment in last 10 years	-1.592	-1.958*
Start-up rate from unemployment is 5 to 6.5% x		
Months in labor market program 3 years before.	0.186***	0.206***
Daily income from last employment x		
Upper secondary school (general)	-.768***	-.901***
GDP per capita between 25,000 and 30,000 EUR x		
Unemployment rate between 6 and 8%	-.781***	-.821***
GDP per capita between 25,000 and 30,000 EUR x		
Higher secondary school and vocational training	-.863**	-.659
Personality traits		
Big Five		
Conscientiousness		0.169***
Extraversion		0.14***
Agreeableness		0.017
Neuroticism		-.057
Openness		0.217***
Other traits		
Readiness to take risks		-.014
Patience		-.099***
Locus of control		0.378***
Impulsiveness		-.086**
General self-efficacy		.107
Constant	1.566	-2.303**
Number of observations	2,452	2,452
Pseudo-R^2	0.204	0.322
log-Likelihood	-1352.541	-1152.516
p-value of Chow-test	0.999	0.992

*Note*: Reported are probit coefficients. ***/**/* denote significance at the 1/5/10 % level. The Chow test assesses whether selection patterns are different among men and women.

Table A.3: Auxillary Regression

|	Leave-one-month-out approval rate
Personality traits	
Big Five	
Conscientiousness	0.008
Extraversion	0.61
Agreeableness	-.765
Neuroticism	-.087
Openness	0.59
Other traits	
Readiness to take risks	-.125
Patience	-.081
Locus of control	0.978
Impulsiveness	0.295
General self-efficacy	0.132
Number of observations	2,452
Regression R^2	0.205
$H_0 : \beta_{personality} = 0$	
F-Statistic	1.49
p-value	0.137
Controls	
Personal characteristics	✓
Regional characteristics	✓

Note: This table reports OLS coefficients on personality traits from an auxiliary regression of the instrument on our measures of personality traits, controlling for personal and regional characteristics. Personality traits are re-scaled, i.e. they all have a standard deviation of one. Inference is based on robust standard errors. The bottom of the table shows the results of an *F*-test of the null hypothesis that personality traits have no conditional correlation with the instrument.

Figure A.1: Propensity Score Distribution – Link Functions
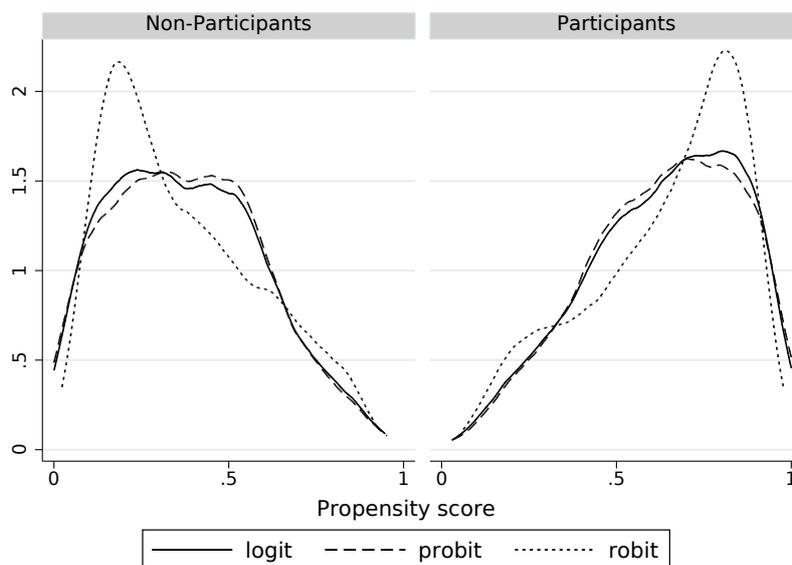
Note: This graph shows kernel density estimates of propensity scores estimated using different link functions. The robit regression makes use of the t-distribution with optimally chosen degrees of freedom. All results are based on the baseline propensity score specification show in Table A.2.

Figure A.2: Propensity Score Distribution – Extended Specification
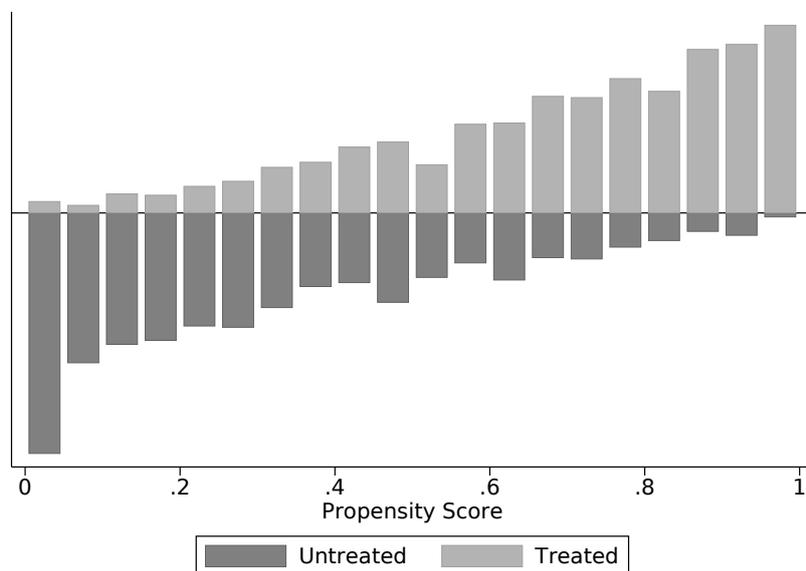
Note: This graph shows the distribution of estimated propensity scores for the treated and comparison group using a probit regression based on the extended specification including information on socio-demographics, human capital, labor market history, intergenerational transmission, regional controls for local labor market conditions and self-employment activity as well as personality traits. For details on the specification and estimated coefficients, see Table A.2.