Economic Systems 47 (2023) 101071

Contents lists available at ScienceDirect

Economic Systems

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

One (program) for all *or* all (programs) for one: Evaluation of the employment program opportunity for all of the federation of Bosnia and Herzegovina

Merima Balavac-Orlic[a,*], Josefina Posadas[b]

[a] *Department of Quantitative Economics, School of Economics and Business, University of Sarajevo, Sarajevo, Bosnia and Herzegovina*
[b] *World Bank, Washington DC, United States*

ARTICLE INFO

Jel classification:
J23
J24
J38
J64
J68

Keywords:
Low wage subsidy
Employment
Unemployment
Duration
ALMPs
Bosnia and Herzegovina

ABSTRACT

Using administrative data from tax records and public employment services, this paper examines whether the largest wage subsidy program deployed in 2014 in the Federation of Bosnia and Herzegovina, *Opportunity for All*, was effective at improving employment outcomes. Given the non-experimental design, the paper relies on propensity score matching estimators. It contributes to the literature on impact evaluations of active labor market policies (ALMPs) by exploiting detailed work histories of jobseekers to identify the control group. In the preferred specification, the program was effective in increasing employment among program participants relative to the control by 13 % 12 months after completing the subsidized period. However, the results are highly sensitive to the assumption of the starting date of the job spell in the control group, which carries information about previous work history. When changing the assumptions about the starting date of the subsidized job spell of the control group, the results remain positive in the short run, but turn out to be either larger or even negative for the medium run. In all cases, heterogeneity analysis reveals that the program is most beneficial for jobseekers of about 40 years of age and older, and for low-skilled workers.

1. Introduction

The debate among researchers and policymakers on the effectiveness of active labor market programs (ALMPs) in promoting employment and other labor market outcomes does not cease. While evidence from rigorous evaluations continues to mount, results remain mixed, with different experts drawing opposite conclusions. Some, seeing the glass as half empty, call for investing public resources in other types of reforms, for example those directed at promoting firm growth (McKenzie, 2017). Others, seeing the glass as half full, argue that ALMPs are a powerful instrument to address unemployment and underemployment (Card et al., 2018). These differences in opinion stem from several causes. The evidence is far from exhaustive, impact evaluations cannot account for the effect of macro factors (such as economic cycles, institutions, labor market regulations, or overall level of unemployment) or general equilibrium effects, and external validity is lacking in many boutique impact evaluations, leading to failures when scaled up or transferred to other settings.

* Correspondence to: School of Economics and Business, University of Sarajevo.
E-mail address: merima.balavac@efsa.unsa.ba (M. Balavac-Orlic).

https://doi.org/10.1016/j.ecosys.2022.101071
Received 8 October 2020; Received in revised form 23 August 2022; Accepted 30 August 2022
Available online 20 December 2022
0939-3625/

Moreover, ALMPs are costly policies, so even when there is a positive impact they might not be cost effective, at least from a fiscal perspective. Very few evaluations include cost-benefit analyses, and even then they only include projections of future fiscal revenues from earnings of improved labor market outcomes and the direct cost of implementing the policy, suggesting they are not worth delivering to older workers and other groups.[1] However, there are other benefits for which assigning a monetary value is more difficult, such as social cohesion, poverty reduction, or even a sense of self-satisfaction (World Bank, 2012). In addition, it is crucial to carry out a thorough heterogeneity analysis as part of the impact evaluation to adjust the cost-benefit calculations for different groups of beneficiaries.

In spite of the controversy regarding the policies' impact and cost-effectiveness, one policy recommendation is clear: ALMPs should *always* be followed by rigorous impact evaluations. However, even though running an impact evaluation is cost-effective most of the time given the high cost of these programs, evaluations, let alone evaluations using randomized control trials (RCTs), are still not standard practice. In Bosnia and Herzegovina, for example, none of the ALMPs have been evaluated since 2001,[2] even though entity governments spent around 50 million Bosnian convertible Marks (BAMs), or 0.167 % of GDP, in 2016 alone to support about 17,000 jobseekers.

This paper aims to contribute to the debate on the effectiveness of ALMPs and, more importantly, to effective policy design in the Federation of Bosnia and Herzegovina (FBH) by running a non-experimental impact evaluation of the 2014 *Opportunity for All* program (OfA). This program consisted of a wage or employment subsidy given to employers when they hired registered unemployed jobseekers.[3] The subsidy lasted six months and covered social contributions, which in the FBH are very high at 51 % of the labor cost.[4] All registered jobseekers were eligible,[5] but most participants were low-skilled and had been unemployed for at least six months. The program benefited 5299 jobseekers and 3672 employers.[6] In spite of its high cost—10,431,867 BAM[7] in 2014—it was not designed to include an experimental impact evaluation, allowing only matching techniques on observable characteristics to construct a control group. Since then, the government of FBH has been allocating larger budgets to support ALMPs, starting with 16 million BAM in 2018 for ALMPs, to 28 million BAM in 2021. Throughout this period, there have only been a few impact evaluations of any ALMPs.

This paper is structured as follows. The next section reviews the relevant empirical evidence on the impact of wage subsidies and a few ALMPs implemented and evaluated in the Balkans. Section 3 briefly describes the entity's labor market. Section 4 describes the data and Section 5 discusses the impact evaluation methodology. Section 6 discusses the results, followed by a robustness analysis in Section 7 and heterogeneity in impact in Section 8. The last section summarizes the main findings and provides ideas for next steps.

2. Literature review

ALMPs have been used in transition countries during the last 25 years or more to support the economic transformation, which has raised levels of unemployment, underemployment and informal employment. Moreover, since the global financial crisis, ALMPs have been expanded in OECD and transition economies as a means of coping with the employment effects of the crisis. They are also being considered globally to support the fast pace of change of labor markets due to the 4th industrial revolution and other megatrends. However, the appropriateness of these policies has become very controversial among researchers and policymakers due to the uncertainty around their impact on labor outcomes as well as their high cost.

ALMPs have been found to have mixed results on employment and earnings, the two outcomes usually measured. The various reviews of rigorous evaluations conducted in the last 20 years (Dar and Tzannatos, 1999; Martin and Grubb, 2000; Betcherman et al., 2004; OECD, 2015; Card et al., 2010, 2018; Crépon and van der Berg, 2016; McKenzie, 2017; Kluve et al., 2016) allow one to draw some conclusions. Overall, ALMPs are more effective in high-income European countries than in other parts of the world (OECD, 2015; Betcherman et al., 2004; Card et al., 2018). In general, youth employment programs are less likely to be effective, except in Latin American and Caribbean countries (Betcherman et al., 2004; Kluve et al., 2018). The impact of ALMPs varies with the type of program, with human capital programs more likely to be effective in the medium and long term and job-search support showing higher impact in the short run (Card et al., 2018). There is substantial heterogeneity in impact; in particular, human capital programs tend to be effective for women and the long-term unemployed (Card et al., 2018). In addition, programs tend to have larger average effects during recessions (Card et al., 2018). However, drawing general conclusions is challenging, as one is unlikely to find ALMPs that are similar in *every* design feature.

[1] Often the cost-benefit analysis only includes the additional cost of public employment services in implementing ALMPs, such as hiring training providers or providing subsidies, and fails to account for the operational costs of the services that are an integral part of the policies.

[2] Benus et al. (2001) evaluate the impact of the Emergency Demobilization and Reintegration Project in Bosnia & Herzegovina. Impact evaluations are rare across sectors, with one prominent exception being the evaluation of the FIRMA project that supports SMEs.

[3] The terms *employment* and *wage subsidy* will be used interchangeably in this paper.

[4] The government of the FBH has put forward a proposal, currently in parliament, to reduce the tax wedge by making the personal income tax more progressive (in spite of making allowances part of the tax base) and reducing social security contributions. With this policy, the tax wedge could go as low as 23 % and 35 % for those at the bottom of the wage distribution.

[5] Hence the name of the subsidy program ("for All"), as every single registered jobseeker had equal access to it.

[6] A total of 4477 unique employers' applications were received by FEI. However, some employers applied more than one time, with a maximum of 89 applications made by one firm. In our sample, we work with 4670 jobseekers and 1939 employers, as further described in Section 4.

[7] Equivalent to 6283,878 USD (in 2014).

Employment or wage subsidies are only one type of ALMP, and are the preferred instrument of many countries. They are liked by governments because they affect both labor demand and supply. Wage subsidies aim to encourage employers to hire new workers or maintain the jobs of existing employees by reducing labor costs (Almeida et al., 2014; Kluve, 2010). They might be interpreted as a payment to the employer for the cost of initial training, and thus be expected to increase the skills of the beneficiaries by program completion (Heckman et al., 2002; Brown and Koettl, 2015).[8] Moreover, wage subsidies can contribute to decreasing the information asymmetry between the employer and the jobseeker about their productivity (Almeida et al., 2014). They can also promote labor force participation by narrowing the wedge between the reservation wage and the productivity of the beneficiary. Finally, for many policymakers, their implementation is easier and more transparent. Usually wage subsidies are delivered by direct transfers to either employers (either as cash or as a reduction of payments for social security contributions) or workers.[9]

However, employment subsidies can also bring undesirable side effects such as substitution, displacement and deadweight losses. Subsidies may cause some workers to lose their jobs due to changes in relative wages (a substitution effect with subsidized workers replacing non-subsidized workers) or because subsidies may push some firms out of the market, destroying jobs (displacement effects). Moreover, many wage subsidy beneficiaries would have been hired regardless of the wage subsidy, leading to large deadweight losses or windfall effects.[10]

Like other types of ALMPs, wage subsidies are not always found to have a positive impact on labor market outcomes. The meta-analysis by Card et al. (2018) shows that the impact of subsidy programs on employment is positive and increases over time.[11] The same study finds that wage subsidy programs are equally effective as training programs. However, some of the skepticism about the effectiveness of subsidy programs lies in the fact that carefully designed evaluations with randomly selected control groups do not find any positive impacts. One such evaluation of a program for young women transitioning from school to work found large impacts eight months after program completion (36-ppt higher employment probability for the treatment group) but *no* effect one year and nine months after program completion (Groh et al., 2016).

3. Labor market and active labor market policies in FBH

Labor markets in the Federation of Bosnia and Herzegovina (FBH) have been weak. In 2014, the year the analyzed program was carried out, labor force participation was low, particularly among women, and the unemployment rate high (Fig. 1). Labor force participation was only 42 %, 30.5 % points below the EU average.[12] While current labor markets have improved,[13] in 2014 the unemployment rate was 28 % and almost 60 % among youths aged 15–24. This is one of the highest observed unemployment levels not only in the region but also in the world.

The high level of unemployment and the generous benefits delivered through employment services result in an extremely large number of persons registered as jobseekers, about twice as many as those actively seeking a job.[14] Fig. 2a shows that in 2014, there were 391,000 persons registered with the public employment service (PES).[15] Besides intermediation services aiming to help the unemployed get back into formal employment, the PES administers unemployment benefits and health insurance[16] (Fig. 2b).

[8] We did not have quantitative data on the skill level of individuals in our sample. However, we did have interviews with some employers that participated in the program. Most employers argue that they invest in firm-specific skills and emphasize that employment contributes to the development of soft skills of participants (primarily motivation and work ethics), as most employees had been experiencing long periods of unemployment or inactivity before program participation.

[9] See Almeida et al. (2014) for a theoretical discussion of wage subsidies.

[10] There are a large number of design parameters that need to be defined in a subsidy program. Some researchers give specific names to programs according to some of these features. For example, for some, "employment subsidies" refer to programs that are used for firms to maintain a certain level of employment (transfers going to firms to keep workers/avoid layoffs) while "hired subsidies" refer to transfers made to firms/workers for each newly unemployed person that is being hired. In spite of these subtle differences, most of the time people simply refer to them as wage subsidies, independently of the modality, since in all cases the instrument consists of a transfer that will be used to pay part of the gross wage of a worker.

[11] The size of the employment impact is 1.1 % points (ppts) (of the control) in the short term (less than 1 year after the completion of the program), 6.2 ppts in the medium term (1–2 years after program completion), and 21.1 ppts in the long term (2 or more years after program completion). The values just given correspond to the mean program effect on probability of employment ($\times$100).

[12] The average labor force participation rate for the EU (28 countries) was 72.5 % (OECD database) https://data.oecd.org/emp/labour-force-participation-rate.htm.

[13] In 2017, the unemployment rate was 20 %.

[14] In the FBH, there are two reported measures of unemployed persons: one from the Federal Employment Institute, which counts the number of registered unemployed, and one from the Labor Force Survey, which uses the ILO definition of unemployment.

[15] The PES refers to all levels of public employment services, both at the entity level (the Federal Employment Institute, FEI) and the cantonal level.

[16] Eligible beneficiaries of the unemployment benefit are those who were formally employed for the last 8 months. It is distributed for the following periods: For 3 months if the person was employed between 8 months and 5 years; for 6 months if they were employed between 5 and 10 years; for 9 months if they were employed between 10 and 15 years; for 12 months if they were employed between 15 and 25 years; for 15 months if they were employed between 25 and 30 years; for 18 months if they were employed between 30 and 35 years; and for 24 months if they were employed more than 35 years. If the person already received unemployment benefits, only employment after that is considered as the base for the next unemployment benefit. All registered unemployed can benefit from health insurance.

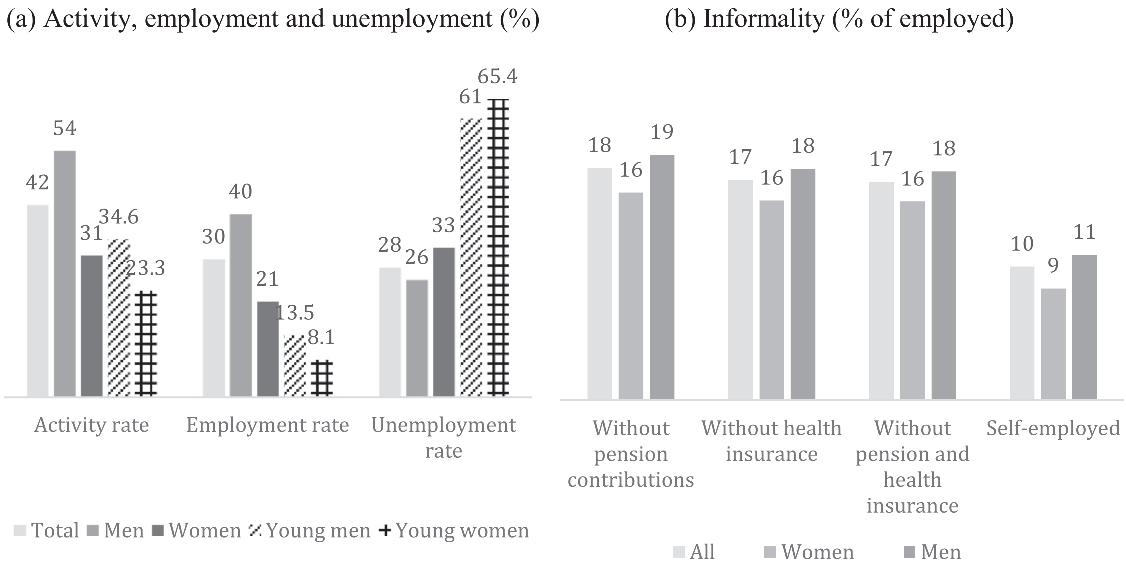


Fig. 1. Key labor market indicators, 2014. (a) Activity, employment and unemployment (%). (b) Informality (% of employed). Source: Labor Force Survey Report for 2014 (http://www.bhas.ba).

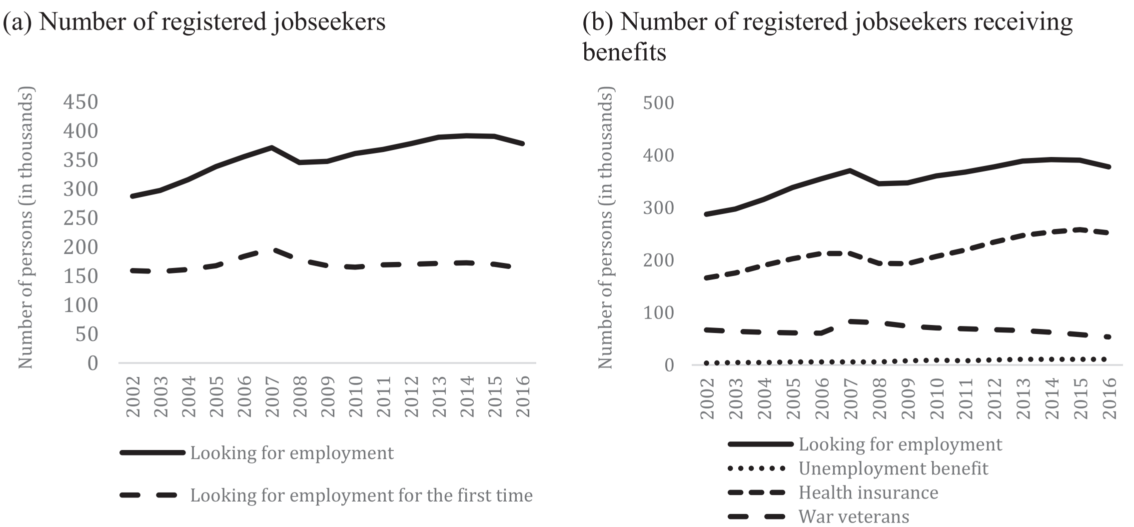


Fig. 2. Registered jobseekers and benefits received, 2002–2016. (a) Number of registered jobseekers. (b) Number of registered jobseekers receiving benefits. Federal Employment Institute.

3.1. Active labor market programs in the federation of Bosnia and Herzegovina

Labor market programs to support jobseekers are designed and financed both at the cantonal (sub-entity) and the federal level, although they are always implemented at the local cantonal level. In general, cantons use common procedures and forms, and the same portfolio of active measures, although the mix may be somewhat different from one canton to another. The Federal Employment Institute (FEI) has been managing the largest ALMPs in FBH since 2002, and OfA was the largest employment program (see Table A.1 in the Appendix).[17]

OfA is a wage subsidy program that aims to help unemployed individuals in the FBH to (re)integrate into the labor market and avoid long-term unemployment. The subsidy is directed at companies whose only eligibility condition is that they must be up to date with taxes and social contributions payments.[18]

[17] Active labor market policy in FB&H is implemented in accordance with the Employment Strategy in B&H and the Law on Professional Rehabilitation, Training and Employment of Persons with Disabilities of the Federation of Bosnia and Herzegovina (FBH), adopted in 2010.

[18] More specifically, employers are excluded if they have evaded taxes and social security contributions; or do not have an agreement with the Federation Tax Administration on debt settlement based on contributions; or are included in the Register of Fines due to a debt of unpaid fines

Firms can hire any unemployed person regardless of age, gender or qualification. The only eligibility condition for program participants is that they must have been registered with the employment offices for at least 90 days before the firm submits its application to the program.[19] Employers can apply to the program as long as there are still funds available. For the 2014 program, enrollment occurred between May 5, 2014, and April 5, 2015.

4. Data

This paper relies on administrative data from the Tax Administration (TAD) and the Federal Employment Institute (FEI). A person is considered unemployed if they are registered with the FEI. Note that this definition is less demanding than that of the ILO, which also requires job-search effort and readiness to start a job. A person is considered employed if an employer pays social security contributions registered in the TAD.

The administrative data of the FEI allows one to identify OfA beneficiaries and can be linked to the TAD. Like many other administrative databases, the TAD provides unemployment and employment histories in a way that resembles longitudinal data but with some distinctions. It shows a series of episodes or spells that end each time the one responsible for paying social security contributions to the TAD changes. Each spell includes the exact start and end date, and a code associated with the responsible contributor that allows identifying if the spell corresponds to employment or unemployment. The data is therefore a series of episodes or spells of employment and unemployment, with one spell ending when the next one begins. It covers registered unemployed and formal workers since its creation in 2009 until April 2016 when data was extracted for the last time. [20], [21].

A number of persons in the FEI database do not have records in the TAD and are hence excluded from the study. An unemployed person registered with FEI for job search support, for whom the health insurance is not paid by the FEI (for example because it is covered by their spouse) and who never held a formal job after 2009 will not be recorded in the TAD database.[22] Table 1 shows the number of observations that are kept in the process of merging both sources of administrative data.

The episode data from TAD is used to construct a series of variables that represent the work history prior to program eligibility and determine program eligibility over time. Work histories can be thought of not only as a proxy for skills/employability but also as a measure of effort in job search and program participation (Bryson et al., 2002). The following variables were constructed: length of last unemployment spell prior to program participation, number of unemployment spells between 2009 and program starting date, and percentage of time jobseekers were employed since 2009. In addition, individual demographic characteristics of the jobseekers were extracted from the TAD and FEI administrative data.[23]

5. Impact evaluation methodology

This paper employs propensity score matching (PSM) and leverages large and detailed administrative data to define a control group and estimate the impact of the OfA program. It has been widely discussed that estimates from this identification strategy could still be subject to biases arising from differences in unobservable variables (Ravallion, 2007; Gertler et al., 2016). Specifically to wage subsidy programs, Schunemann et al. (2015) estimate both intention to treat, relying on a regression discontinuity design (RDD), and difference-in-difference (DiD), and compare it to PSM for the BHI[24] program in Germany. The authors claim that PSM biases are large enough to lead to the wrong conclusion that the program was effective. The biases stem from the fact that it is difficult for the analyst to separate the effect of being employed from the effect of the subsidy relying only on observable characteristics. While it is accurate that PSM cannot take into account unobservables, in practice biases are not necessarily found to be this large. In fact, Schunemann et al.'s (2015) conclusions have been challenged in the literature on several grounds. First, the RDD estimates cannot reject the hypothesis of a positive impact of the BHI program due to the lack of precision. Second, their control group was benefiting from other ALMPs (Wolff and Stephan, 2013). Third, their results are also subject to other selection issues as the unemployed with 11 months of unemployment may prefer to wait to enter into employment to access the BHI program, which was very generous (Cseres-Gergely et al., 2015). Instead, recent evidence suggests that concerns about PSM biases might not be of large consequence in settings like this, in particular when the analyst can rely on rich data (Caliendo et al., 2017). More broadly, Card et al. (2018) show that the choice of evaluation method (RCTs versus other non-experimental) does not bias their meta-analysis results. Furthermore, subsequent

(*footnote continued*)
because they did not conclude a contract with employees (i.e. informal employment); or used credit or nonrefundable funds from the FEI and cantonal programs and did not comply with the program obligations (e.g., if they did not employ the envisaged number of employees and did not return regularly received funds).

[19] This period is too short and very often people start working informally while waiting these 3 months

[20] This research team plans to request an additional data extraction to track progress over a longer period.

[21] If the person was employed in 2009 when the TAD's new system was created, the records go back to the date the worker started the job held at the time the system was created. Thus, the work history is unbalanced, covering longer periods of time for those who had longer job tenures in 2009.

[22] In other words, for a registered jobseeker to be in the TAD it is necessary that (i) the FEI pays for their health insurance (because the TAD has a unique record for the health insurance contributions paid by the FEI to the Health Insurance Fund), or (ii) have had a previous formal job. Persons who have spells of inactivity or informal employment have gaps in their timeline in the TAD.

[23] Actually, at this stage, the individual background characteristics in the sample come from the FEI for the treatment and from the TAD for the control.

[24] The official German name is *Beschäftigungshilfen für Langzeitarbeitslose* (BHI).

Table 1
Sample size.
Source: Administrative data from FEI and TAD.

	Sample size
Number of registered job seekers in FEI in May 2014	388,000
— and, of those, who are also registered in TAD	72,850
Number of job seekers participating in OfA	5299
Number of OfA participants registered with TAD	4666
Number of Ofa participants after trimming[(a)]	4180
Number of Ofa participants after matching	3966

Notes: 66,180 are non-OFA participants who are in both databases. (a) Outliers in terms of length of unemployment were removed from the sample.

publications using ITT to estimate the impact of wage subsidy programs have found positive effects (Pasquini et al., 2019; Sjögren and Vikström, 2015; Desiere and Cockx, 2021).

Propensity score matching (PSM) aims to identify control cases such as the probability of being in the treatment or the control group being statistically equal, conditional on observables.[25] Note that because OfA is an employment subsidy program, participation becomes conditional on accepting a job offer from an eligible employer. Hence, having equal probability of being in the treatment and control groups also requires that the eligible nonparticipants have the same probability of finding a job as treated individuals had the subsidy not been in place. Shünermann et al. (2015) refer to this as "double selection," the unemployed needs to be selected for a job (selection into employment) and into the program (selection into treatment), these being joint events.

The PSM is run in five steps. First, we estimate the probability of participating in OfA, in this case using a Logit model, to explain the probability of being in the program using observable characteristics. Second, we check if there is common support between the treatment and control groups. This means that the estimated propensity scores take positive values for each person regardless of being in the treatment or control group (Heckman and Vytlacil, 2005). Third, we match every jobseeker in the treatment group with one in the control group using 4:1 nearest-neighbor matching with replacement and bias adjustment and exact matching on female. The large and rich dataset allows choosing this option, which is appealing because it does not assume any specific functional form of either outcome or treatment model and because it increases efficiency by including up to four matches (as in Abadie et al., 2004). The robustness analysis includes testing various matching algorithms to ensure that our results are not sensitive to the choice of matching algorithm.

Next, we check that, conditional on the propensity score, the treatment and (matched) control group have the same distribution of observed characteristics, to ensure that observable characteristics do not further explain the probability of being in the program, relative to the propensity score. This is done by regressing the probability of participating in the program on the propensity score and the observed characteristics (Rosenbaum and Rubin, 1985). Lastly, we compare the mean in outcomes between the treatment and the (matched) control group to measure impact. This stage is repeated for different subgroups to explore heterogeneous effects across the group of beneficiaries.

To estimate the probability of participating in OfA we want to take advantage of the multiple work history variables. Since the OfA program allowed enrollment from May 2014 to May 2015, two issues need to be considered in defining the variables used in the PSM. First, there is seasonality in the labor market affecting the probability of receiving a job offer over time. For example, between June and September 2014, there could have been more openings due to high demand in some sectors, such as construction and tourism, but also a season when many jobseekers migrated to take temporary jobs in Croatia or Germany. Second, recent work history information weighs into the probability of receiving a job offer. Hence, it becomes important to carefully define the length of the last unemployment spell and number of jobs, or equivalently selecting a fixed date from which to start computing the length of the last unemployment spell. For example, those who started an OfA job later in the year may have taken another job earlier on, which helped them find the OfA job. *Third,* a person in the control group in May 2015 could become part of the treatment group if they accepted an OfA job later in the year.

The treatment group comprises all OfA participants in the 2014 call for employers' applications, and the control group will be selected from jobseekers registered with the FEI and TAD in May 2014, when the call for applications was announced. Most employers applied during the first two months (Fig. 3a) and most contracts (80 %) were signed in 2014, with July 2014 being the busiest month. Thus, the day OfA participants formally initiated employment (referred to later as "starting date of subsidy spell") varied over time. Fig. 3b shows the distribution of the starting date of the subsidy spell by month.[26]

Addressing these dynamic issues is not straightforward. To our knowledge, the only papers that tackle them are Vikström (2017) and Cronin et al. (2020). Vikström (2017) examines dynamic treatment effects by taking into account the probability that a person in the control group moves to the treatment group later in time. We do not enter into an exploration of these dynamic considerations. Instead, similar to Cronin et al. (2020), we focus on assigning a comparable point of reference in time to calculate work history

[25] Mathematically, we can represent this as $\Pr(\text{OfA}=1|X,t=1)=\Pr(\text{OfA}=1|X,t=0)$.

[26] Moreover, there is variation in the month when employers claimed the subsidy. Some employers were claiming the reimbursement of the subsidy as soon as they paid the first wage, while others did so later in the year, even after the whole subsidy period was completed. Figure A.1 in the Appendix shows the distribution over time of the month when employers presented the first invoice to claim the subsidy.

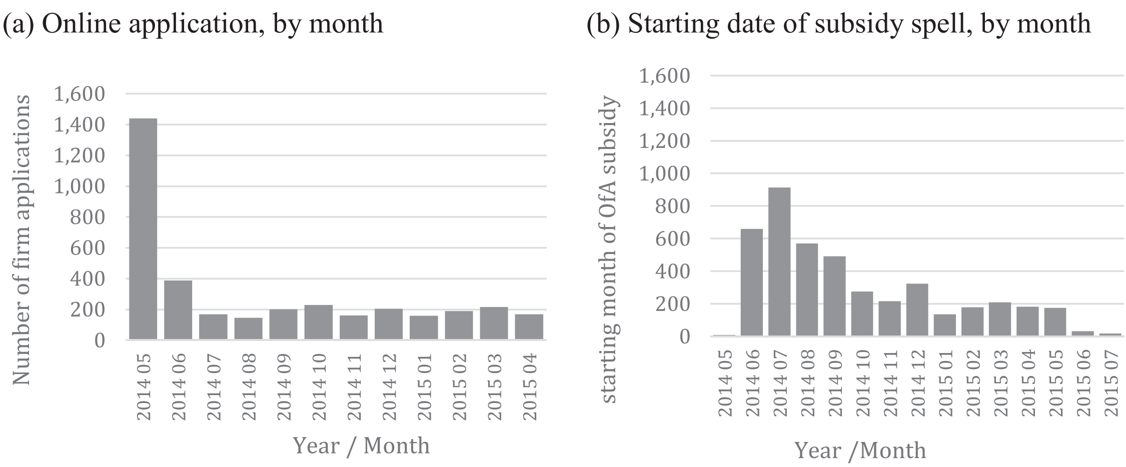


Fig. 3. Distribution over time of applications to OfA and signing of contracts. (a) Online application, by month. (b) Starting date of subsidy spell, by month. Source: FEI administrative data.

variables—or the starting date—to jobseekers who could be selected to form the control group. If we were to assign May 2014 as the starting date for all persons in the control group, we would be "allowing" them less time to find a job as compared to those who initiated the OfA program close to the end of the program. If we were to assign May 2015 as the starting date for all persons, we would be giving them a longer time to find a job or a better job match compared to those who joined the OfA close to the beginning of the program. More importantly, either approach would be equally arbitrary and creates room for different biases.[27] Hence, we assume that the counterfactual starting date to compute work history variables is November 5, 2014, the midpoint date of the program. In a six-month window around this date, there could be eligible non-participants of OfA who have or have not found a job. For those who were unemployed all along, we stick to using November 5 as the starting date. For those who found a job in that window, we use the first job spell of that window as the starting date.

Jobseekers who benefited from OfA have different observable characteristics from those who did not. Table 2 shows the means for the covariates for jobseekers in both groups, which are statistically different. While there is a higher proportion of women in the control group, the magnitude of the difference is very small, with 40.2 % of program participants and 42.8 % of nonparticipants being women. Program participants are on average 35 years old, mostly low skilled (88 % of them without higher education and 39.5 % with less than secondary school), have been unemployed for almost three years, and have on average 1.5 unemployment spells. Nonparticipants are more educated, with a higher proportion of secondary and tertiary graduates, but have been out of a job for two and a half months longer and are slightly more likely to have had an unemployment spell, on average. The geographic distribution of jobseekers differs as well. Table 2 also shows that OfA participants exhibit better labor market outcomes than OfA non-participants, before matching them.

While the differences in the observable characteristics are small and do not point toward either a positive or a negative selection into the program, cream skimming in the assignment of subsidies cannot be ruled out. First, it is well known that cream skimming also depends on unobservable characteristics. Second, the selection of jobseekers is done by the employers both for treatment and control, and they are obviously interested in getting the best candidates. However, there is no information to track how much public employment service intermediation was used to match jobseekers in the treatment and control groups with firms. It could well be the case that jobseekers in the treatment group were more likely to receive counseling and mediation, having more time to seek a job than those in the control group. Wage subsidies benefit both employers and employees.[28] Beneficiary firms are small but not start-ups. These firms have been in business for around 15 years, more than 60 % of them belonged in part or completely to the Government of FBH, most are concentrated in the service sector (70 %) and to a lesser extent in manufacturing (27.8 %), and most of them have 10 or fewer employees.

6. Results

We follow the steps described above to apply exact matching using a limited number of explanatory variables so as not to overburden the balance conditions. Once the control is identified, we explain the outcome variables using a larger set of explanatory variables.

[27] See Fig. 3 and A.1 in the Appendix for a graphic representation of the problem.

[28] As explained in Section 2, how much employers and employees benefit depends on the elasticity of the demand of labor.

Table 2
Characteristics and labor market outcomes of jobseekers and firms, for OfA participants and nonparticipants.
Sources: FEI and TAD.

	OfA Participants	OfA Non-participants	Difference	
Jobseekers				
Female (percentage)	40.2	42.83	-2.63	***
Age (years)	35.34	36.01	-0.67	***
Education (percent distribution)				
Nonqualified	9.24	5.95	3.29	***
Primary school	30.24	18.87	11.38	***
Secondary school	48.55	54.68	-6.13	
Higher diploma	1.15	1.87	-0.72	***
Higher education	10.82	18.63	-7.81	***
Length of last unemployment spell (months)	34.798	37.17	-2.53	***
Number of unemployment spells since 2009	1.45	1.55	-0.1	***
Cantons (percent distribution)				
Una-Sana Canton	14.15	9.24	4.91	***
Posavina Canton	4.07	1.22	2.85	***
Tuzla Canton	29.98	24.65	5.33	***
Zenica-Doboj Canton	14.35	20.71	-6.36	***
Canton Goražde	1.71	0.67	1.04	***
Central Bosnia Canton	14.69	6.28	8.41	***
Herzegovina-Neretva Canton	4.95	8.73	-3.78	***
West Herzegovina Canton	3.17	2.39	0.78	***
Sarajevo Canton	10.13	24.08	-13.9	***
Canton 10	2.81	2.04	0.77	***
Number of job seekers (N)	4666	68,186		
Beneficiary firms				
Age (years)	14.95			
Private firm (percentage)	38.58			
Economic sector (percent distribution)				
Agriculture	2.20			
Manufactures	27.78			
Services	70.02			
Size (percent distribution)				
Micro (1–10 employees)	69.87			
Small (10–50 employees)	21.74			
Medium (50–250 employees)	7.13			
Large (more than 250 employees)	1.27			
Mean number of employers	218.6			
Number of beneficiary firms	2477			
Jobseekers' employment outcomes 6 months after program completion				
Employed (%)	63.31	48.15	15.16	***
No. days worked	296.14	177.94	118.2	***
Employed w/same firm (%)	50.83	17.02	33.81	***
Number of jobs	1.12	0.9	0.22	***
Average length of unemployment[(b)]	52.99	78.13	-25.14	***
Number of observations	4666	68,186		

Notes: The sample includes participants and nonparticipants for workers but only means for participant firms as we could not get hold of all firm administrative data at this stage.

6.1. Propensity score matching and the identified control group

First, the propensity score, which measures the jobseeker's probability of participating in OfA (and jointly receiving a job offer), is estimated using a logit model and variables that capture skills, the local labor market context, and effort in the job search. To capture skills/employability we use age and age squared (as a proxy of experience), level of education, length of last unemployment spell (measured in months), and number of unemployment spells since 2009. We also control for the jobseeker's sex, canton of residence, and education level, interacted with canton dummies. Table A.2 in the Appendix shows the regression results of the probability model, and Table A.3 shows the mean values of the characteristics for the treatment and control groups constructed using the PSM method described above. The matching improves the quality of the control group, indicating that selection into the program is not random. Table 3.

The balancing test indicates that each outcome variable is balanced after matching, with the mean standardized bias being only 0.3 %, which is considerably smaller than the 3 or 5 % threshold (Smith and Todd, 2005). Similarly, the results of the likelihood ratio (Chi2) show that the joint test of significance between the characteristics of the treatment and the control group after matching cannot be rejected. Thus, there is enough confidence to regard the matching as performing well, and to consider the sample well balanced between treated and nontreated individuals for all observed characteristics (which are applied in all the tables in this section).

Table 3
Descriptive statistics for treatment and control after matching.
Sources: FEI and TAD, using administrative data.

	Treatment	Control	Difference	P-value	Percent bias reduction
Jobseekers					
Female (percentage)	40.89	40.89	0	1.000	100
Age (years)	34.49	34.42	0.07	87.7	0.737
Age squared	1276.2	1270.1	6.1	82.7	0.703
Education level	2.04	2.04	0	100	1
Length of last unemployment spell (months)	29.1	28.8	0.3	86.1	0.644
Number of unemployment spells since 2009	1.271	1.27	0.001	99.5	0.933
Cantons					
Posavina	4.11	4.11	0.00	1.000	100
Tuzla	30.00	30.00	0.00	1.000	100
Zenica-Doboj	15.1	15.1	0.00	1.000	100
Goražde	1.94	1.19	0.00	1.000	100
Central Bosnia	15.63	15.63	0.00	1.000	100
Herzegovina-Neretva	4.94	4.94	0.00	1.000	100
West Herzegovina	3.4	3.4	0.00	1.000	100
Sarajevo	10.29	10.29	0.00	1.000	100
Canton 10	4.11	4.11	0.00	1.000	100
Number of jobseekers (N)	3966	3966			

Notes: The control group corresponds to the *static with varying starting dates* model. Una-Sana Canton is an omitted canton. The education level is measured using an integer variable, which takes three values: 1 for non-qualified workers; 2 for primary school and secondary school; and 3 for TVET and higher education.

Table 4
Program impact on employment probability after a fixed number of months of subsidy termination.

	Employed at program completion (dummy)	Employed 3 months after the program (dummy)	Employed 6 months after the program (dummy)	Employed 9 months after the program (dummy)	Employed 12 months after the program (dummy)
	(1)	(2)	(3)	(4)	(5)
Treatment	0.275 * **	0.119 * **	0.083 * **	0.018 * **	0.131 * **
(st.error)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Mean of treatment	0.926	0.724	0.673	0.567	0.683
Mean of control	0.657	0.613	0.588	0.563	0.567
Treatment as a % of mean of control	41.78	19.47	14.07	3.12	23.06
Number of observations[(a)]	55,695	55,695	55,695	55,695	55,695
Matched sample					
In control	3966	3966	3966	3966	3966
In treatment	3966	3966	3966	3966	3966

Notes: The assumptions correspond to the specification assuming a starting date using a 3-month window around November 5th, 2014. The control variables are sex, age, age squared, education level, canton dummies, length of last unemployment spell, education level interacted with canton dummies, and number of unemployment spells since 2009. Abadie–Imbens (2006) standard errors are in parentheses; The exact matching on discrete variables (female) (teffect command in STATA with option *ematch* on female only) has been used. (a) Counts all the observations in the common support

6.2. Program results

The results show that OfA has a positive effect on the probability of being employed. Table 4 shows the average effect on probability of employment after program completion. The coefficients represent the average percentage point difference in the probability of being employed between the treatment (the program participants) and the control (the nonparticipants), when the control has similar observable characteristics to the treated. Six months after the completion of the subsidy, OfA participants were 8 % points more likely to be employed than jobseekers in the control group.

The positive treatment effect becomes smaller at longer durations, but remains statistically significant. This result is aligned with what is observed for some programs evaluated in the literature (Hujer et al., 2004; Venetoklis, 2004; Card and Hyslop, 2005; Kvasnicka, 2009; David and Houseman, 2010; Groh et al., 2016; De Mel et al., 2016, Galasso et al., 2004; Betcherman et al., 2010; Burger et al., 2022). There are many possible explanations for decreasing results over time. The comparison of our results with the means estimated by Card et al. (2018) is not straightforward. On the one hand, our medium-term results seem larger compared to those obtained for wage subsidies in this relevant meta-analysis, which are around 12 %. On the other hand, most studies of the meta-analysis have a longer term horizon of 24–36 months after program completion, and in many instances the wage subsidy programs were combined with job intermediation and/or skills training programs.

Table 5
Program impact on work intensity, job turnover, job-search duration and tenure in the subsidized job.

	Number of days worked 12 months after the program (1)	No. of jobs 12 months after completion (2)	Average unemployment length after 12 months (3)	Employed with the same firm 12 months after program completion (dummy) (4)
Treatment	103.64 * **	0.08 * **	-7.26 * **	0.322 * *
(st.error)	(3.18)	(0.01)	(2.44)	(0.009)
Mean of treatment	406.8	1.277	58.29	0.771
Mean of control	308.5	1.276	71.49	0.599
Treatment effect as a % of mean of control	33.6	6.57	-10.15	53.75
Number of observations[(a)]	55,695	55,695	55,695	55,695
Matched sample				
In control	3966	3966	3966	3966
In treatment	3966	3966	3966	3966

Notes: The assumptions correspond to the specification assuming a starting date using a 3-month window around November 5th, 2014. The control variables are sex, age, age squared, education level, canton dummies, length of last unemployment spell, education level interacted with canton dummies, and number of unemployment spells since 2009. Abadie–Imbens (2006) standard errors are in parentheses. The exact matching on discrete variables (female) (teffect command in STATA with option *ematch* on female only) has been used.

On the employers' side, the subsidy lowers the unit labor cost of the new hires for the duration of the program. As a result, during this period the firm has incentives to open additional temporary vacancies up to the point where the value of the marginal product of labor equals the lower marginal labor cost. Once the subsidy ends—and the marginal cost of labor goes back to its pre-subsidy level—employment also returns to its pre-subsidy equilibrium level, and some firms will destroy the vacancy. However, if the cost of hiring/firing/rehiring and the cost of training jobseekers are positive and larger than the wage subsidy, firms may prefer to keep the worker. The fact that the impact of the program decreases over time tells us that it is likely that the employers taking advantage of OfA have large costs due to hiring/firing/rehiring. On the workers' side, wage subsidies might help human capital accumulation while also contributing to weakening future employability. For example, the fact that the probability of being employed increases between 9 and 12 months after program completion might suggest that OfA participants gained skills that facilitated their re-employment.[29]

To better understand whether the above mentioned explanations could be at play, we next look at work-history trajectories over the two years of data. It is argued by some researchers (Caliendo and Schmidl, 2016; Caliendo et al., 2017) that wage subsidies could contribute to enhancing employability by raising the skills of the worker. If that is the case, we should expect program participants to be more likely to find another job (that is, in a shorter time), which would then lead to them having more days of work and fewer/ shorter unemployment spells following the program. Table 5 shows the impact of the program on the number of days of work. In line with the previous results, the number of days employed is larger for the treatment than for the control group. On average, twelve months after the end of the subsidy, OfA participants have worked 103 more days than the control group, that is, 33 % more. Another way of exploring whether program beneficiaries have gained skills is to look at the number of jobs held and the average duration of unemployment spells. OfA participants had more jobs on average than people in the control group. Twelve months after OfA completion, the treatment is 8 % more likely to have an additional job than the control (Table 5).

As a consistency check, we also measure the average length of unemployment between spells, which should always be shorter for those who participated in OfA given the results above. The time it takes a jobseeker to find another job if they participated in OfA is 7 days less compared to the control twelve months after the program (Table 5). This suggests that those who participated in OfA improved their skills and their job search efforts and/or adjusted their reservation wage and accepted a job offer more quickly than those in the control group.

Finally, the wage subsidy is expected to benefit the firm as well. The firm is expected to invest in building the specific human capital of the worker. In these cases, if the cost of investing in specific human capital and hiring is positive, a firm will be incentivized to maintain the worker. However, if the subsidy is larger than those costs, firms might prefer to dismiss the worker and rehire. It seems reasonable to think that firms whose cost for training human capital, whether specific or general, is low will tend to need or use less skills. Thus, it is likely that a worker who benefited from a job that ended with the subsidy did not gain many skills. Conditional on being employed, OfA participants were 32 % points more likely to stay with the same employer 12 months after program completion. To our knowledge, no other study has explored this outcome before, and this finding is quite large.

[29] Other studies have found the opposite effect. For example, Oskamp et al. (2008) argue that low-wage subsidies may have the expected direct effect by enhancing the employment of low-skilled workers, but instead create disincentives for workers to accrue human capital by reducing wage premiums between high- and low-skilled workers. This is explained, on the one hand, because wage subsidies artificially lead to wages that are above the productivity of low-skilled workers, and on the other hand because wage subsidies reduce the wage of high-skilled workers when subsidies are financed through labor taxes and contributions paid by this group. Another explanation is that wage subsidies are not necessarily used by firms that invest in the skills of program beneficiaries.

Table 6
Program impact after 12 months of subsidy termination, assuming the starting date as the beginning of the OfA.

	Employed 12 months after the program (dummy) (1)	Number of jobs 12 months after (2)	Average unemployment length after 12 months (4)	Employed with the same firm 12 months after program completion (%) (5)
Treatment	0.142 ***	-0.115 ***	-28.673 ***	0.147 ***
(st.error)	0.009	(0.009)	(2.617)	(0.008)
Mean of treatment	0.729	0.44	48.23	0.302
Mean of control	0.594	0.562	83.53	0.146
Treatment as a % of control	23.89	-20.47	-34.33	100.06
No. of observations	55,695	55,695	56.19	55,695

Note: The control variables are sex, age, age squared, education dummies, canton dummies, length of last unemployment spell, and number of unemployment spells since 2009.

7. Robustness checks

Three types of robustness checks were conducted. First, we explore what happens when we change the definition of the outcome variables by changing the assumption of the starting date for jobseekers in the control group. We repeat the estimates, assuming that all jobseekers in the control group would have had access to the subsidized job on May 5, 2014. Second, we check whether the results are dependent on the matching method. Third, we assume that jobseekers in the treatment and control groups are optimizing over a fixed period of time, starting in May 2014 when the call for applications was opened and ending in April 2016 when data was extracted from TAD. We will refer to this assumption as the *full optimization* model.

7.1. Changing the assumption regarding the starting date for the control

One concern is that the results are sensitive to the assumption of the assigned starting date of what would have been the subsidized employment for the control group. In the previous section, we assumed that the starting date of the counterfactual spell to the subsidized employment for the control was the midpoint of the program duration. However, this is not realistic, since the control does not have an equivalent job and jobs start every day. Two alternative extreme assumptions are to consider the beginning (May 5, 2014) or the end of the program period (May 5, 2015). Each assumption could create a bias that would go in both directions. On the one hand, if we assume that the starting date for the program for the control is May 5, 2014, then we would be stricter with those in the control by giving them less time to optimize relative to the treatment. If instead we assume the starting date to be May 5, 2015, we are giving people in the control one year to either be choosy about the match or to change some characteristics so that they are not exactly matched with the treatment.

Knowing that none of these assumptions are perfect, Table 6 presents the results assuming May 5, 2014, as the starting date.[30] The results after 12 months are similar under this assumption. The dynamics concerning the number of jobs and the average unemployment spell are maintained, as is tenure with the same employer.

7.2. Treatment effects estimators

To check the sensitivity of the results to the matching algorithm, we tested a battery of methods including: (i) propensity score matching (PSM); (ii) one-to-one nearest neighbor matching (NNM) with replacement and caliper 0.001; (iii) kernel matching; (iv) regression adjustment; (v) inverse probability weighting; and (iv) inverse-probability-weighted regression adjustment. The results are shown in Table A.4 in the Appendix for selected outcome variables. Overall, the average treatment on treated effects is invariant to the choice of the matching algorithms. If correctly specified, different estimators should produce similar results. In particular, similar results should be found in a large sample, following Smith (2000).

7.3. Full optimization model

This approach is appealing as neither the analyst nor the public employment services determine who is going to be in the treatment or the control group among all eligible jobseekers. Some jobseekers may know as early as May 2014 that they could have access to a subsidized job and start working informally with the employer, while others might reject job offers while waiting for a subsidized job, and still others might even reject a subsidized job to find better opportunities somewhere else. Thus, by comparing outcomes by April 2016, we are giving all persons the same amount of time to obtain the best job possible, one that would put them on a trajectory to sustainable employment.[31] Applying this comparison will tell us if participating in OfA was positive for those who wanted it and could benefit from it.

[30] We also estimated the results assuming May 5, 2015, as the starting date; these are available upon request.

[31] A trajectory to *sustainable* employment is not the same as *continuous* employment with the same employer. For example, working in a big firm and learning the craft (whether as a mechanic or a lawyer) could lead to future jobs in other companies or to self-employment.

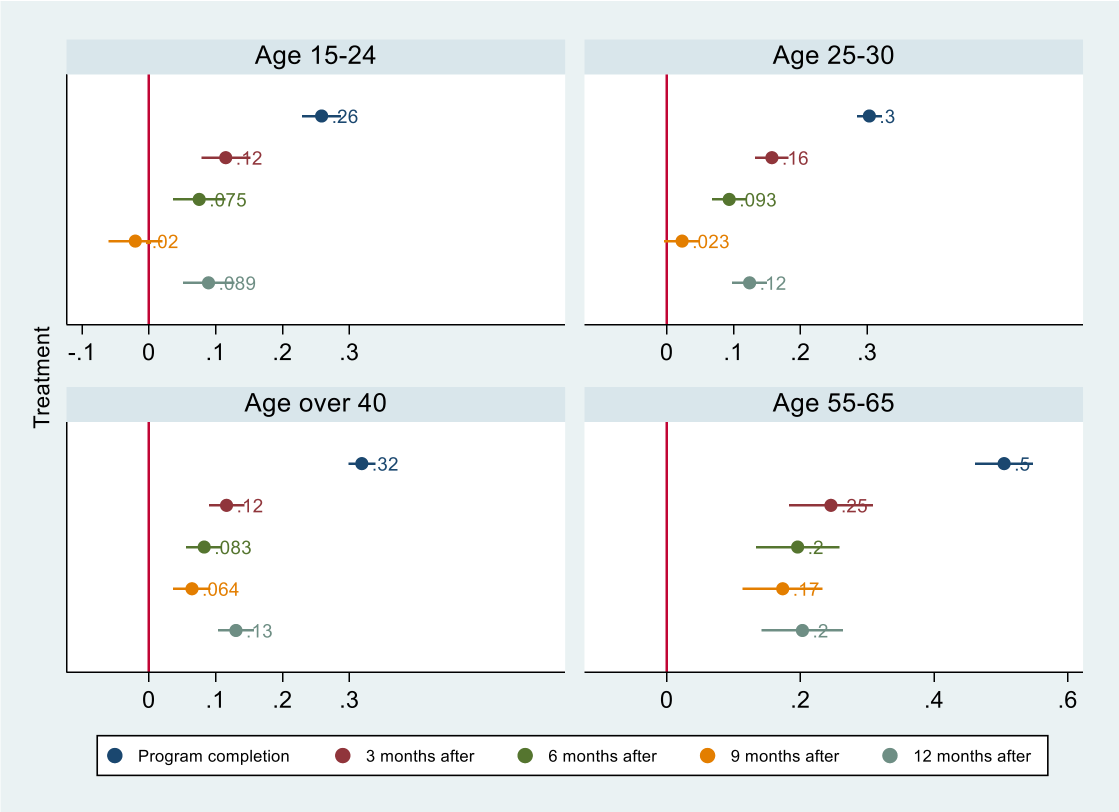


Fig. 4. Impact by age. *Sources*: FEI and TAD. *Note*: Static model with varying starting dates.

The results for selected outcome variables are shown in Table A.5 in the Appendix. By April 13, 2016, almost two years after the call for applications for employers was made, OfA participants were 12 % more likely to be employed than the jobseekers in the control group. OfA participants also worked 121 days more than jobseekers in the control group and were 27 % more likely to stay with the same employer and thus have fewer job spells and shorter unemployment spells. These results are quite positive and relatively large. However, they could also be subject to various biases. For example, employers might be opening vacancies for nonsubsidized jobs only after they have filled the subsidized jobs. In that case, the OfA would be causing a negative externality in the control group, and the counterfactual spell could end up happening later in the year. In the same vein, OfA participants can also be more skilled in un-observables and that is why they are selected by employers who benefited from the subsidy.

8. Heterogeneous effects

In this section, we look at the impact on different population groups. Since the program under investigation is not narrowly targeted—as the name “Opportunity for All” reminds us—there could be sizable heterogeneity in the subpopulation. Hence, we estimate treatment effects for the following groups: (i) youth, looking separately at those between ages 15 and 24 and those between ages 25 and 30; (ii) those close to retirement (ages 55–65); (iii) late-mid career (ages 40 and older); (iv) women; (v) long-term unemployed, defined as having been out of a job for one year or longer; (vi) low-skilled and high-skilled, defined by level of education; and (vii) new entrants to the labor market, defined as those who never held a job; and some combinations of the above. We also explore whether there are regional variations.

Fig. 4 shows the impact of OfA for different age groups. The main message is that wage subsidies are more effective for older workers, especially those close to retirement. For this last group, the impact is always positive. As with other groups, the size of the impact decreases over time. However, this may be because the impact is largest at program completion; even if it loses power, it remains positive 12 months later. One could imagine various reasons driving the larger impact for older workers. For example, older workers may not be that interested in developing a career and/or have more pressures to maintain their families, hence making them more likely to take the first job offered and not to invest in switching jobs if the job conditions are not the most desirable. At the same time, it could be more difficult for older workers to find a job, and thus, those in the control group do not perform as well, while those in the treatment group—knowing that—make more of an effort to keep the job. Another concern is that older workers may have more variation in unobservable characteristics that cannot be perfectly picked with the work history information, which in any case is truncated.

The estimated treatment has a positive influence on females’ employment outcomes; with the exception of outcomes measured nine months after the program, program participation has statistically significant effects among women (Fig. 5). For

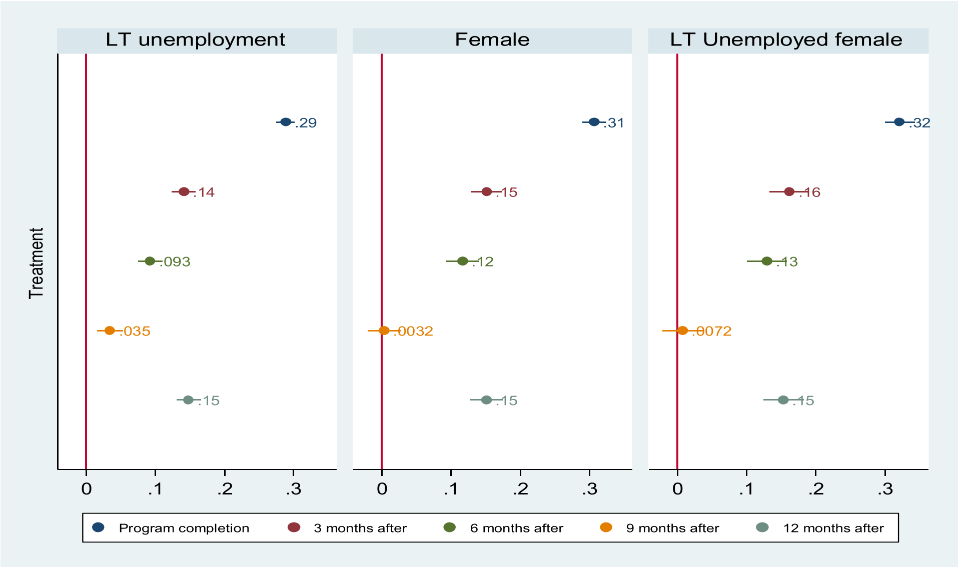


Fig. 5. Impact for women and long-term unemployed. *Sources*: FEI and TAD. *Note*: Static model with varying starting dates, using the 3-month window around November 5th 2014.

women, the program loses impact over time, and switches from positive to not different from zero by the ninth month after OfA completion.

We also examine whether the program is more beneficial for the long-term unemployed. The literature argues that the long-term unemployed (one year or longer) are less productive since they may have lost some skills while unemployed or have a very limited initial set of skills (Bell et al., 1999). There is an argument that wage subsidies to the long-term unemployed and substitution of the long- to short-term unemployed may result in beneficial general equilibrium effects on the labor market, with imperfect competition (Richardson, 1997). It has been argued that the long-term unemployed are labor market outsiders who put little downward pressure on nominal wages. The transformation of these “outsiders” into labor market participants means that the equilibrium rate of unemployment is reduced due to the increase in effective labor supply. Therefore, in the long run, the natural rate of unemployment is reduced (Bell et al., 1999). Our results for the long-term unemployed are very similar to the results from the entire population.

While the general pattern of the wage subsidy loses power over time regardless of the level of education of the beneficiary, low-skilled workers with primary education or less benefit more than the rest, being more likely to be employed than other groups. In particular, the low-skilled who have been unemployed for more than a year will benefit the most. Fig. 6 shows the impact by level of education for the whole population (Panel A) and for the long-term unemployed (Panel B).

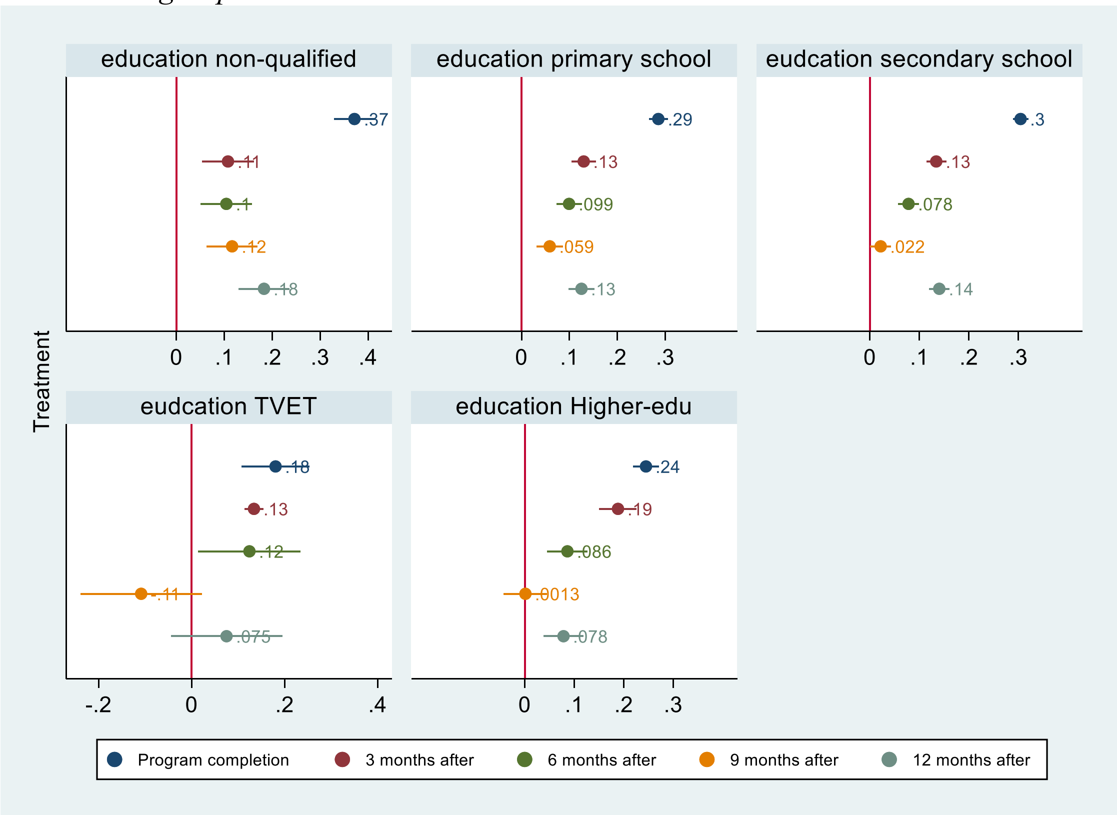
Next, we examine first-time entrants to the labor market separately. These are jobseekers who never held a job. While the impact is slightly larger, it still fades away with time as is observed for the total population (Fig. 7) in the short run; the impact of the program disappears a year after the program concluded.

Finally, we explore whether heterogeneity may arise by cantons. Table A.6 shows that wage subsidies lose impact over time in many cantons, but not in all of them. After the program is ended, the impact on the likelihood of being employed deteriorates most rapidly in Bosansko-Podrinjski, Srednjobosanski and Zapadno-Hercegovacki. In these cantons, the impact is positive, but only at program completion, and negative or insignificant in every follow-up period. Instead, for Unsko-Sanski, Posavski, Hercegovacko-Neretvanski, Sarajevo and K10, treatment is still positive three months after the program; it loses its effectiveness afterwards but never becomes negative. Finally, treatment is the most effective in the Tuzla and Zenicko Dobojski cantons. However, the overall trend is similar, with the treatment losing its effect over time but being surprisingly significant and positive nine months after the program ended. The results might be explained by the sectoral composition of employment in these cantons. The first group has a relatively large proportion of agricultural employment (more than 5 % in Central Bosnia). In the second group, employment is predominantly in services. And in the last group manufacturing is the main employment sector.

8.1. Heterogeneous effects by industry

Part of the heterogeneity could also be related to firms and industries. Firms in different economic sectors may have different needs and training costs, as well as different rates of churning. For example, in the construction sector, inherently, there is more churning because the activity varies with the season and the economic cycle. Certain manufacturing firms invest in the skills of new hires and understand that the skills need to be learned mostly on the job. Since the OfA program is open to all firms and all jobseekers, participants can work in any sector, in the same way as the jobseekers from the control group who find a job. Given that we cannot know in which economic sector a placement for a treatment or control jobseeker would occur, we adjust the sample to compare

Panel A. All groups



Panel B. By education for the long-term unemployed

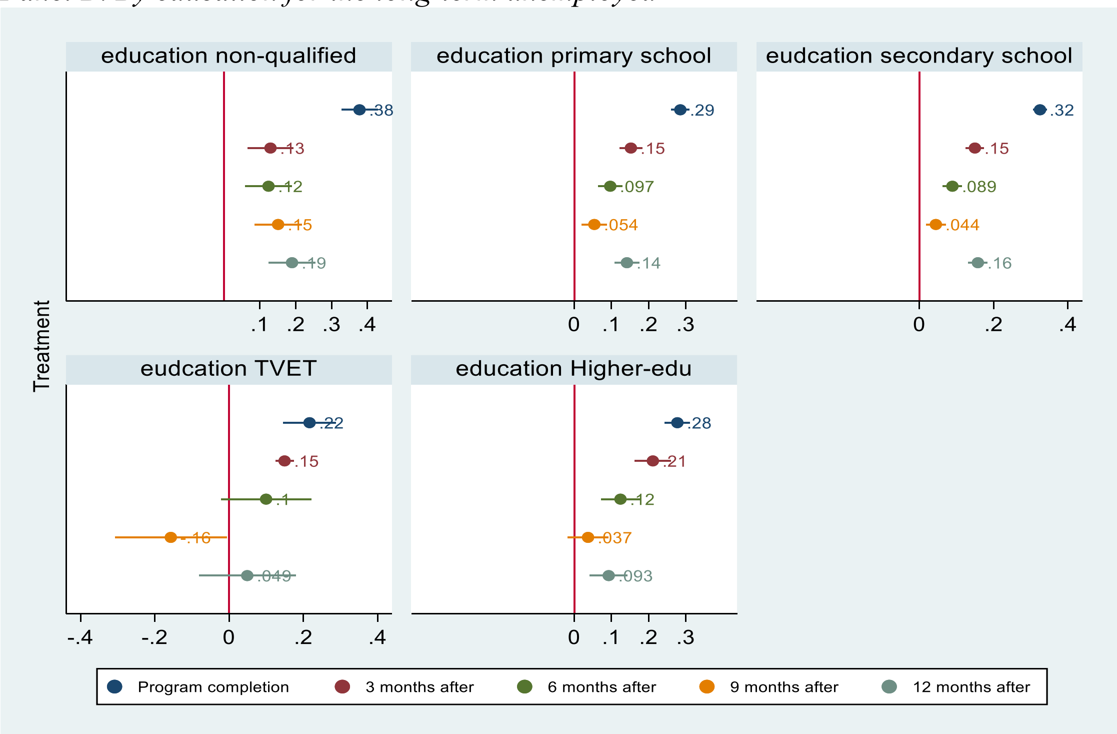


Fig. 6. Impact by level of education. *Panel A. All groups. Panel B. By education for the long-term unemployed. Sources*: FEI and TAD. *Note*: Static model with varying starting dates, using the 3-month window around November 5th 2014.

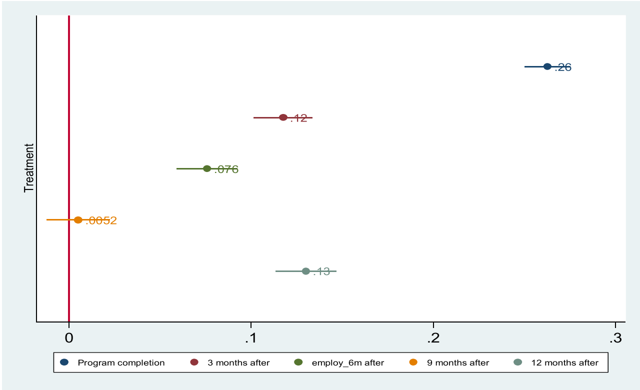


Fig. 7. First-time entrants to the labor market. *Sources*: FEI and TAD. *Note*: Static model with varying starting dates, using the 3-month window around November 5th 2014.

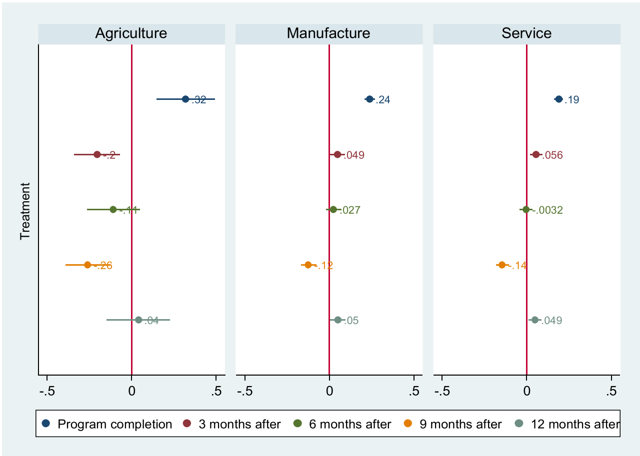


Fig. 8. Probability of being employed, conditional on having a first job and matching on economic sector, and by economic sector.
Sources: FEI and TAD. Note: Static model with varying starting dates, using the 3-month window around November 5th 2014.

persons in the treatment (OfA participants) with jobseekers in the control who have found a job. We match them as before, but also add the sector (defined at 1 digit of NACE) of employment of the first job. Fig. 8.

9. Summary of findings

This study finds that ALMPs, including wage subsidy programs, yield mixed results in terms of their impact on labor market outcomes. That fact, combined with the high cost of this type of policy, makes it imperative that the Federation of Bosnia and Herzegovina invest in robust impact evaluations. Given the weak labor market outcomes of the entity, neither the government nor the citizens can afford programs that are not effective, or even those that require improvement.

Within that context, this study is the first serious attempt to conduct an evaluation in the entity. Given the reticence of the government to carry out experimental evaluations, this study relies on a nonexperimental evaluation, using propensity score matching on observables and a reach dataset that combines administrative data from the Tax Administration office and the Federal Employment Institute. The evaluation is thus not only relevant for the country, given the centrality of the jobs agenda for the government, but it is also innovative in its use of work histories to match treatment and control jobseekers and to assess a richer set of outcomes.

Three main results come out of this exercise. First, based on the analysis conducted so far, the program *Opportunity for All* (OfA) has a positive impact on both the probability of being employed and the intensity of employment. Under the preferred specifications, OfA participants are between 8.3% and 13.1% more likely to be employed 6–12 months after program completion than comparable jobseekers who did not benefit from the program.

Second, the heterogeneity analysis reveals that certain groups of jobseekers benefit more from the program than others. In particular, the program had a higher impact on employment outcomes for older workers and especially for those closer to retirement (age 55–65), the long-term unemployed, and low-skilled workers. For these persons, the impact of OfA on both the probability of being employed (shown) and the intensity of employment (not shown) was higher than for other groups of beneficiaries.

To our knowledge, the only published papers that attempt to make better use of this type of data are Vikstrom (2017) and Cronin et al. (2020). However, there are various additional subjects that this line of research should continue to examine. For example, the dynamic elements of trajectories into and out of employment during the period when jobseekers could access the program could be further explored. One way of doing this would be to work with inverse probability weighting schemes for the matching, which takes into account the probability of changing the labor force status. Another way could be to work with hazard functions, both for a single transition or in a competing model estimation.

Thus, at this stage, while the results obtained so far are informative, they also call for further research. First and foremost, this study could be extended to further explore the dynamic aspects of individuals' work history in the matching and in the impact of the program. Second, it would be extremely valuable to have access to additional administrative data to examine general equilibrium effects. Given the information that the tax administration data contains, it seems possible to measure how many new jobs ALMPs create, or in other words to net out substitution and displacement effects.

Finally, and most importantly, the government of the FBH cannot further delay investing more in robust experimental impact evaluations. Given the amount of resources the government channels to ALMP programs, and given the labor market outcomes, the government would benefit from having a serious evaluation of all ALMPs, comprising various evaluation questions and using various methods, including randomized control trials and other robust methodologies.

Data Availability

Data in support of the findings of this study is available from the corresponding author upon reasonable request.

Appendix A

See Fig. A.1 and Tables A.1–A.6.

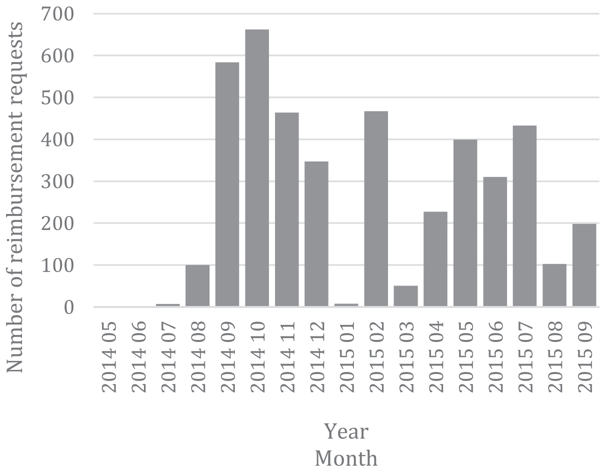


Fig. A.1. Month of claim for first reimbursement. *Note*: Month of claim for reimbursement represents the month when employers presented the first invoice to claim the subsidy.

Table A.1
ALMPs implemented in 2014, with the number of participants and budget.
Source: FEI Report 2014.

Program	Number of participants	Budget (KM)
Opportunity for All	5299	10,431,867
First Working Experience	1890	7128,942
Training Program	698	1206,036
Voucher for Employment Program	439	1300,564
Co-financing of Seasonal Employment	333	280,820
Employment of Roma Population	73	680,000

Table A.2
Probability of participating in Opportunity for All.

treatment	Coef.	Std. Err.	z	P > z	95% Confidence Interval	
Age	0.154	0.014	11.310	0.000	0.127	0.180
Age2	-0.002	0.000	-10.840	0.000	-0.002	-0.002
Female	0.077	0.035	2.210	0.027	0.009	0.145
Education	-0.467	0.106	-4.410	0.000	-0.674	-0.259
Length of last unemployment spell	0.000	0.001	-0.580	0.562	-0.002	0.001
Number of unemployment spells	-0.536	0.032	-16.680	0.000	-0.599	-0.473
Cantons						
Posavski (PK)	-0.063	0.493	-0.130	0.898	-1.030	0.903
Tuzla(TK)	-0.846	0.261	-3.240	0.001	-1.358	-0.334
Zenicko-Dobojski (ZDK)	0.044	0.287	0.150	0.878	-0.519	0.607
Bosansko-Podrinjski (BPK)	0.925	0.544	1.700	0.089	-0.141	1.991
Central Bosnia (SBK)	1.695	0.307	5.530	0.000	1.094	2.296
Herzegovina-Neretva -Neretvanski (HNK)	-2.230	0.395	-5.640	0.000	-3.005	-1.455
West-Herzegovina (ZHK)	0.861	0.521	1.650	0.098	-0.160	1.882
Sarajevo (KS)	-1.281	0.317	-4.040	0.000	-1.903	-0.659
Canton 10	0.499	0.522	0.960	0.339	-0.525	1.523
Canton *Education						
PK*education	0.430	0.238	1.810	0.071	-0.037	0.897
TK*education	0.380	0.124	3.060	0.002	0.137	0.624
ZDK*education	-0.364	0.141	-2.580	0.010	-0.640	-0.087
BPK*education	-0.160	0.266	-0.600	0.546	-0.681	0.360
SBK*education	-0.644	0.151	-4.270	0.000	-0.940	-0.348
HNK*education	0.542	0.178	3.040	0.002	0.193	0.890
ZHK*education	-0.446	0.243	-1.830	0.067	-0.924	0.031
KS*education	0.008	0.148	0.050	0.956	-0.282	0.298
K10 *education	-0.246	0.241	-1.020	0.309	-0.719	0.227
PK*education	0.000					
TK*education	0.000					
_cons	-3.337	0.327	-10.220	0.000	-3.977	-2.697
Pseudo R2	0.0711					

Table A.3
Employment outcomes for OfA participants and non-participants after matching.

	OfA Participants	OfA Non-participants	Difference	p-value
Extensive margin				
Employed at program completion (%)	92.68	65.26	27.42	* **
Employed 3 months after program completion (%)	72.42	60.49	11.93	* **
Employed 6 months after program completion (%)	67.29	59.04	8.25	* **
Employed 9 months after program completion (%)	56.83	55.1	1.73	* **
Employed 12 months after program completion (%)	68.31	55.3	13.01	* **
No. days worked 3 months after program completion (days)	242.46	153.11	89.35	* **
No. days worked 6 months after program completion (days)	310.79	60.49	250.3	* **
No. days worked 9 months after program completion (days)	385.36	278.45	106.91	* **
No. days worked 12 months after program completion (days)	406.77	303.28	103.49	* **
Employed w/same firm 3 months after program completion	72.39	60.19	12.2	* **
Employed w/same firm 6 months after program completion	53.86	21.72	32.14	* **
Employed w/same firm 9 months after program completion	42.23	18.63	23.6	* **
Employed w/same firm 12 months after program completion	48.97	16.87	32.09	* **
Number of jobs 3 months after program completion	1.11	0.94	0.17	* **
Number of jobs 6 months after program completion	1.174	1.03	0.144	* **
Number of jobs 9 months after program completion	1.23	1.11	0.12	* **
Number of jobs 12 months after program completion	1.276	1.19	0.086	* **
Average length of unemployment 3 months after completion (days)	48.59	60.52	-11.97	* **
Average length of unemployment 6 months after completion (days)	55.69	65.87	-10.18	* **
Average length of unemployment 9 months after completion (days)	53.78	70.22	-16.44	* **

Table A.4
Employment outcomes for OfA participants and non-participants with different matching algorithms.

	Outcomes				
Outcome variable	Employed at the program completion (dummy)	Employed 3 months after the program (dummy)	Employed 6 months after the program (dummy)	Employed 9 months after the program (dummy)	Employed 12 months after the program (dummy)
	(1)	(2)	(3)	(4)	(5)
Propensity score matching	0.292 * **	0.126 * **	0.083 * **	0.0325 * **	0.136 * **
(st.error)	(0.0095)	(0.0112)	(0.0114)	(0.0117)	(0.0114)
One-to-one NNM with replacement and caliper 0.001	0.291 * **	0.1208 * **	0.089 * **	0.0286 * **	0.1205 * **
(st.error)	(0.0067)	(0.0113)	(0.0079)	(0.0116)	(0.0115)
Kernel matching	0.2916 * **	0.204 * **	0.0856 * **	0.0302 * **	0.125 * **
(st.error)	(0.0048)	(0.0096)	(0.0079)	(0.0083)	(0.0078)
Regression adjustment	0.292 * **	0.132 * **	0.0857 * **	0.035 * **	0.128 * **
(st.error)	(0.005)	(0.0076)	(0.0079)	(0.0083)	(0.0078)
Inverse probability weighting	0.295 * **	0.136 * **	0.086 * **	0.0322 * **	0.1277 * **
(st. error)	(0.005)	(0.007)	(0.0079)	(0.0083)	(0.0078)
Inverse-probability-weighted regression adjustment	0.295 * **	0.136 * **	0.089 * **	0.0347 * **	0.1305 * **
(st. error)	(0.005)	(0.007)	(0.0079)	(0.00828)	(0.0079)

Table A.5
Program impact on employment outcomes by April 13, 2016.

	Employed by April 13, 2016 (dummy)	Number of days employed by April 13, 2016 (days)	Employed with the same employer by April 13, 2016 (dummy)	Number of jobs by April 13, 2016	Average length of unemployment spell by April 13, 2016
	(1)	(2)	(3)	(4)	(5)
Treatment	0.1207 * **	121.83 * **	0.2739 * **	-1.3044 * **	-14.05 * **
(st.error)	(0.009)	(3.66)	(0.009)	(0.039)	(2.47)
Mean of treatment	0.668	444.2	0.668	1.347	1.7
Mean of control	0.565	328.4	0.414	1.357	24
Treatment effect as a % of control	21.37	37.10	66.18	-96.12	-58.58
No. of observations	55,688	55,688	55,688	55,688	55,688
After matching					
In control	3966	3966	3966	3966	3966
In treatment	3966	3966	3966	3966	3966

Notes: The assumptions correspond to the *Full optimization* model. The control variables are sex, age, age squared, education level, canton dummies, length of last unemployment spell, the education level interacted with canton dummies, and number of unemployment spells since 2009. Abadie–Imbens (2006) standard errors are in parentheses. The exact matching on discrete variables (female) (teffect command in STATA with option *ematch* on female only) has been used.

Table A.6
Impact by canton for select outcome variables.

Cantons	Outcomes	Employed at program completion (dummy)	Employed 3 months after the program (dummy)	Employed 6 months after the program (dummy)	Employed 9 months after the program (dummy)	Employed 12 months after the program (dummy)
		(1)	(2)	(3)	(4)	(5)
Una-Sana Canton	Treatment	0.332 * **	0.119 * **	0.056 *	-0.084 * *	0.170 * **
	(st.error)	(16.49)	(4.80)	(2.15)	(−3.29)	(6.60)
	N	5276	5276	5276	5276	5276
Posavina Canton	Treatment	0.339 * **	0.211 * **	0.0913	-0.0613	0.132 * *
	(st.error)	(7.57)	(4.28)	(1.79)	(−1.21)	(2.66)
	N	885	885	885	885	885
Tuzla Canton	Treatment	0.312 * **	0.152 * **	0.150 * **	0.0731 * **	0.128 * **
	(st.error)	(28.26)	(10.11)	(9.77)	(4.57)	(8.11)
	N	14,510	14,510	14,510	14,510	14,510
Zenica-Doboj Canton	Treatment	0.278 * **	0.131 * **	0.134 * **	0.0856 * **	0.151 * **
	(st.error)	(17.61)	(6.24)	(6.23)	(3.82)	(7.11)
	N	12,813	12,813	12,813	12,813	12,813
Bosansko-Podrinjski Canton	Treatment	0.0808	-0.0429	-0.0581	-0.239 * **	-0.0265
	(st.error)	(1.39)	(−0.61)	(−0.86)	(−3.30)	(−0.40)
	N	464	464	464	464	464
Central Bosnia Canton	Treatment	0.274 * **	0.132 * **	0.0408	-0.0194	0.154 * **
	(st.error)	(15.56)	(5.68)	(1.70)	(−0.78)	(6.66)
	N	4245	4245	4245	4245	4245
Herzegovina-Neretva Canton	Treatment	0.223 * **	0.0383	0.00488	-0.0397	0.0605
	(st.error)	(9.54)	(1.04)	(0.13)	(−0.98)	(1.63)
	N	5890	5890	5890	5890	5890
West Herzegovina Canton	Treatment	0.130 * **	-0.0582	-0.0699	-0.0265	0.0717
	(st.error)	(3.88)	(−1.23)	(−1.46)	(−0.53)	(1.50)
	N	5890	5890	5890	5890	5890
Canton Sarajevo	Treatment	0.223 * **	0.0615 *	0.0326	0.0591 *	0.0793 * *
	(st.error)	(12.77)	(2.46)	(1.28)	(2.33)	(3.25)
	N	15,309	15,309	15,309	15,309	15,309
Canton 10	Treatment	0.285 * **	0.221 * **	0.0641	0.000584	0.239 * **
	(st.error)	(7.41)	(4.68)	(1.11)	(0.01)	(4.47)
	N	1222	1222	1222	1222	1222

References

Abadie, Alberto, Drukker, David, Jane, Leber Herr, Imbens, Guido, 2004. Implementing matching estimators for average treatment effects in Stata. Stata J. 4, 290–311.

Abadie, Alberto, Imbens, Guido, 2006. Large sample properties of matching estimators for average treatment Effects. Econometrica 74 (1), 235–267.

Almeida, Rita, Orr, Larry, Robalino, David, 2014. Wage subsidies in developing countries as a tool to build human capital: design and implementation issues. IZA J. Labor Policy 3 (1), 12.

Bell, Brian; Blundell, Richard; Van Reenen, John , 1999: Getting the Unemployed Back To Work: The Role of Targeted Wage Subsidies. The Institute for Fiscal Studies Working Paper No. W99/12. London, UK.

Benus, Jacob, Rude, James, Patrabansh, Satyendra, 2001. Impact of the emergency demobilization and reintegration project in Bosnia & Herzegovina. *Impaq International*.

Betcherman, Gordon, Daysal, N.Meltem, Pagés, Carmen, 2010. Do employment subsidies work? evidence from regionally targeted subsidies in Turkey. Labour Econ. 17 (4), 710–722.

Betcherman, Gordon; Olivas, K.; Dar, Amit , 2004): Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries. Social Protection Discussion Paper No.0402. Washington, DC: World Bank.

Brown, Alessio J.G., Koettl, Johannes, 2015. Active labor market programs-employment gain or fiscal drain? IZA J. Labor Econ. 4 (1), 12.

Bryson, A.; Dorsett, R.; Purdon, S. , 2002: The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies. Policy Studies Institute and National Centre for Social Research Working Paper No. 4, ISBN 1 84388 043 1 ISSN 1476 3583.

Burger, A., Kluve, J., Vodopivec, M., Vodopivec, M., 2022. A comprehensive impact evaluation of active labour market programmes in Slovenia. Empir. Econ. 62 (6), 3015–3039.

Caliendo, Marco, Schmidl, Ricarda, 2016. Youth unemployment and active labor market policies in Europe. IZA J. Labor Policy 5, 1.

Caliendo, Marco, Mahlstedt, Robert, Mitnik, Oscar A., 2017. Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labor market policies. Labour Econ. Volume 46 (2017).

Card, D., Hyslop, D.R., 2005. Estimating the effects of a time-limited earnings subsidy for welfare-leavers. Econometrica 73 (6), 1723–1770.

Card, David, Kluve, Jochen, Weber, Andrea, 2010. Active labor market policy evaluations: a meta-analysis. Econ. J. 120, F452–F477.

Card, David, Kluve, Jochen, Weber, Andrea, 2018. What works? a meta-analysis of recent active labor market program evaluations. J. Eur. Econ. Assoc.

Crépon, Bruno, Van den, Berg, Gerard, J, 2016. Active labor market policies. Annu. Rev. Econ 8, 521–546.

Cronin, Hugh; Ferrara, Antonella; Geraci, Andrea; Hardiman, Saidhbhín; Judge, Ciaran; Mazzarella, Gianluca; Santangelo, Giulia. (2020): JobPlus Evaluation. JRC Technical Report. European Comission.

Dar, Amit, Tzannatos, Zafiris, 1999. Active Labor Market Programs: A review of the Evidence From Evaluations. Social Protection. World Bank,, Washington.

Galasso, Emanuela, Ravallion, Martin, Salvia, Agustin, 2004. Assisting the transition from workfare to work: a randomized experiment. Int. Labor Relat. Rev. 58 (1), 128–142.

Gertler, Paul J., Martinez, Sebastian, Premand, Patrick, Rawlings, Laura, 2016. Impact Evaluation in Practice, second ed. Inter-American Development Bank and World Bank,, Washington, DC https://openknowledge.worldbank.org/handle/10986/25030 License: CC BY 3.0 IGO.

Groh, Matthew, Krishnan, Nandini, McKenzie, David, Vishwanath, Tara, 2016. Do wage subsidies provide a stepping-stone to employment for recent college graduates? evidence from a randomized experiment in Jordan. Rev. Econ. Stat. 98 (3), 488–502.

Heckman, James; Lochner, Lance; Cossa, Ricardo ,2002): Learning-by-doing vs. on-the-job training: Using variation induced by the EITC to distinguish between models of skill formation. National Bureau of Economic Research Working Paper No. W9083.

Heckman, James, Vytlacil, Edward, 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73 (3), 669–738.

Kluve, Jochen, 2010. The effectiveness of european active labour market policy. Labour Econ. 17 (6), 904–918.

Kluve, Jochen; Puerto, Susana; Robalino, David; Romero, Jose Manuel; Rother, Friederike; Stöterau, Jonathan; Weidenkaff, Felix; Witte, Marc , 2016: Do Youth Employment Programs Improve Labor Market Outcomes? A Systematic Review. IZA Discussion Paper No. 10263.

Martin, John P., Grubb, D., 2000. What works and for whom: a review of OECD Countries' experiences with active labour market policies. Swed. Econ. Policy Rev. 8 (2), 9–60.

McKenzie, D., 2017. How effective are active labor market policies in developing countries? a critical review of recent evidence. World Bank Res. Obs. 32 (2), 127–154.

OECD , 2015: Activation policies for more inclusive labour markets", in OECD Employment Outlook 2015, OECD Publishing, Paris, https://doi.org/10.1787/empl_outlook-2015–7-en.

Oskamp, Frank; Snower. Dennis J. , 2008: The Effect of Low-Wage Subsidies on Skills and Employment. Kiel Institute for the World Economy (IfW) Working Paper No. 1292.

Ravallion, Martin , 2007: Evaluating Anti-poverty Programs. In Handbook of Development Economics, vol. 4 (3787–3846). Elsevier.

Richardson, J. , 1997: Can Active Labour Market Policies Work? Some theoretical considerations. Center for Economic Performance Discussion Paper No. 331 Rosenbaum, Paul R., and Donald B. Rubin. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." The American Statistician 39, no. 1 (1985): 33–38.

Smith, Jeffrey A., 2000. A critical survey of empirical methods for evaluating active labor market policies. Schweiz. Z. fur Volkswirtsch. und Stat. 136 (3), 1–22.

Smith, Jeffrey A., Todd, Petra E., 2005. Does matching overcome lalonde's critique of nonexperimental estimators? J. Econ. 125 (1–2), 305–353.

Vikström, Johan, 2017. Dynamic treatment assignment and evaluation of active labor market policies. Labour Econ. 49, 42–54.

World Bank (2012): World Development Report 2013: Jobs. World Development Report. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/11843 License: CC BY 3.0 IGO.