ROMANIAN JOURNAL OF LABOUR AND SOCIAL STUDIES
Volume 1, No.1, 2016

This is the online published version

Available from Romanian Journal of Labour and Social Studies website
http://www.rjlss.ro/view-current-issue/

Can internship program reduce unemployment among youth in a low job creation environment? The evidence from Macedonia[1]

Nikica Mojsoska-Blazevski[a] and Marjan Petreski[b]

Abstract

The aim of this paper is to assess the effectiveness of an internship program implemented in Macedonia, a country with high unemployment, as part of the active labour market programs. We evaluate two years of program implementation, as to see if some changes made to the program brought better effects. The key research question was whether participation in the internship program increased the probability of participants to find and retain gainful employment. However, in addition to this main outcome, we include other outcome variables such as: inactivity, unemployment, wages, changes in the prospects for employment after the program, changes in financial status, subjective labour market status, etc. To answer such questions (i.e. to assess the program outcomes) we employ a quasi-experimental impact evaluation method. The method, in essence, measures and compares the differences in labour market outcomes between those who participated in the programs (treatment group) and those who did not (control group).

The evaluation (for both years) shows that the program is effective in terms of employment outcomes of the program participants, despite the low overall job creation in the country. Particularly, the program has no significant impact on the participants after the program end (meaning after the end of the internship), but has large positive effects on the subsequent employment. Similarly, the subjective employment is higher for the treatment group. The comparison between the program in 2010 and 2012 suggests that effects are similar, but of a higher magnitude in 2012, when an obligation was introduced for the participating employers to employ at least 50% of the participants. The 2012 program also found significant financial gains for participants, which is not the case with the 2010 program. The results from this analysis provide strong argument for policymakers for increasing the financing and the scale of the program in future years.

Keywords: impact evaluation, internship, unemployment, Macedonia

[a] [b]University American College Skopje

1. Introduction

This paper presents the findings of the impact evaluation of an internship program implemented in Macedonia, as part of the overall active labour market programs (ALMPs) implemented by the Employment Service Agency (ESA). Given that the design of the internship program has been changed between 2010 and 2012, we conduct the evaluation for both years and compare the findings. In particular, in the internship program 2012 an obligation was introduced for the host companies to employ at least 50% of the interns after the termination of the program.

The internship programs targets young people in Macedonia aged up to 29 years, with competed secondary and tertiary education. It aims to provide young people with work experience which should increase their employment probabilities. The position of young people in the Macedonian labour market is quite vulnerable. Whereas the labour market, in general, is characterised by high unemployment (28% in 2014), young people face unemployment rate of 50%. Even more so, their employment rate is very low, 15% in 2014.

In this regard, the key research question is whether participation in the internship increased the probability of participants to find and retain gainful employment. However, in addition to this main outcome, we include other outcome variables such as: inactivity, unemployment, wages, changes in the prospects for employment after the program, changes in financial status, subjective labour market status, etc. Data for the evaluation were gathered through a one-to-one survey that was run during October 2014, covering the participants (treatment groups) and non-participants (control groups). The survey was implemented by the Institute for Political Research-Skopje (IPIS). From methodological viewpoint, we implement an impact evaluation approach by which we compare the employment (and some additional outcomes) of the participants in the program and non-participants.

The structure of the paper is as follows. Section 2 provides a brief description of the Macedonian labour market and the role of the active labour market programs (ALMPs) in general. Section 3 explains the impact evaluation methodology and survey design. Section 4 presents some descriptive statistics and groups' comparisons. Section 5 examines the main findings from the study. Section 6 concludes.

2. **Labour market developments and active labour market programs**

The environment in which active programs operate is very important for their overall effectiveness. In particular, many studies show that ALMPs are less effective (if at all) in a situation of low labour demand (Betcherman et al., 2004), like the Macedonian environment is. On the other hand, some active programs (such as subsidised employment and public works) can contribute towards bringing greater demand for workers, hence increasing overall employment.

The labour market in Macedonia is characterized with relatively high unemployment, and low employment and activity (Table **1**). Unemployment has been on a declining path for almost a decade, but at a slow pace, being at 28% in 2014. At the same time, there has been a steady path of increase in the employment rate, from about 42% in 2008 to 47% in 2014. The activity is still low, especially compared to the EU countries, but with large gender difference, such that males in Macedonia have comparable activity to the males in the EU countries.

Table 1 – Labour market developments 2008-2014, by gender

	Activity rate			Employment rate			Unemployment rate		
	Total	Males	Females	Total	Males	Females	Total	Males	Females
2008	63.5	76.6	50.2	41.9	50.7	32.9	34.0	33.7	34.3
2009	64.0	77.6	50.0	43.3	52.8	33.5	32.3	32.0	32.9
2010	64.2	77.7	50.4	43.5	52.8	34.0	32.2	32.1	32.5
2011	64.2	76.8	51.2	43.9	52.3	35.3	31.6	31.9	31.0
2012	63.9	76.6	50.8	44	52.4	35.3	31.2	31.6	30.5
2013	64.9	76.8	52.7	46	54.5	37.3	29.1	29.1	29.2
2014	65.3	77.7	52.5	46.9	56.1	37.4	28.1	27.8	28.7

Source: Eurostat database

The gender differences are also quite pronounced in the employment rate, where the gap equals close to 20 p.p. The low employment rates of females are related to the large inactivity (only half of the working-age females are engaged in the labour market), and rooted in the still traditional society where women hold the role of second bread-winners, taking care of the household, children and the elderly.

Besides females, young workers are also facing a vulnerable position in the Macedonian labour market. Their unemployment rate reaches 50%, meaning that every

second young person searching for a job cannot find one. The employment rate of young persons aged 15-24 in 2014 was only 15%. Studies show that young people in Macedonia exhibit very difficult school-to-work transition. For instance, the study of Elder et al. (2013) showed that 80% of youth search for a job more than a year (i.e. are long-term unemployed), and that the average time spent in the transition to a job is six years (71.6 months). In the light of the high youth unemployment, the Government has initiated a specific agenda for fight against youth unemployment, through the Action Plan on Youth Employment. The first Action plan was adopted in 2010, and recently the second one has been enacted for the period 2016-2020. *Inter alia*, the Plan puts the young persons (aged 15-29) in the forefront of the ALMPs.

The process of planning, design and implementation of active labour market programs (ALMPs) in Macedonia has been streamlined from 2007 onwards. The planning and implementation of the programs is organized through annual Operational Plans for ALMPs (OPs). The ESA is the implementing body for the OPs. The spending on ALMPs in Macedonia is comparable to that in the countries in the Western Balkan region, and to some of the EU Member States (for instance, Slovakia and Estonia), although being at a generally low level (below 0.1% of GDP)[2]. If spending is compared to the overall unemployment (active jobseekers), Macedonia spends about EUR 50 per jobseeker, whereas lowest spenders among the EU countries in 2012 were Romania (EUR 53 per jobseeker), Croatia (EUR 210) and Bulgaria (EUR 214).

Throughout years, as the country (policymakers) increased its experience with the design and implementation of ALMPs, there has been a growing need (both nationally and from the international community) for monitoring and evaluation of the active programs. The assessment of the effectiveness of active measures is regularly conducted through employment outcomes of different programs/interventions. However, such assessment fails to take into account several important elements: the cost-effectiveness of the programs; potential dead-weight loss; substitution effect; comparison of the employment outcomes with unemployed individuals who have not participated in the program; etc.[3] Only after consideration of these elements, an assessment will show whether the programs are bringing "value for money", i.e. whether taxpayers' money are used effectively, as well as whether the

---

[2] In 2009, as part of the anti-recession package of the Government, the spending on ALPMs was increased to 0.2% of GDP, mainly through increased scope and funding of public works.
[3] The substitution effect indicates whether improved employment prospects for the participants in an ALMP comes at the expense of worsened employment prospects for the non-participants. Dead-weight loss represents a situation when a participant obtains employment through participation in ALMP that s/he would have still got without the program.

internship program can bring real difference to the participating youth. This is a first study of such type for Macedonia.

3. **The impact evaluation methodology and survey design**

3.1. Design and sample

The main objective of this study is to evaluate the effectiveness and efficiency of the internship program implemented in Macedonia. We aim to show if internship program has a potential to reduce youth unemployment in a country with high unemployment. For this purpose, we employ an impact evaluation approach by which we measure the differences in i) labour market outcomes (employment at the time of survey and employment at any time between the end of the active program and the survey date) and ii) subjective wellbeing outcomes (subjective evaluation of the change in the financial situation and chances to find a job before and after program participation) between those who participated in the active programs (*treatment group)* and those who did not (*control group)*.[4]

For a more precise estimate of program effects, it is necessary to "compare the comparable" (Heckman et al. 1999). This means that program participants need to be compared only to those non-participants who could have participated in the program (i.e. had an equal chance to be selected for participation as those who were actually treated). The control group has been chosen out of those who applied on the ESA open call for participation in the internship program, but who were ultimately non-selected. This is plausible, given that the non-selected applicants not only satisfy the conditions of the call, but also they may be similar to the selected participants based on some unobservable characteristics (such as, motivation to participate in a program, or persuasion that program participation is beneficial). According to ESA, after all applicants passed through the eligibility filters/criteria for participation in the program, the selection into actual program participation is done randomly.[5] Out of such defined potential control group, the particular control groups are subsequently selected by means of a matching approach.

---

[4] Both groups being registered as unemployed with the ESA.

[5] In case when an employer asks for interns of certain profile, ESA sends to the employer all the CVs which match the request. The employer chooses an applicant from the CVs. Even if the employer makes his decision by observable characteristics, we address the selection on observables in this analysis, and hence this does not impose a bias. If, however, the employer makes the decision based on unobservables, this would impose a bias, which we do not address. However, even if the employer imposes unobservable factors into the decision, we believe this is not a general case.

In order to be able to identify the causal impact of the ALMPs (namely, the intervention) we created credible control groups through matching non-participants with the participants on relevant observable characteristics. By doing so, the comparison of the employment, wage prospects and subjective well-being of the treatment and the control group enables us to isolate the impact of the program on those prospects. To ensure as much as possible precision in the evaluation, we survey the whole population of the program participants, which was 199 persons in 2010 and 99 in 2012. The size of the control groups (selected as previously explained) was 210 individuals in 2010 and 204 in 2012. The response rates were over 55% with exception of the control group in 2012 where the response rate was low, at 13% (with relatively large rejection, but also incorrect cell phones and/or home addresses).

3.2 Survey questionnaire and outcome variables

Data were collected through questionnaires designed by the researchers, based on an ILO template and the questionnaires used in a similar study implemented by the FREN, in Serbia. Broadly, there are two questionnaires, one for the control groups and one for the treatment groups. In addition to the questions for the control group, program participants (treatment group) answered questions related to their subjective assessment of the programs' usefulness for their future employment. Draft questionnaires were sent for comments to the stakeholders, pre-tested on the pilot sample and adjusted in line with the comments received. Final questionnaire(s) were administered to the selected sample of the treatment and the control groups.

The outcome variables examined in the study (through the questionnaires) are the following:

1. *Employment (i.e. current employment)*: The employment is defined on the basis of the ILO definition, that is all individuals who, in the reference week, performed some work for at least one hour for a remuneration (in cash or in-kind) and employed individuals who in the reference week were absent from work. To this definition, we added farmers and contributing family members.

2. *Non-employment-at-any-time*: Individuals who were unemployed (according to the above definition) at any time after the program's end (including those currently unemployed).

3. *Employed at program's end but currently unemployed:* Individuals who were employed at program's and (and certain period thereafter), but are currently unemployed.

4. *Inactivity:* Individuals who have not searched for a job in the last four weeks.

5. *Hourly wage*: Wage per hour earned on the current job.

6. *Changes in the prospects of employment* after program participation /cut-off point: This indicator is based on the subjective assessment of the respondent. Respondents rated the level of change on a three-point scale, from 1 ("Prospects are better") to 3 ("Prospects are worse").

7. *Changes in financial status* after program participation/cut-off point: Based on the subjective assessment of the respondent. Respondents rated the change on a three-point scale, ranging from 1 ("Financial situation is better") to 3 ("Financial situation is worse").

8. *Subjective labour market status:* This is a self-assessment of the interviewed persons on their current labour market status. The status includes: employed, seasonal/temporary work, self-employed, trying to start-up a business, searching for a job, non-employed but not searching for a job, and involved in education/training.

3.3 Impact evaluation methodology

Any impact evaluation research has to deal with the problem of the counterfactual. This arises because it is impossible to directly observe a single individual in two different statuses (participation and non-participation in a program). Therefore, the main task of an impact evaluation study is to find a valid estimate of the counterfactual.

There are two methods to estimate the counterfactual: randomized experiments and non-experimental (also called quasi-experimental) methods. In principle, randomized experiments provide the most robust method to construct the counterfactual. In randomized experiments, individuals eligible for participation are randomly assigned to the treatment and control group. Since these two groups do not differ from each other (on average) both in observable and unobservable characteristics (i.e. the control group can be considered as "identical" to the treatment group), the average difference in outcomes between the two groups provides a simple answer to the counterfactual question. Often, however, randomized experiments are politically or socially unfeasible and they are not entirely free of estimation difficulties (Heckman et al., 1999).

The internship program of the ESA was not designed as randomized experiments, which substantially lowered the chances to obtain *ex post* a control group with the same average characteristics as the treatment group. Still, the choice of a control group from those who applied (and hence were eligible for the program), but did not participate could mimic a natural experiment and the possibility of finding the treatment and the control group with essentially the same average characteristics was not excluded *a priori*. However, a more

realistic assumption would be that − if additional characteristics did play a role in determining the chances to participate in the internship program − one could not consider the treatment and the control group as "identical". In this case, a simple comparison of mean outcomes (such as employment rates) between the two groups would be insufficient. The evaluation may also give biased results in case of substantial differences between the number of planned and accomplished interviews (i.e. low response rates) in both groups, since the selection of the control group was based on planned, rather than on accomplished interviews.

To assess whether program participation could be regarded as quasi-random, the characteristics of participants and non-participants were compared. Initially, statistical tests of the hypothesis of random assignment to participation were performed (i.e. random differences between the treatment and control group). In particular, we tested statistically whether the means of important socio-demographic characteristics and labour market outcomes were significantly different between treatment and control group. If the hypothesis of random assignment is rejected, it may be actually misleading to compute program's net effects as the difference in the average outcomes between participants and non-participants.

A common technique to solve the evaluation problem when participants and non-participants are not randomly assigned to a labour market program is the matching approach. This approach mimics a randomized experiment *ex post* by constructing a control group that resembles the treatment group as closely as possible. After matching, the probability to be selected for participation in the program of the control group' individuals, on the basis of their observable characteristics, is comparable to the probability of the individuals from the treatment group.

In the dataset there are many variables that presumably influence both the selection into the program and labour market outcomes. Hence, it appears reasonable to assume that selection into the program and labour market outcomes are independent conditional on these observables (the so-called conditional independence assumption). Under this assumption we apply one-to-one *nearest neighbour matching with replacement* and *the nearest neighbour matching with caliper*. Both approaches consist of two steps: (i) an estimation of the individual probabilities to participate in the program, depending on a set of observable characteristics; (ii) matching of participants and non-participants on the basis of these estimated probabilities. One-to-one matching means that each member of the treatment group is matched with a single member from the control group. Nearest neighbour matching means that the pairs are matched according to the minimum distance of the predicted probabilities of program participation, matching with replacement means that the data on individuals in the

control group may be used more than once, provided that they are the nearest neighbour of an individual in the treatment group, and finally, matching with caliper means that control's propensity score belongs to a pre-defined radius.

4. **Descriptive analysis of the sample and the outcome indicators**

This section presents a descriptive analysis and comparison between the program participants (treatment group) and the control group. As previously explained, the analysis in this section presents the raw effects of the programs. Given that the treatment and control groups are likely to differ in their observable and unobservable characteristics, a comparison of their employment outcomes can be biased. In other words, program participants may have better employment outcomes not because of the effectiveness of the programs but because of, say, their better characteristics (higher education levels, prior work experience, shorter unemployment spells prior to program participation, etc.).

Table 2 presents comparative information on several demographic characteristics of the treatment and control group. Average age of the program participants is 31 years in both years, and 26 and 30 years for the control groups, in 2010 and 2012, respectively. Females dominate both in the treatment and control group.

Table 2 - Demographic characteristics of treatment and control group

Indicator	IN10		IN12	
	TG	CG	TG	CG
Age (mean)	31	30	31	26
Gender (in %)				
Males	35.7	43.4	47.4	35.1
Females	64.3	56.6	52.6	64.9
Urban/rural (in %)				
Urban	94.6	88.5	80.8	70.3
Suburban	2.7	4.4	5.1	5.2
Rural	2.7	7.1	14.1	24.5
Highest education today (%)				
Primary and less	0	0	0	0
Secondary, 3-year	0.9	1.8	3.8	1.8
Secondary, 4-year	24.1	32.7	17.9	16.4
General high	4.5	4.4	6.4	2.9
Baccalaureate	59.8	57.5	62.8	67.8

Master's or PhD	10.7	3.5	9	11.1

Source: Own calculations based on survey data.

Urban population dominates in all programs, with shares from 70% (for the control group in 2012), up to 95% for the treatment group in 2010. About quarter of young people in both groups are married. The education structure of the two groups is determined by the eligibility rules by which only secondary and tertiary education graduates can participate in the program. Most of the participants and non-participants in the internship program are holding tertiary education degree (either baccalaureate or master/PhD), followed by those with secondary education.

Data on labour market histories prior to entering the programs/cut-off date (Table 3) show that most of the participants in the program as well as the non-treated individuals had very low unemployment spell prior to program participation. In particular, most of them (more than 85%) were unemployed less than 1 month when entered the program/cut-off date. Though, 4-5% of the participants and the applicants were unemployed for more than 2 years. As expected, few of the participants had some previous work experience.[6]

Table 3 - Labour market characteristics before program's entry/cut-off point, %

	IN10		IN12	
	TG	CG	TG	CG
Unemployment spell (in % of the group)				
< than month	87.5	85	94.9	93
1-6 months	0	0	0	0
6-12 months	6.3	5.3	1.3	0.6
12-24 months	1.8	3.5	0	0.6
24+ months	4.5	5.3	3.8	5.2
Prior work experience (in % of the group)				
Yes	22.3	23.9	10.2	9.9
No	77.7	76.1	89.8	90.1

Source: Own calculations based on survey data.

---

[6] The program design allows participation of young persons with previous work experience, but only if the previous job(s) did not match their educational profile.

Program participants (treatment group) have better employment outcomes relative to the control group in all programs, except in the subsidized employment (

Table **4**). In terms of the current employment (at the interview date), 63% and 67% of the 2010 and 2012 program participants (respectively) were employed were employed at the time of the interview (currently employed). These employment rates are about 20 p.p. higher than those of the non-treated. Program participants are also doing better in terms of the other two outcome indicators: they are less likely to experienced unemployment at any time after program end, and less likely to be employed at the program end but currently unemployed.

Table 4 – Labour market status

Year of program	Current employment		Not employed at any time after/cut-off		Employed at program end, currently not employed	
	TG	CG	TG	CG	TG	CG
2010	63.4	40.7	25.9	43.4	5.4	15.9
2012	66.7	38.6	24.4	43.6	5.1	17.4

Source: Own calculations based on survey data.

Besides the generally better employment opportunities, program participants show more positive attitude towards changes in well-being, as assessed i) by the change in the financial situation from the program beginning/cut-off date until the interview and ii) change in the employment prospects (see Table 5). The exception is the lower share of participants in the 2012 program stating that they experienced positive change in the employment prospects after program participation/cut-off date. Similarly, the subjective (self-reported) labour market status shows much higher employment rate among program participants (above 63% in both years) than the control group (about 40% in both years). These subjective employment rates are very close to the actual employment rates reported in

Table **4**. Regarding the wages, there is no clear pattern of better paid job among participants compared to non-participants.

In summary, analyses in this section show that the internship program in both years deliver positive effects for the participants, both measured by actual employment status, as well as by subjective measures of employment. Though, we do not find any significant difference in the reported wage rates.

Table 5 - Self-assessment of the changes in employment prospects and the financial situation, %

	IN10		IN12	
	TG	CG	TG	CG
Change in financial situation (in % of respondents)				
Better	19.1	12.5	20.5	15.8
Same	65.5	67	68	67.8
Worse	15.5	20.5	11.5	16.4
Change in employment prospects (in % of respondents)				
Better	17.1	5.7	8.7	15.5
Same	70.7	81.4	87	73.8
Worse	12.2	12.9	4.4	10.7
Subjective labour market status (in % of respondents)				
Employed	63.4	40.7	66.7	39
Not employed	36.6	59.3	33.3	61.1

Source: Own calculations based on survey data.

5. Impact evaluation: results and discussion

For a valid measurement of the effects of the internship program, we compare program participants – the *treatment group* – to the non-participants (*control group*), i.e. those who had an equal chance to be selected for participation in the program as the actually treated.

We conduct the analysis and present the results and findings in the following five steps.

First, to assess whether program participation can be regarded as quasi-random, we perform statistical tests of the hypothesis of random assignment to participation. Specifically, we test whether the means of important socio-demographic characteristics and labour market

outcomes are significantly different between the two groups. If the means are statistically different (i.e. the hypothesis of random assignment has been rejected), we move to the second step (below). In case they are not, we observe the difference in the outcome variable between the treated and the non-treated, i.e. we jump to step five.

Second, in order to mimic a randomized experiment *ex post,* we construct a control group that resembles the treatment group by applying *one-to-one nearest neighbour matching with replacement*. This method comprises two steps: (i) an estimation of the individual probabilities to participate to the program, depending on a set of observable characteristics; and (ii) the matching of participants and non-participants on the basis of these estimated probabilities. The first step is conducted through using standard *probit* regression on the treated and the non-treated, whereby the estimated coefficients will provide insights in the factors influencing selection into treatment, but may also capture factors of attrition from the survey, i.e. factors explaining differential non-response rates in the treatment and in the control group.

Third, we apply the one-to-one nearest neighbour matching with replacement by using the estimated parameters from the probit regression of the previous step to predict the probability to participate in a treatment – the so-called *propensity score* – for each individual in the treatment and comparison groups. The propensity scores are used to match participants with comparable non-participants. For each treated individual, we look for the one individual among non-participants who is the closest neighbour in terms of the predicted probability of being treated. To ensure that the matched pairs have reasonably similar probabilities to be treated, we exclude participants for whom the predicted probability to be in the program is larger than for any individual in the comparison group. In this way we achieve *common support*. Alternative matching procedures are used as robustness checks.

Fourth, we conduct evaluation of the matching quality. A way to do so is to compare the standardized bias before matching (SB^{b}) to the standardized bias after matching ($SB^{a)}$. The standardized biases are defined as:

$$SB^b = \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5(V_1(X) - V_0(X))}}; SB^a = \frac{(\overline{X}_{1M} - \overline{X}_{0M})}{\sqrt{0.5(V_{1M}(X) - V_{0M}(X))}}$$

Where $X_1(V_1)$ is the mean (variance) in the treated group before matching and $X_0(V_0)$ is the analogue for the comparison group. $X_{1M}(V_{1M})$ and $X_{0M}(V_{0M})$ are the corresponding values after matching (Rosenbaum and Rubin, 1985). We also re-estimate the propensity score on the matched sample to compute the pseudo-R^2 before and after matching (like in Sianesi, 2004).

Fifth, we study the causal impact of the social programs on labour market outcomes, and the subjective well-being variables. The outcome variables are based on the labour market status after the participation in the program/cut-off date or at the time of the interview, base don the outcomes explained in section 3.3.

We evaluate the internship program in two years 2010 and 2012. As argued in the previous sections, the design of the program slightly changed in-between. In particular, in the internship program 2012 an obligation was introduced for the host companies to employ at least 50% of the interns after the termination of the program. Hence, besides evaluating the impact of the program for participants against the non-participants, we opt to evaluate which of the two designs delivered better results. Therefore, we first analyse each of the two programs separately, while in the third sub-section we analyse them together.

5.1 Impact of 2010 internship program

Table 6 provides early evidence that the treatment and the control samples are systematically different in a couple of aspects; in addition, the Hotelling test rejects the null of joint equality between the means of the two groups at the 10%. In particular, interns are on average 2 years older than the non-treated; more frequently from urban areas; they live in smaller households, in particular where the number of unemployed members in the household is almost twice lower that of the control group; and larger percent of treated own a house. The two samples are different in terms of the outcome variables, treated being better positioned on the labour market than non-treated.

Table 6 - Means' comparison – Internship 2010 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	29.91	32.11	0.04	**
Education today	7.28	7.65	0.15	
Educational level before treatment/cut-	7.23	7.48	0.50	
Gender (1=male)	0.43	0.35	0.46	
Ethnicity (1=Macedonian)	0.87	0.87	0.90	
Ethnicity (1=Albanian)	0.10	0.09	0.92	
Marital status (1=married)	0.33	0.33	0.66	
Geography (1=urban)	0.88	0.97	0.10	*
Geography (1=suburban)	0.04	0.03	0.96	
Geography (1=rural)	0.06	-	0.02	**
Children (1=person has at least one	0.25	0.24	0.59	
Number of children	0.39	0.31	0.31	

Partner's employment status (1=if	0.25	0.29	0.98	
Number of household members	4.12	3.60	0.01	***
Number of household members below	0.50	0.31	0.15	
Number of employed household	1.52	1.87	0.17	
Number of unemployed household	1.74	1.05	0.00	***
Number of retired household members	0.33	0.36	0.95	
House ownership (1=owned)	0.87	0.96	0.02	**
House size (sq. meters)	76.23	73.12	0.32	
Work experience (1=has work	0.24	0.28	0.71	
Work experience (number of months)	96.00	10.29	0.96	
Contract (1=has written contract)	0.13	0.17	0.56	
Salary before treatment/cut-off (MKD)	2,176.99	1,986.67	0.57	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.08	1.96	0.14	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	2.07	1.95	0.23	
Subjective employment status	0.41	0.63	0.00	***
Employment status (1=employed)	0.41	0.63	0.00	***
Unemployed at any time after program's end/cut-off (1=unemployed)	0.43	0.26	0.01	***
Employment status (1=inactive)	0.12	0.10	0.54	
Employed at program's end, but	0.16	0.05	0.01	***
Wage per hour (MKD)	27.00	34.81	0.27	
Hotelling T-squared test H0: The vector of means of all variables are equal across groups				
p-value				0.0572

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

Given the systematic difference between the two samples, d as an intern.

Table 7 presents the results of a probit regression, to check which characteristics of a person lead to higher or lower probability that he/she is selected in the program. The results suggest that older persons have higher probability to be selected as interns, up to a certain age. An additional household member reduces the chance that one is selected as an intern.

Table 7 - Matching – Internship 2010 program

Variable	Marginal effects	p-value	sig

Age	0.10	0.00	***
Age squared	(0.00)	0.00	***
Education today	0.04	0.33	
Gender (1=male)	(0.04)	0.61	
Ethnicity (1=Macedonian)	(0.04)	0.87	
Ethnicity (1=Albanian)	0.07	0.78	
Marital status (1=married)	(0.06)	0.56	
Geography (1=urban)	0.24	0.13	
Geography (1=suburban)	0.16	0.50	
Number of children	(0.04)	0.60	
Number of household members	(0.06)	0.05	**
Work experience (1=has work experience)	(0.06)	0.47	
# Observations	224		
Log-pseudolikelihood	-142.15842		
Pseudo R^2	0.0844		

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

The propensity score obtained from the previous regression is used to match the treatment and control samples, so as to obtain the unbiased impact of the internship programs on the outcomes. These, along the evaluation of the matching quality are presented in Table 8.

Results robustly suggest that the internship 2010 program played sizeable role for the employment prospects of the involved individuals. In particular, the program led to current employment gains of about 25 p.p., both in terms of the actual and perceived employment status. Similarly, the program reduced the incidence of unemployment over the entire post-program period by 21 p.p., compared to non-participants, who actually face almost twice larger probability to stay unemployed over a prolonged period of time. The fact that the program does not exert influence on the employment rate at program's end suggests that the benefits of the program are reaped over longer time horizon only, which in our framework is about four years. Also, findings suggest that the program does not make difference in financial terms, as neither the financial condition nor the wage per hour was found

systematically different between the treatment and control group. The program likely works through a couple of channels: the networking effects; the prevention of certain skills to erode; practicing and acquiring new skills, in particular, soft skills; and building a sense of organizational culture; all hence increasing the chances for faster and likely more suitable employment later. One should note here also that the targeted individuals within the internship program are spread across all income groups of the households in Macedonia.

Table 8 - Effects – Internship 2010 program

Outcome variable	Unmatched					One-to-one matching with replacement					Matching with caliper (0.01)				
	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.96	2.08	(0.12)	(1.50)		1.95	2.11	(0.16)	(1.56)		1.94	2.11	(0.12)	(1.50)	
Bias reduction						Mean	-41%	Median	-41%		Mean	-36%	Median	-47%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	2.07	(0.12)	(1.29)		1.86	2.03	(0.17)	(1.20)		1.81	2.00	(0.19)	(1.35)	
Bias reduction						Mean	36%	Median	45%		Mean	43%	Median	35%	
Subjective employment status (1=employed)	0.63	0.41	0.22	3.42	***	0.64	0.41	0.23	2.34	**	0.64	0.41	0.23	2.38	**
Bias reduction						Mean	-39%	Median	-40%		Mean	-35%	Median	-49%	

Employment status (1=employed)	0.64	0.41	0.23	3.57	***	0.65	0.40	0.25	2.57	**	0.65	0.40	0.25	2.61	***
Bias reduction						Mean	-38%	Median	-40%		Mean	-34%	Median	-49%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.25	0.43	(0.18)	(2.90)	***	0.25	0.46	(0.21)	(2.14)	**	0.25	0.47	(0.21)	(2.16)	**
Bias reduction						Mean	-37%	Median	-40%		Mean	-34%	Median	-49%	
Employment status (1=inactive)	0.10	0.12	(0.02)	(0.59)		0.10	0.09	0.01	0.16		0.10	0.08	0.02	0.32	
Bias reduction						Mean	-50%	Median	-19%		Mean	-45%	Median	-27%	
Employed at program's end, but unemployed now	0.05	0.16	(0.11)	(2.57)	**	0.06	0.14	(0.09)	(1.21)		0.05	0.14	(0.09)	(1.22)	
Bias reduction						Mean	-45%	Median	-40%		Mean	-41%	Median	-49%	
Wage per hour (MKD)	35.12	27.00	8.13	1.15		36.15	29.37	6.79	0.59		36.86	28.83	8.02	0.70	

Bias reduction					Mean	-47%	Median	-40%		Mean	-42%	Median	-49%	

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

5.2 Impact of 2012 Internship

We iterate the same procedure and provide the estimates for the 2012 internship program. Table 9 provides evidence that the treatment and the control samples are systematically different in a couple of aspects; the same has been proven with the Hotelling test.

Table 9 - Means' comparison – Internship 2012 program

Socio-demographic characteristics	mean control	mean treated	p-value	sig
Age (years)	25.91	30.55	0.01	***
Education today	7.70	7.55	0.27	
Educational level before treatment/cut-	7.64	7.26	0.01	***
Gender (1=male)	0.35	0.47	0.07	*
Ethnicity (1=Macedonian)	0.97	0.87	0.02	**
Ethnicity (1=Albanian)	0.02	0.09	0.06	*
Marital status (1=married)	0.23	0.24	0.85	
Geography (1=urban)	0.70	0.81	0.07	*
Geography (1=suburban)	0.05	0.05	0.97	
Geography (1=rural)	0.24	0.14	0.06	*
Children (1=person has at least one	0.13	0.14	0.78	
Number of children	0.17	0.19	0.80	
Partner's employment status (1=if	0.17	0.10	0.14	
Number of household members	3.96	4.27	0.05	**
Number of household members below	0.26	0.21	0.53	
Number of employed household	1.82	2.09	0.07	*
Number of unemployed household	1.61	1.60	0.96	
Number of retired household members	0.21	0.35	0.08	*
House ownership (1=owned)	0.90	0.94	0.27	
House size (sq. meters)	78.21	86.19	0.04	**
Work experience (1=has work	0.10	0.10	0.93	
Work experience (number of months)	2.94	3.64	0.78	
Contract (1=has written contract)	0.06	0.08	0.59	
Salary before treatment/cut-off (MKD)	813.95	788.46	0.95	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	2.01	1.91	0.22	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	1.96	0.96	
Subjective employment status	0.39	0.67	0.00	***
Employment status (1=employed)	0.39	0.67	0.00	***
Unemployed at any time after program's	0.44	0.24	0.00	***

end/cut-off (1=unemployed)				
Employment status (1=inactive)	0.05	0.09	0.31	
Employed at program's end, but	0.17	0.05	0.00	***
Wage per hour (MKD)	21.91	34.95	0.02	**
Hotelling T-squared test H0: The vector of means of all variables are equal				
p-value				0.0000

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

Interns are on average 5 years older than the non-treated; with lower education before treatment, more frequently males or Albanians, from urban areas, they live in larger households (which is the opposite of the 2010 groups), in particular where the number of employed and retired members in the household is higher that of the control group; and live in larger houses. The two samples are different in terms of the outcome variables, treated being better positioned on the labour market than non-treated, on average.

Given the systematic difference between the two samples, we also a probit regression (as for the 2010 program), to check which characteristics of a person lead to higher or lower probability that he/she is selected in the program. We only find that urban inhabitants have higher probability to be selected for participation in the program.

The propensity score obtained from the previous regression is used to match the treatment and control samples, so as to obtain the unbiased impact of the internship programs on the battery of outcomes. These, along the evaluation of the matching quality are presented in the following Table 10. Results robustly suggest that the internship 2012 program also played sizeable role for the employment prospects of the involved individuals. In particular, the program led to current employment gains of about 31 p.p., both in terms of the actual and perceived employment status. Interestingly though, results suggest that interns had quite lower employment rate at the end of the program than compared to the non-treated persons over the same period of time. Still, this may be in line with the finding of the 2010 program whereby unemployed at any time after program's end has been reduced for participants, in a sense that participants face larger employment opportunities in the long run only, i.e. they are not immediately hired by the host company at program end. That this may be the case is confirmed by the last finding whereby participants exhibit larger earnings today compared to non-participants, despite the latter were more frequently employed at program's-end time. Earnings are found more than twice larger, which is a sizeable impact in terms of the financial status that could be ascribed to the program.

5.3 Has Internship 2012 led to further gains over Internship 2010?

The quantitative assessment of the internship program impact in both 2010 and 2012 suggests that both programs brought gains for interns, predominantly over the longer time horizon. In terms of employment, both subjective and actual, both designs led to very favourable results, increasing participants employability by 25 to 31 p.p. (in 2010 and 2012, respectively) compared to non-participants. The better employment probability of participants is slightly higher in the 2012 program. This is probably related to the introduced obligation to host companies to hire at least 50% of the interns. Both designs did not produce short-term impacts. The 2012 program also found significant financial gains for participants.

Although we give a (slight) preference to internship program in 2012 in the above discussion, we still perform a quantitative evaluation to check which of the two exerted stronger impact. To do so, we perform the following exercise: we compare the treated in 2012 with the treated in 2010 and derive the difference in a similar manner as we do when comparing treated versus control groups.

Table 10 - Effects – Internship 2012 program

	Unmatched					One-to-one matching with replacement					Matching with caliper (0.01)				
Outcome variable	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig	treated	control	difference	T-stat	sig
Financial condition after treatment/cut-off (1=better; 3=worse)	1.91	2.01	(0.10)	(1.23)		1.94	2.14	(0.19)	(1.48)		1.91	2.12	(0.10)	(1.23)	
Bias reduction						Mean	-29%	Median	-61%		Mean	-31%	Median	-50%	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.96	1.95	0.01	0.04		1.95	1.90	0.05	0.29		1.95	1.90	0.05	0.28	
Bias reduction						Mean	-22%	Median	-27%		Mean	-5%	Median	23%	
Subjective employment status (1=employed)	0.67	0.39	0.27	4.15	***	0.65	0.35	0.31	2.89	***	0.66	0.34	0.32	3.01	***
Bias reduction						Mean	-29%	Median	-63%		Mean	-31%	Median	-53%	

Employment status (1=employed)	0.67	0.39	0.28	4.24	***	0.65	0.36	0.29	2.79	***	0.66	0.35	0.31	2.90	***
Bias reduction						Mean	-30%	Median	-63%		Mean	-32%	Median	-53%	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.24	0.44	(0.20)	(2.98)	***	0.25	0.42	(0.17)	(1.57)		0.26	0.43	(0.17)	(1.57)	
Bias reduction						Mean	-35%	Median	-63%		Mean	-38%	Median	-53%	
Employment status (1=inactive)	0.09	0.05	0.04	1.11		0.08	-	0.08	2.54	**	0.09	-	0.09	2.55	**
Bias reduction						Mean	-29%	Median	-61%		Mean	-31%	Median	-50%	
Employed at program's end, but unemployed now	0.05	0.18	(0.12)	(2.67)	***	0.06	0.22	(0.17)	(2.23)	**	0.05	0.22	(0.17)	(2.26)	**
Bias reduction						Mean	-28%	Median	-63%		Mean	-31%	Median	-53%	
Wage per hour (MKD)	34.95	22.04	12.91	2.24	**	36.08	16.46	19.62	2.70	***	35.70	14.30	21.40	2.86	***

Bias reduction					Mean	-28%	Median	-63%		Mean	-29%	Median	-53%	

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

Designed in this way, the treatment dummy captures two influences: i) potential differential results between Internship 2010 and Internship 2012; and ii) differential macroeconomic and labour market conditions.

Table 11 - Means' comparison – Internship 2012 versus Internship 2010 programs

Socio-demographic characteristics	mean IN10	mean IN12	p-value	sig
Age (years)	31.21	30.55	0.72	
Education today	7.55	7.55	0.99	
Educational level before treatment/cut-	7.41	7.26	0.30	
Gender (1=male)	0.36	0.47	0.11	
Ethnicity (1=Macedonian)	0.89	0.87	0.66	
Ethnicity (1=Albanian)	0.08	0.09	0.82	
Marital status (1=married)	0.32	0.24	0.24	
Geography (1=urban)	0.95	0.81	0.01	***
Geography (1=suburban)	0.03	0.05	0.41	
Geography (1=rural)	0.03	0.14	0.01	***
Children (1=person has at least one	0.24	0.14	0.08	*
Number of children	0.32	0.19	0.13	
Partner's employment status (1=if	0.27	0.10	0.00	***
Number of household members	3.68	4.27	0.00	***
Number of household members below	0.29	0.21	0.31	
Number of employed household	1.78	2.09	0.05	*
Number of unemployed household	1.22	1.60	0.03	**
Number of retired household members	0.34	0.35	0.94	
House ownership (1=owned)	0.96	0.94	0.57	
House size (sq. meters)	76.19	86.19	0.03	**
Work experience (1=has work	0.22	0.10	0.02	**
Work experience (number of months)	8.82	3.64	0.14	
Contract (1=has written contract)	0.14	0.08	0.14	
Salary before treatment/cut-off (MKD)	1,419.64	788.46	0.22	
Outcome variables				
Financial condition after treatment/cut-off (1=better; 3=worse)	1.96	1.91	0.53	
Chances to find a job after treatment/cut-off (1=better; 3=worse)	1.95	1.96	0.96	
Subjective employment status	0.63	0.67	0.64	
Employment status (1=employed)	0.63	0.67	0.64	
Unemployed at any time after program's end/cut-off (1=unemployed)	0.26	0.24	0.81	
Employment status (1=inactive)	0.10	0.09	0.84	
Employed at program's end, but	0.05	0.05	0.94	
Wage per hour (MKD)	34.81	34.95	0.98	

Hotelling T-squared test H0: The vector of means of all variables are equal	
p-value	0.0114

Source: Authors' calculations based on survey.

*Note: *, ** and *** signify that the difference is statistically significant at the 99, 95 and 90 percent level, respectively.*

Table 11 presents the tests of the means differences and provides evidence that the two groups are still different in terms of the average observable demographics. This may suggest that the targeting of the two programs actually changed between the two years, or simply the program became more popular, so that the amount and variety of (selected) applicants increased. However, the insignificance of the differences in the outcome variables may actually suggest that, even if the targeting may have changed, the outcomes may have not.

Still, to verify this quantitatively, we match the two groups and then provide the average treatment effect. The calculations are presented in Annex 1. We find some evidence of changed targeting of the program. Younger individuals were more frequently a target of the 2012 program than compared to the 2010 one. Then, individuals with more education, of other ethnicity than Macedonian and Albanian, and from households with more household members had higher chance to be selected in 2012 than compared to 2010. If ESA did not change the targeting intentionally, this may indeed suggest that the announcement of the 2012 program had larger reach and more "different" applicants (both interns and companies) applied and became part of the program. Results after the matching suggest that the impact of the 2012 internship program has been hardly stronger than the one in 2010, in quantitative terms (see Table A1.2 in Annex 1). When the matching with caliper is pursued, results suggest that the 2012 program is better in terms of employment, both actual and subjective, but the finding is only significant at the 10%. In addition, the caliper matching does not lead to any important efficiency gains over the usual nearest-neighbour matching.

In qualitative fashion, still 2012 internship produces slightly more gains for interns than compared to the 2010 design and hence should be maintained. Note that the results disprove the claim that macro and labour-market conditions in 2012 significantly differed from those in 2010, since had they differed it would have been in positive direction only (i.e.

2012 cannot be claimed to have been worse economic year than 2010). Hence, it is unlikely that worse economic conditions compensate the potentially positive effect of 2012 program over the one in 2010. The qualitatively better results for the 2012 program may be actually revealing the better administration of the program (learning curve) and confirming our initial notion that over time the pool of applicants (and selected applicants) and host companies may have become more diverse in both observable and unobservable characteristics. Though, we cannot claim that the better outcomes of the 2012 program are related to the obligation introduced in 2012 that employers hire at least 50% of the interns at the program end.

6. Conclusion

This report presents the findings of the impact evaluation of the internship program implemented by the Employment Service Agency (ESA) of Macedonia in 2010 and 2012. We evaluate two years of program implementation, as to see if some changes made to the program brought better effects. The key research question was whether participation in the internship program increased the probability of participants to find and retain gainful employment. However, in addition to this main outcome, we include other outcome variables such as: inactivity, unemployment, wages, changes in the prospects for employment after the program, changes in financial status, subjective labour market status, etc.

To answer such questions (i.e. to assess the program outcomes) we employ a quasi-experimental impact evaluation method. The method, in essence, measures and compares the differences in labour market outcomes between those who participated in the programs (*treatment group*) and those who did not (*control group*).

The evaluation (for both years) shows that the program is effective in terms of employment outcomes of the program participants, despite the low overall job creation in the country. Particularly, the program has no significant impact on the participants after the program end (meaning after the end of the internship), but has large positive effects on the subsequent employment. Similarly, the subjective employment is higher for the treatment group. There is difference between the two groups in employment after program end/cut-off date only in 2012. The comparison between the program in 2010 and 2012 suggests that

effects are similar, but of a higher magnitude in 2012. The 2012 program also found significant financial gains for participants, which is not the case with the 2010 program.

These results show that employers do not immediately hire the interns, but once there is a need for a worker, their former interns are first to hire.[‡‡] In addition, the interns have higher probability of employment relative to young persons who were not taking part in this program as employers are favouring workers with previous work experience. In general, better employment prospects come from different channels such as networking while on internship, new skills gained, initial on-the-job training received, but also maybe the signalling effect (interns signalling to employers that they are of better quality, more motivated, etc. compared to their peers). Our results do not show the presence of the so-called internship "trap" whereby young persons are trapped into internships, moving from one to next, without being hired, and only substituting the regular staff. Given that the programs in the two years have similar effects, we propose that the obligation for employers to hire 50% of the interns is lifted, as it may be a constrain to the interest of employers to demand interns. As a word of notion, though, the program is currently running on a very small scale as to make larger effect on the overall employment rate in the country. However, the results from this analysis provide strong argument for policymakers for increasing the financing and the scale of the program in future years.

[‡‡] We cannot assess with the data if interns are hired by the same company where they were on internship.

References

Betcherman, G., Olivas, K. and Dar, A. (2004) *Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries.* World Bank Social Protection Discussion Papers Series, No. 0402.

Elder, S., Novkovska, B. and Krsteva, V. (2013) Labor Market Transitions of Young Women and Men in the former Yugoslav Republic of Macedonia. *Work4Youth Publication Series*, No. 1.

Heckman, J., LaLonde, R. and Smith J. (1999) The Economics and Econometrics of Active Labour Market Policy, in Ashenfelter, O. and Card, D. (eds.), *The Handbook of Labour Economics*, Volume III, Amsterdam: Elsevier Science.

Rosenbaum P. R. and Rubin D.B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *The American Statistician,* 1985, Vol.39, No1).

Sianesi, B. (2004): An Evaluation of the Active Labour Market Programs in Sweden, *The Review of Economics and Statistics*, 86(1), 133-155.