Vocational training and labour market outcomes: Evidence from Youth Guarantee in Latvia

Massimiliano Bratti, Corinna Ghirelli, Enkelejda Havari, Giulia Santangelo*

European Commission, Joint Research Centre

in collaboration with

Jānis Leikučs, Normunds Strautmanis†

Ministry of Finance, Latvia

June 7, 2017

Preliminary, not for quotation or circulation

*Centre for Research on Impact Evaluation (CRIE), European Commission, Joint Research Centre (JRC)

†Evaluation Division, EU Funds Strategy Department, Ministry of Finance of the Republic of Latvia

Abstract

The aim of this study is to evaluate the impact of a vocational training (VT) programme implemented in Latvia in 2014 and targeted at youth unemployed. The training programme is part of the Youth Guarantee schemes supporting young people aged 15-29 who are not in education, employment or training (NEETs). For the first wave of the programme, we expoit the eligibility rules, in a Fuzzy Regression Discontinuity Design (FRDD), where only NEETs aged up to 24 years could enroll in VT courses, to identify the causal effect of VT. For the second wave of the programme, which targets NEETs up to 29 years, we use and compare Propensity Score and Coarsened Exact Matching estimators. Estimated results on the employability of the youth show that in both waves the participation in the vocational training programme has positive but statistically insignificant effects.

Foreword

The Counterfactual Impact Evaluation (CIE) of the vocational training programme implemented in Latvia under the Youth Guarantee schemes was carried out within the "Data Fitness Initiative for CIE," launched in February 2016 by the Directorate General Employment, Social Affairs and Inclusion (DG EMPL) and the Centre for Research on Impact Evaluation (CRIE) to promote the use of CIE for the assessment of European Social Fund (ESF) interventions. Based on the quality of the data and on the policy relevance of the intervention proposed, in June 2016 the Latvian data were selected by CRIE to establish a collaboration agreement with the Evaluation Division of the EU Funds Strategy Department of the Ministry of Finance of the Republic of Latvia, and work together on the evaluation of the programme. This collaboration resulted very fruitful, both for strengthening interactions between the ESF Managing Authorities and the European Commission, and in terms of the scientific contribution to the evidence on the impact of ESF interventions.

Acknowledgements

CRIE would like to thank the Evaluation Division of the EU Funds Strategy Department of Latvia for granting access to and collecting the data used in this report from the State Employment Agency and the State Revenue Service.

1 Introduction

Given the high risk of unemployment faced by youth, active labor market policies (ALMP) targeted at this specific group, and the related evaluation studies, are widespread in Europe. In the European countries, in fact, in the post-crisis period the youth unemployment rates have reached, and stabilised at, around 20% on average. From the second quarter of 2008, the youth unemployment rate has taken an upward trend peaking at 23.9% in the first quarter 2013, before receding to 19.7% at the end of 2015. In 2013 it reached the highest and lowest values in Greece (58.3%) and Germany (7.8%) respectively, while in Latvia it attained the 23.2%.[1]

In order to reduce the levels of youth unemployment in the worst affected regions, since 2013 collective and centralised efforts of the European Union added to the national initiatives, as the English New Deal for Young People (NDYP), the Danish Youth Unemployment Program (YUP) and the German Jugend mit Perspektive (JUMP).

In February 2013 the European Council created the Youth Employment Initiative (YEI) Package to increase the EU financial support available to the regions and individuals suffering most from youth unemployment and inactivity. The YEI typically subsidises the provision of apprenticeships, traineeships, job placements and further education leading to a qualification. It exclusively supports young NEETs, including long-term unemployed youngsters or those not registered as job-seekers, in regions experiencing youth unemployment rates above 25%. The YEI package can also be used to support the implementation of the Youth Guarantee (YG) schemes. One of the common aspects within the current YEI in Europe is for instance the preparation of customised analyses of the needs of unemployed young people, together with the crucial role played by the Public Employment Services (PES) in providing these services.

Although the YG national experiences are quite well documented (Cabasés Piqué, Pardell Veá and Strecker 2015, Pastore 2015, Escudero and López Mourelo 2015, among others), evidence of the effectiveness of the recent ALMP financed through the YEI in Europe has not been yet established in the literature. These measures were introduced very recently and the data collection process needed to perform a rigorous impact evaluation is still ongoing.

The aim of this study is to evaluate the impact of a vocational training programme implemented in Latvia in 2014 and targeted at unemployed youth. It is part of the Lavian YG, which are financed by YEI, ESF and Latvian State Budget and managed by the Latvian State Employment Agency.[2]

[1] Eurostat, Employment and unemployment: Labour force survey, 2017.

[2] In Latvian the programme of interest is called "JG Profesionalas apmacibu programmas". The group of YG

The scientific knowledge on the YG measures mostly relies on the evaluation studies conducted on the Nordic countries who activated these programmes in the 1980s and the 1990s. These programs in general entail age-specific eligibility rules. Sweden, introduced the first youth guarantee in 1984, Norway in 1993, and Denmark and Finland in 1996. A similar scheme, known as New Deal for Young People (NDYP), was implemented in the UK in 1998 to target unemployed youth in the age group 18–24.

Studies on the Nordic countries find moderate effects in the short-term and negligible effects in the long term. For Sweden, Carling and Larsson (2005) show that the reform passed in 1998 contributed to positive effects in youth employability in the short term but had no impact in the long term. A very recent paper by Halmalaine et al. (2014) examines the Youth Guarantee programme introduced in Finland in 2005. The reform consisted of an early intervention, monitoring and individualized job search plans for unemployed young persons. Using the age threshold set at 25 years, they find that the YG adopted in 2005 moderately increased unsubsidized employment while having a negligible impact on unemployment in the age range of 23-24. Further, estimates by educational level show that the reform did not improve the labour market prospects of unskilled youth.

One of the few papers that show positive effects both in the short and the long-run (though moderate) is Blundell et al. (2004) who evaluate the NDYP programme implemented in the UK. This programme introduced extensive job assistance and wage subsidies to employers and affected several million of young people. It was first piloted in given areas and then extended to others. The authors exploit particular features of the programme such as area and age-eligibility criteria that vary across individuals, and find that the impact of the program significantly raised transitions to employment by about 5 percentage points.

In most of the aforementioned programmes intensified counselling during a job-search period has been proved to positively affect employment. On the other hand, the literature has highlighted that incentives to increase participation in ALMP also matter. Indeed, one plausible reason why people fail to participate is that they do not foresee high returns from attending the training programmes.

activities (professional training programs) to which the programme belongs is referred to as "Profesionala izglitiba, auto un traktortehnikas vaditaja apliecibas iegusana".

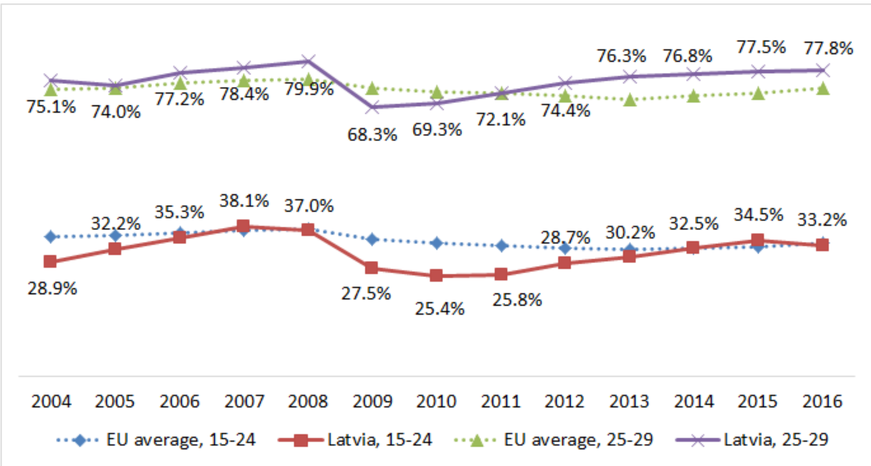
2 Youth unemployment in Latvia

Within the first three quarters of 2016, in Latvia there were on average 44,000 of young people in the age group 15–29 not in employment, education or training (NEETs), of which half in the age group 15–24.[3]

NEETs aged 15–24 and NEETs aged 15–29, correspond respectively to the 11.2% and 13.3% of the total population of the same age group. The peak was registered in 2009, as a consequence of the global financial crisis where the share of NEETs in the age group 15–29 reached 20.8%. In the years after there was a decreasing trend with the rate of NEETs declining from 19.1% in 2011 to 13.8% in 2015.

In general, labour market conditions of youth in Latvia have improved in the recent years, with an increase in the employment rate for all the age groups. As shown in Figure 1, the employment level of youth aged 15-24 has increased by 5 percentage points (pp) since 2012, falling behind the EU average of 33.7% for 15-24 by only 0.5 pp at the end of the third quarter of 2016. By contrast, the employment rate of those aged 25-29 has increased by 3 pp within the same period, thus exceeding the EU average of 73.1% for 25-29 by 5 pp.

Figure 1: Employment rate of youth aged 15-24 and 15-29 in Latvia compared to the EU average.



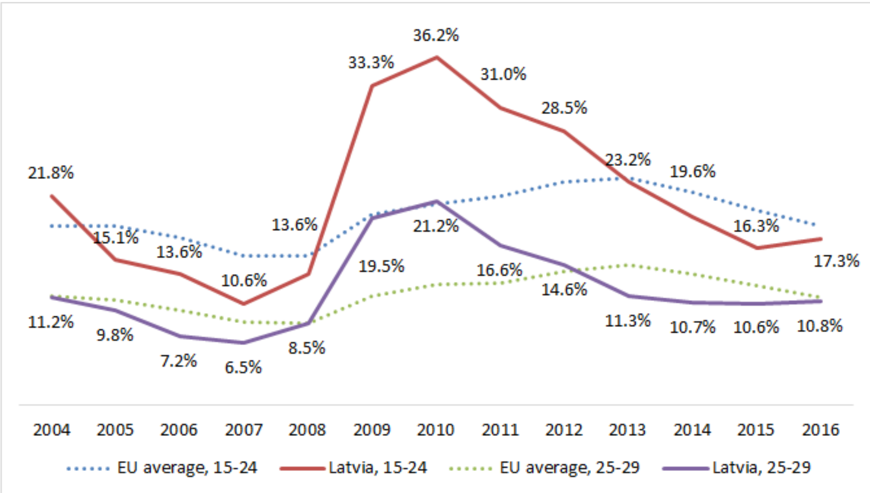
Data source: Eurostat.

From Figure 2 we can see that the unemployment rate of youth aged 15-24 has decreased by 11 pp since 2012 remaining 1 pp below the EU average of 18.7% for 15-24 at the end of the third quarter of 2016. Likewise, the unemployment rate of youth aged 25-29 has decreased by 4 pp within

[3] In 2015 NEETs counted 6% more compared to 2016 and in 2014 they counted 18% more with respect to 2016.

the same period of time, again exceeding the EU average of 11.2% for 25-29 by almost 1 pp.

Figure 2: Unemployment rate of youth aged 15-24 and 15-29 in Latvia compared to the EU average.



Data source: Eurostat.

Different features emerge when comparing the unemployment rate of youth aged 15–24 with the total unemployment rate. First, the youth unemployment rate is about twice as large as the total unemployment rate, and this is true for Latvia as well (17.7% and 9.5% respectively in the third quarter of year 2016). This can be partly explained by the official statistical definition of unemployment rate, calculated as the ratio on economically active individuals. Although the number of unemployed youth is lower than the total number of unemployed, when compared to the corresponding number of active people in the corresponding age-group, the rate of unemployed youth turns out to be higher than the total rate of unemployed since a large number of youngsters are in fact economically inactive (namely they are in education, and are not considered as being economically active), while in the total population a large part is economically active (meaning they finished their studies and started to work or to look for a job, becoming part of the labour force). Second, in recent years, despite a decline in the number of unemployed youth, the rate of youth unemployment in Latvia increased, going from 16.3% in 2015 to 17.3% in 2016.

As for the Latvian labour market in general, the increase in the total employment rate observed in the last years (reaching 72.3% in 2015 for people aged 20–64), despite a decreasing labour supply, could be explained by a sizeable decline of 15% of the working age population (the highest value registered in the EU). This is due to both a negative natural demographic balance (i.e. the difference between births and deaths) and high emigration of youth (mostly as a result of better labour market conditions abroad). Youth outward migration is considerable, as more than 40% of the emigrants are in the age group 20–35.[4]

Concerning the level of education of unemployed youth, a relatively large proportion of young people enter the Latvian labour market without any professional qualification (i.e. with only basic or general secondary education). According to the authors' view, this could indicate shortcomings in the career guidance system or limited access to post-secondary education. Although the tertiary education attainment rate is high (well above the Europe 2020 target of 34–36%), the supply of university graduates in knowledge intensive sectors engineering and mathematics (or more generally in STEM) fields (17.9% in 2013) remains among the lowest in the EU. With respect to the vocational education and training (VET) system, this has been reformed over the years but the challenges remain in terms of updating of work-based learning components and curricula. Moreover, apprenticeship type schemes are considered to be underdeveloped in Latvia.[5]

In numbers, labour force statistics by age and educational attainment level (International Standard Classification of Education, ISCED 2011) show that in Latvia in 2016 the unemployment rate for the age group 15–24 was 27.2% for people with less than primary, primary and lower secondary education (ISCED levels 0-2), 14.7% for people with upper secondary and post-secondary non-tertiary education (ISCED levels 3 and 4) and 16.2% for people with tertiary education (ISCED levels 5-8). Analogously, the unemployment rate for the age group 25-29 was 16.3% for ISCED levels 0-2, 14.4% for ISCED levels 3 and 4 and 5.8% for ISCED levels 5-8.[6] For both age groups, therefore, the highest rates of unemployment are observed for people with the lowest levels of education.

Given Latvia's demographic challenges, the activation of young people is especially crucial. The field work to reach out to young people not in education, employment and training has been considerably delayed; as a consequence the labour supply potential of young people is not fully utilised.[7]

Coverage of the unemployed by active labour market policy measures remains low. For instance, only 10.4% of the registered unemployed were activated in 2014. The funding for ALMPs was reduced in 2015, but it increased again in 2016. Against the backdrop of declining unemployment, this should increase the coverage of ALMPs for 2016. Currently, the largest share of ALMP spending

[4] European Commission, Country Report Latvia 2016.

[5] Ibid.

[6] Eurostat, European Union Labour Force Survey (EU-LFS) series - detailed annual survey results.

[7] European Commission, Country Report Latvia 2016.

is dedicated to the YG (39%) and other training (39%) measures.[8]

The registered unemployed youth represents on average 60% of all unemployed and 35% of the total number of young NEETs in the country. At the end of 2016, 15,072 unemployed aged 15–29 were (7% or 1,150 persons less compared to year 2015) registered at the State Employment Agency, 40% of whom were aged 15–24. Since the start of the YG, i.e. from 2014 to 2016, more than 111,000 people aged 15–29 took part in the programme's activities.

3 Youth Guarantee and vocational training in Latvia

The programme under analysis is part of the YG programme and is financed by the ESF, YEI and the Latvian budget for a total budget of 9.2 million Euros (Latvian Ministry of Finance).[9]

According to Art. 16 of the ESF Regulation,[10] the YEI shall target "all young persons under the age of 25 NEETs residing in eligible regions, who are inactive or unemployed including the long term unemployed, and whether or not registered as seeking work. On a voluntary basis, MS may decide to extend the target group to include young persons under the age of 30".[11] Given the high unemployment rate of this age group, in Latvia the possibility to participate was extended also to young people aged up to 29 years.

The intervention: In Latvia the Youth Guarantee is the biggest support program for youth aged 15–29. The programme of interest for this evaluation is a Vocational Training Programme (VTP). It is implemented by the Latvian State Employment Agency (SEA). The programme aims at youth acquiring or increasing their vocational qualifications in accordance with the labour market demand. It offers a number of different training courses which are organized in a voucher system: that is, young unemployed receive a voucher which can be spent in one of the vocational education institutions in the country. After passing a final examination, participants receive a certification which confirms the acquired professional qualification. Classes consist on average of 10-12 students. The length of training courses varies from 3 up to 9 months. The starting and ending dates of participation can vary across participants. During the training programme participants receive

[8] Ibid.

[9] The programme is ongoing and it is supposed to continue until January 2018.

[10] The Regulation can be accessed here: link to REGULATION (EU) No 1304/2013, 17 December 2013.

[11] European Commission, Guidance on implementing the Youth Employment Initiative, European Social Fund thematic paper 2014.

a monthly allowance of 100 Euros and eventually a reimbursement of the travel costs related to commuting if they wish to attend a course that it is not available in their area of residence.

The programme started on 1st January 2014.[12] While the programme is ongoing until 2018, this evaluation considers the participation period from the start of the programme (January 2014) until December 2015.

The programme of interest was advertised also through a pilot project implemented from March to December 2015 on outreach and awareness raising activities for young people on YG measures in Finland, Latvia, Portugal and Romania. The activities comprised press publications, meetings with journalists, regional visits and press conferences. Moreover, the Latvian SEA publishes information on different measures available for young NEETs on a regular basis.

Eligibility criteria: The intervention targets young NEETs aged 15-29 years. However, the programme can reach only young NEETs who register as unemployed at the SEA. Since (i) registration in SEA unemployment list is pre-requisite to access the programme and (ii) the unit of observation in the analysis are registered unemployed, in the remainder of the text we refer to young unemployed as our population of interest rather than young NEETs. A young unemployed shall be involved in the acquisition of a vocational education programme if:

- the vocational qualification acquired previously by the unemployed or their professional experience is not demanded in the labour market or it does not conform to the requirements laid down for the relevant profession and, therefore, it is impossible to find appropriate work;
- he/she has lost his or her vocational skills;
- he/she has not previously acquired a vocational qualification.

Since the intervention targets NEETs registered as unemployed, hereinafter participants will be referred to both as NEETs and as unemployed.

3.1 The voucher system: practical implementation

The system was developed in different steps.

[12]According to the Rules of Cabinet laying down the provisions of the measure, the intervention was financed as from 1st January 2014 and first participants were accepted already in February. The regulation was adopted on 28th April 2015.

1. *Registration*: in order to promote efficient and targeted provision of the measures offered by the SEA to the unemployed, the SEA carries out the profiling of unemployed, which includes the determination of the most suitable available active employment measures for the unemployed and the preferable sequence for receiving the measures. Such profiling takes place in a meeting between the unemployed and the SEA officer. During the meeting the unemployed applies for participation in the training measures. The SEA officer checks that the unemployed satisfies the eligibility requirement for participation before registering the application in the SEA database.

2. *Selection of a programme*: the unemployed may choose a suitable programme from the list of training programmes (approximately 75 vocational training programmes and 60 non-formal training programmes).

3. *Voucher receipt*: the SEA officer makes a phone call and invites the unemployed to receive a training voucher. The voucher consists of 2 parts: one is for the training provider and the other should be returned to the SEA officer. It contains information on the maximum amount of expenses covered by the SEA.

4. *Choice of the training provider*: the unemployed selects a training provider within the first 10 working days after the SEA's job search assistance. The choice is made from the list of procured training providers published on the SEA's website.[13] The training provider shall determine the suitability of a person for participating in a training programme.

5. *Bringing back the voucher*: once the unemployed and the training provider have signed the agreement, the latter fills in the voucher, signs it and returns it to the unemployed, who brings it back to the SEA officer within one month before the voucher expires.

6. *Contract*: the SEA officer prepares an agreement with the unemployed and with the training provider. The contract specifies, among others, the provisions and the time of the training, the mutual duties and rights during the training, and the provisions for interruptions and termination of the training. It also specifies the organization of the final examinations.

7. *Training*: The training has to start within one month from the signature of the training

[13] However, other training providers may also be selected, if they are ready to make an agreement with the unemployed and follow the next procurement procedure.

voucher. The SEA officer controls the quality of the training services and the client satisfaction.

4 Data

This study is based on a database obtained by merging data from the SEA with data from the State Revenue Service (i.e. the State Tax Authority).[14] The SEA is in charge of the intervention and provides information about participants in the training programme.

Administrative databases from the SEA provides us with data on the NEETs registered as unemployed in any given period between June 2013 and December 2015. These include those participating in the vocational training programme (*treated units*) and those who did not participate in the vocational training programme nor in any other training programme at SEA (*control units*). Both groups received job-search assistance after the registration in the SEA. The case-workers performed a screening of the profiles of the registered unemployed based on their qualifications, age, etc., in order to check their needs and eventually their eligibility for the training courses.

Data from the SEA include the following information: individual characteristics such as gender, exact birth date, residence, nationality, highest education attained, unemployment starting date.

As for the participants (treated units), information regarding the vocational training programmee includes the starting date and the ending date of the training, the type of the attended training course, whether it was completed or not (dropout), and if one participated in another programme after having completed the training. It is also possible to observe if the individual had participated in another programme under the YG package before participating in the considered training. For participants, the exact day when they found a job after the training is also available. However, this information is not available for the non-participants.

The administrative data from the State Revenue Service reports information on labour market performance of each individual at specific dates.[15] This allows us to define an indicator of formal employment at specific points in time. For individuals who are formally registered as employed in the State Revenue Service database, it is also possible to observe the wage, the sector of activity

[14]We would like to thank the Evaluation Division of the EU Funds Strategy Department of Latvia for granting access to and collecting the data from the SEA and the State Revenue Service.

[15]State Revenue Service database is based on employers' monthly report on employees' insurance, income, working hours, firm sector (Statistical Classification of Economic Activities in the European Community, i.e. NACE category), firm size.

and the size of the firm. This information has been extracted for each individual in the sample at the following dates: January 2012, June 2012, December 2012, June 2013, December 2013, June 2014, December 2014, June 2015, December 2015, June 2016. Since the intervention starts in January 2014, the information collected between January 2012 and December 2013 is used to construct pre-intervention measures of labour market career (e.g., employment status, income, social contributions) of individuals. Data collected in December 2015 and June 2016 serve as outcome variables to evaluate the labour market performance of the individuals after the intervention.

Data from the SEA are hence merged with data from the State Revenue Service to obtain information on labour market status both in the pre- and post-intervention period for all individuals in the sample (the treated and the control group).

The sample consists of youth aged 15–29, registered in the SEA since 2009.

The database contains one unemployment spell for each individual: for all individuals the starting date of the unemployment spell is known. Due to data limitations, we cannot observe the exact duration of the unemployment spell.[16] Hence, we use the information on earnings observed in January 2012, June 2012, December 2012, June 2013, and December 2013 (i.e. until the start of the programme on 1 January 2014) and keep in the sample the individuals who did not earn a wage at the aforementioned dates in the period between the start of the unemployment spell and 1st January 2014, so as to be reasonably sure that we are considering one continuous unemployment spell for each individual.

5 Sample definition

The initial sample is composed of 1,896 treated units and 39,253 control units. The programme started officially on 1st January 2014. Although the programme is ongoing and will continue until 2018, in this analysis we restrict the participation window from January 2014 to December 2015, based on data availability.

Figure 3 and Figure 4 show the distribution of the participation starting date for all treated units and the age at the start of the programme, respectively.

Two features emerge. First, participation in new training courses was interrupted between

[16]As we said, we know individual employment status at some fixed dates only.

Figure 3: Participation starting date

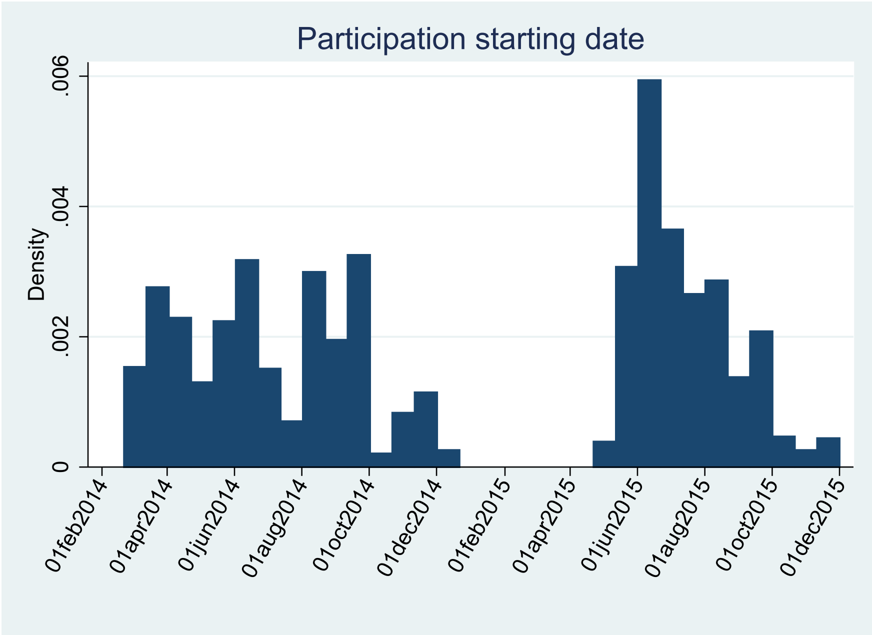
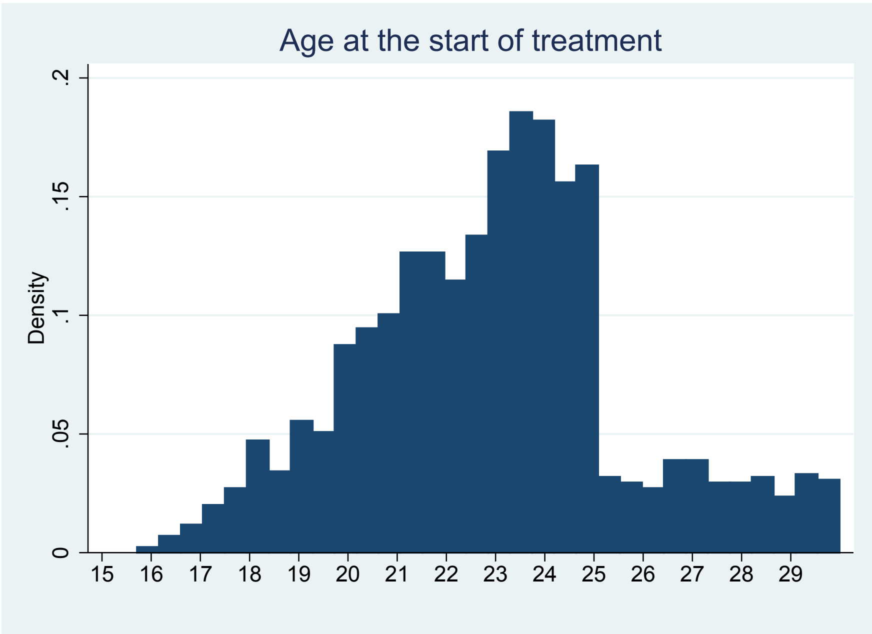


Figure 4: Age measured at participation starting date



January 2015 and April 2015. Hence, there are two waves of participants: the first one started the training programme between May and December 2014, and the second one started the training between May and December 2015.[17] Second, from Figure 4 we see that most of the participants are below age 25 when they start the training course, although there are also participants who start the training between age 25 and 29. This graph suggests that the YG's priority rule (such that priority is given to applicants aged 15–24 years) holds in this case and it is in line with the implementation of the YG in Latvia. However, analysis based on the data at hand show that the priority rule was very strict for the first wave (there are no participants above age 25) but not for the second wave.

Based on the aforementioned feature of the data we will separately analyse the two waves of participants adopting different methodological approaches depending on the characteristics of the treated units and relevant background information.

6 First wave of the programme

6.1 Descriptive statistics

In this section we present descriptive statistics on participants who enrolled in the vocational training programme in the first wave, that is between January and December 2014. We apply some sample restrictions. First, consistently with the Youth Guarantee age eligibility criteria, we retain individuals who are aged between 15 and 29 years. To be conservative, we drop 3 individuals who are younger than 15 on 31st December 2014 (who were certainly younger than 15 at the start of the participation period), and we drop 3,627 individuals who were older than 29 on 1st January 2014 (who were certainly older than 29 during the participation period). Second, we impose common support in the date of entry into unemployment between treated and control units, which results in selecting treated and control units entering unemployment between June 2013 and December 2014. The final definition of treated and control units is described in Box 1.

[17] However, individuals who enrolled in training courses that started before January 2015 were allowed to continue.

Box 1. *Definition of treatment group, control group and outcome*

Treated group: The treated group is composed of individuals who registered as unemployed in the SEA between June 2013 and December 2014, who started the vocational training before the end of December 2015, and who did not earn a wage at some fixed dates (January 2012, June 2012, December 2012, June 2013, and December 2013) in the period between registration in the SEA and 1st January 2014.

Control group: The control group consists of individuals who registered as unemployed in the SEA between June 2013 and December 2014 who have not participated in the vocational training programme and any other programme within the considered participation period (that is, from January 2014 until December 2015), and who did not earn a wage at some fixed dates (January 2012, June 2012, December 2012, June 2013, and December 2013) in the period between registration in the SEA and 1st January 2014.

Outcome: Employment status at some fixed dates (December 2015, June 2016).

The selected sample consists of 932 treated units and 19,849 control units. Table 1 and 2 report descriptive statistics on outcome variables and covariates by treatment status.

Table 1: Descriptive statistics of outcomes by treatment status

Treated group					
Variable	Obs	Mean	Std. Dev.	Min	Max
Income December 2015	932	198.9443	299.0035	0	1657.77
Income June 2016	932	212.6264	320.6676	0	2175.6
Employed December 2015	932	0.403434	0.49085	0	1
Employed June 2016	932	0.412017	0.492462	0	1
Control group					
Variable	Obs	Mean	Std. Dev.	Min	Max
Income December 2015	19,849	288.4349	473.7788	0	16471.7
Income June 2016	19,849	305.074	486.8926	0	8536.37
Employed December 2015	19,849	0.440476	0.496457	0	1
Income June 2016	19,849	0.444708	0.496946	0	1

Table 2: Descriptive statistics of covariates by treatment status

	Control group			**Treated group**				
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Difference in means	$t-$statistics
Average income before treatment	19,849	188.1152	282.3172	932	99.83367	172.4556	88.28***	(9.46)
Nr.years with positive income before 2014	19,849	1.870321	1.71363	932	1.310086	1.592607	0.560***	(9.78)
Female	19,849	0.51524	0.49978	932	0.55794	0.496898	-0.0427*	(-2.55)
Education								
Lower than primary	19,849	0.294876	0.455999	932	0.387339	0.487404	-0.0925***	(-6.03)
General secondary	19,849	0.293869	0.455544	932	0.368026	0.482527	-0.0742***	(-4.84)
Professional secondary	19,849	0.232153	0.422217	932	0.175966	0.380995	0.0562***	(3.99)
Higher education	19,849	0.178447	0.382899	932	0.06867	0.253027	0.110***	(8.66)
Nationality								
Belorussian	19,849	0.009925	0.099131	932	0.015022	0.121703	-0.00510	(-1.52)
Hebrew	19,849	0.000302	0.017384	932	0	0	0.000302	(0.53)
Latvian	19,849	0.635599	0.481274	932	0.674893	0.468666	-0.0393*	(-2.44)
Lithuanian	19,849	0.007305	0.08516	932	0.004292	0.065407	0.00301	(1.07)
Polish	19,849	0.011436	0.10633	932	0.013949	0.11734	-0.00251	(-0.70)
Rome	19,849	0.007053	0.083689	932	0.002146	0.046299	0.00491	(1.78)
Russian	19,849	0.177289	0.381922	932	0.166309	0.372558	0.0110	(0.86)
Ukrainian	19,849	0.00932	0.096094	932	0.001073	0.032756	0.00825**	(2.61)
Not specified	19,849	0.135221	0.341968	932	0.120172	0.325337	0.0150	(1.32)
Residence								
Capital city	19,849	0.233563	0.423108	932	0.103004	0.304128	0.131***	(9.31)
Local center	19,849	0.057736	0.233249	932	0.046137	0.209895	0.0116	(1.49)
National center	19,849	0.211144	0.408131	932	0	0	0.211***	(15.79)
National dev. center	19,849	0	0	932	0.233906	0.42354	-0.234***	(-77.84)
Regional center	19,849	0.122575	0.327958	932	0.215665	0.411504	-0.0931***	(-8.36)
Rural area	19,849	0.374981	0.48413	932	0.401288	0.490422	-0.0263	(-1.62)

*** $P-val$ <0.01, ** $P-val$ <0.05, * $P-val$ <0.1.

As shown by the results of the t-tests on the difference in the means, the treated and the control units are only partially balanced in terms of nationality. On the one hand, the proportion of Latvians is higher in the treated group than in the control group (the difference between the two means is negatively ad statistically different from zero at the 1% level). On the other hand, the proportion of other nationalities (Belorussian, Hebrew, Lithuanian, Polish, Rome, Russian and Ukrainian) is balanced between the two groups (being the differences between the two means not statistically different from zero). The average income and the number of years with positive income in the pre-treatment period (<2014) are higher in the control group compared to the treated one.[18] The two groups also differ with respect to gender, educational level and area of residence: first, the proportion of female unemployed is higher in the treated group. Second, treated unemployed are on average less educated: the proportion of unemployed with lower than primary or general secondary education is higher in the treated group than in the control one, while the proportion of unemployed with professional secondary or higher education is higher in the control group. Finally, the proportion of unemployed living in the capital city or national centers is higher in the control group compared to the treated one, while treated unemployed reside more often in the national development centers or regional centers than the controls. However, treated and controls reside in the same proportion in local centers and rural areas. All in all, these statistics suggest that individuals participating in vocational training may be the least employable in terms of observable characteristics (e.g. past work experience, education) and perhaps also of unobservable characteristics (motivation, job search effort, etc.).

Besides the variables included in table 2, we also account in our regression models for other characteristics such as the starting date of the unemployment spell.[19]

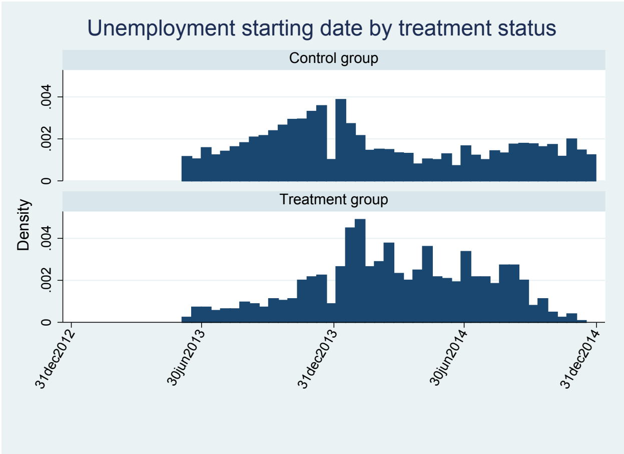
Figure 8 shows the starting date of the unemployment spell by treatment status. The distribution is quite different between the two groups in the selected sample. For the treatment group, the starting date of the unemployment spell is more concentrated between December 2013 and December 2014. By contrast, the starting date of the unemployment spell of control units is more spread over time. For this reason, we select the control group based on the distribution of the unemployment starting date for the treated group.

Figure 6 and 7 show the participation starting date and the age of participants at the start

[18] Although we imposed that treated and control units must have earned no income between SEA registration and January 2014, individuals may still differ in earned incomes before this period.

[19] As we already mentioned, we cannot use the unemployment duration in months to match the treated and control units because it is available only for the treated group.

Figure 5: Starting date of unemployment spell by treatment status



of the treatment. Differently from Figure 4, in the first wave of the programme the priority rule strictly applies to all participants as they are less than 25 years old at the start of the vocational training. Hence, for the first wave we can exploit the discontinuity at age 25 and adopt a regression discontinuity design approach. Since not all individuals below 25 participated in the programme, we adopt a fuzzy regression discontinuity design (FRDD) approach. In the next session we discuss the characteristics of this CIE method and the empirical specification in our particular setting.

Figure 6: Participation in vocational training starting date

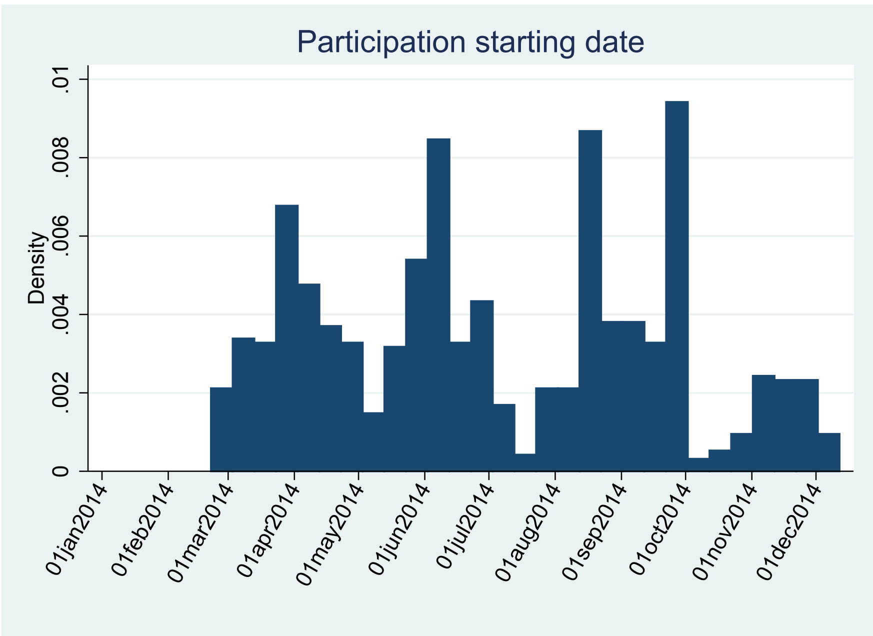
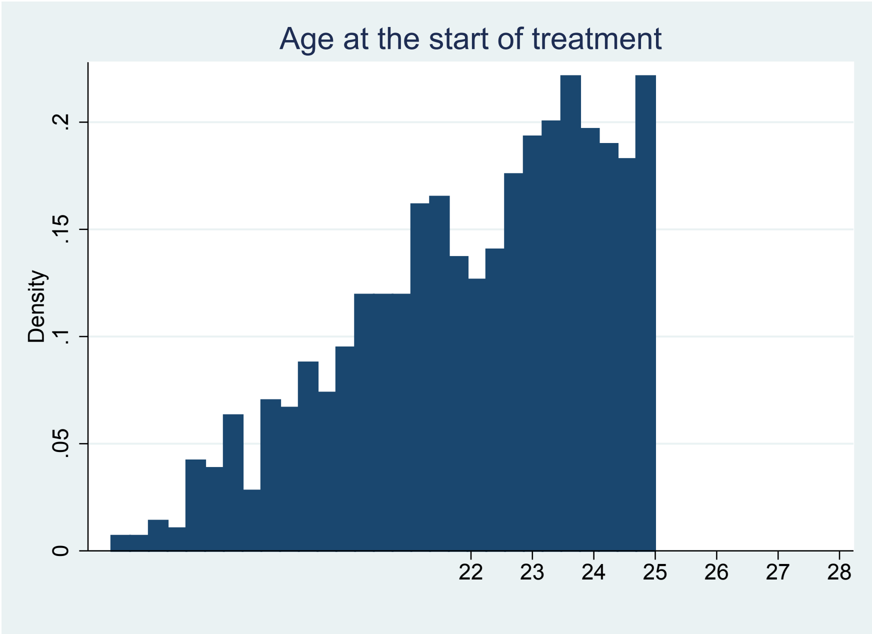


Figure 7: Age at participation in vocational training starting date



6.2 Counterfactual impact evaluation analysis: Fuzzy Regression Discontinuity Design

Our empirical strategy is based on a FRDD, which exploits the priority rule given to unemployed youth aged less than 25 years. This methodological approach can be used whenever there is a threshold in a given individual attribute that determines the assignment into treatment (Sharp RDD) or the likelihood of being treated (in the case of the FRDD). This setting can be seen as a sort of randomisation at the threshold value, since one can assume that around the threshold assignment to treatment is 'as good as random,' or said in other words that treated and control units have identical observable and unobservable characteristics. In our case, age is the running variable determining the probability of being treated, and 25 years represents the threshold, or cut-off point, in the eligibility criteria.

FRDD allows for imprecise allocation of treatment around the cut-off point, for settings where, although individuals with the running variable (i.e. age) above or below a given threshold are eligible to a programme, actual participation is voluntary. Hence, the participation in a given programme may depend both on the compliance with the eligibility conditions and on the individual motivation. Since this latter is likely to be associated with the outcomes of interest, identifying the causal effect of participating in the programme requires netting out it from the impact that motivation (or other unobservable attributes) may have on the choice to participate. This can be done trough the use of an instrumental variable approach and a two-stage least squares (2SLS) estimation strategy. In fact the discontinuity — in our case the discontinuity in the probability of participating given by the age — becomes an instrumental variable for treatment status instead of deterministically switching treatment on or off as in the case of a sharp RDD (Angrist and Pischke 2008).[20]

The 2SLS estimation strategy consists of two equations: the first one is set to estimate the effect of the instrumental variable on the probability of participating in the treatment, hence exploiting the exogenous variation in the instrument to predict the participation. In the second stage equation, the estimated probability in the first stage regression is used to compute the effect of the treatment on the outcome variable of interest.

In addition to explaining the choice of participating in the programme (relevance), the exogeneity requirement for the instrumental variable requires that it is not related with the unobserved individual characteristics such as motivation, which may affect the outcome as well. This is controlled for in the FRDD by including a flexible polynomial in age in the first stage of 2SLS, so as there are no specific reasons to expect motivation to have a discontinuity exactly at age 25, i.e. the age at which eligibility changes. Finally, the instrument is assumed to have only an indirect effect on the outcome variable, only through the probability of participation. This assumption is referred to as exclusion restriction, and it is ensured by the inclusion of a flexible polynomial in age in the second stage of 2SLS.

The 2SLS procedure allows to estimate the causal effect of the treatment on the compliers, defined as the individuals whose treatment status changes as we move the value of the running variable from just to the left of threshold to just to the right of threshold (Hahn, Todd, and van der Klaauw 2001), i.e. age 25 in our case.

In our baseline specification the treatment variable is a binary indicator that takes value one for participating in the vocational training programme under YG, and zero otherwise. The instrument is a binary indicator for being subject to the priority rule set up by the Government of Latvia at a given moment, i.e. being aged less than 25 years.

Our main equation is:

$$Y_i = \beta T_i + f(age_i - c) + X_i B + \epsilon_i \tag{1}$$

where Y_i is the employment status of individual i in December 2015, T_i is the treatment status,

[20]In particular, since the priority rule was strictly observed in the first wave of the programme, the RDD is sharp on the right of the cut-off point (age 25) and fuzzy on the left.

$f(age_i - c)$ are linear and/or quadratic polynomials in the running variable, which is the age of the individual measured at different points in time and normalised at age 25 (the cut-off point). Eq. (1) is estimated through 2SLS. The corresponding first stage equation is as follows:

$$T_i = \gamma_1 below\_age25 + g(age_i - c) + X_i\Gamma + \eta_i \quad (2)$$

Note that in Eq. (2) the relationship between T_i and age_i has the same slope on both sides of the cut-off (set at age 25). Function $g(\cdot)$ represents a linear or quadratic polynomial in the running variable.

It must be noted that unlike a textbook version of the FRDD, in which treatment assignment must be evaluated in a given point in time, VT courses were organized continuously across the whole 2014, so as in principle eligibility should be evaluated at each course's starting date and the FRDD applied to each single course. However, the number of observations is not large enough to present course-specific FRDD estimates, and we have to pool all courses and assess individual eligibility when the first VT courses started (i.e. January 2014).

Alternatively, we can estimate Eq. (2) using 2SLS instead of a FRDD and an instrument based on the number of months one is exposed to the priority rule, i.e. mo_{exp} (i.e. the number of months one is below age 25 during the entire participation period). As opposed to the dichotomous variable $below\_age25$, mo_{exp} is continuous. In addition, it is not a function of the calendar date when the age is measured. Nevertheless, even in this case we need to define a specific window over which to count the number of months spent under the priority rule, which is set between January and December 2014.

Table 3: Priority rule over time

	month in 2014												
birth month	1-Jan	31-Jan	28-Feb	31-Mar	30-Apr	31-May	30-Jun	31-Jul	31-Aug	30-Sep	31-Oct	30-Nov	31-Dec
1-Dec-88	25.08	25.17	25.24	25.33	25.41	25.49	25.58	25.66	25.75	25.83	25.91	26.00	26.08
1-Jan-89	25.00	25.08	25.16	25.24	25.33	25.41	25.49	25.58	25.66	25.74	25.83	25.91	26.00
1-Feb-89	**24.91**	25.00	25.07	25.16	25.24	25.33	25.41	25.49	25.58	25.66	25.74	25.83	25.91
1-Mar-89	**24.84**	**24.92**	25.00	25.08	25.16	25.25	25.33	25.42	25.50	25.58	25.67	25.75	25.83
1-Apr-89	**24.75**	**24.84**	**24.91**	25.00	25.08	25.16	25.25	25.33	25.42	25.50	25.58	25.66	25.75
1-May-89	**24.67**	**24.75**	**24.83**	**24.91**	25.00	25.08	25.16	25.25	25.33	25.42	25.50	25.58	25.67
1-Jun-89	**24.59**	**24.67**	**24.74**	**24.83**	**24.91**	25.00	25.08	25.16	25.25	25.33	25.42	25.50	25.58
1-Jul-89	**24.50**	**24.59**	**24.66**	**24.75**	**24.83**	**24.91**	25.00	25.08	25.17	25.25	25.33	25.42	25.50
1-Aug-89	**24.42**	**24.50**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**	25.00	25.08	25.16	25.25	25.33	25.42
1-Sep-89	**24.33**	**24.42**	**24.49**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**	25.00	25.08	25.16	25.25	25.33
1-Oct-89	**24.25**	**24.33**	**24.41**	**24.50**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**	25.00	25.08	25.16	25.25
1-Nov-89	**24.17**	**24.25**	**24.33**	**24.41**	**24.49**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**	25.00	25.08	25.16
1-Dec-89	**24.08**	**24.17**	**24.24**	**24.33**	**24.41**	**24.50**	**24.58**	**24.66**	**24.75**	**24.83**	**24.91**	25.00	25.08
1-Jan-90	**24.00**	**24.08**	**24.16**	**24.24**	**24.33**	**24.41**	**24.49**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**	25.00
1-Feb-90	**23.92**	**24.00**	**24.07**	**24.16**	**24.24**	**24.33**	**24.41**	**24.49**	**24.58**	**24.66**	**24.74**	**24.83**	**24.91**
1-Mar-90	**23.84**	**23.92**	**24.00**	**24.08**	**24.16**	**24.25**	**24.33**	**24.42**	**24.50**	**24.58**	**24.67**	**24.75**	**24.84**

Note. Each cell represents the individual age at each birth-month and calendar-month combination. Younger individuals have a longer exposition to the priority rule, and therefore a higher probability of participation in VT courses.

6.3 Results

6.3.1 Regression results

We now comment on a series of regression tables summarising the main results. Table 4 shows the estimates of the effect of participating in the vocational training on the employment status in June 2016, one year and a half after the completion of the first wave. We show the results when age is measured in January 2014, at the official start of the programme. In line with the graphical analysis, we show the results when using a quadratic specification of the polynomial in age.

In Column 1 we report the ordinary least square (OLS) results which suggest that participation in the treatment increases the probability of being employed in June 2016 by 0.7 pp. The coefficient is not statistically significant at conventional levels. However, these results may hinder positive or negative selection. In case of positive selection we would expect that the more motivated individuals enroll in the programme, while in case of negative selection we expect that the least employable (e.g. those with lower education) participate. To tackle this problem we use the Fuzzy RDD approach which basically leads to an IV estimate, given by the ratio of two estimands: the first stage estimand that captures the effect of the priority rule (the instrument) on the participation in the vocational training (endogenous variable) and the reduced form, i.e. the effect of the priority rule on employment status (outcome).

Results on the first stage are reported in Column 2. Being subjective to the priority rule (that is being younger than 25 at the start of the programme) increases the probability to participate in the programme by 3.7 pp. The $F-$statistic on the excluded instrument is 55 (not reported in the table), which rules out the the possibility that our analysis suffers from a weak instrument problem.[21] Females have a higher probability to participate in the programme compared to males and individuals with a secondary education have a higher probability to participate in the programme compared to individuals with a university degree or more.

Results on the reduced form are reported in Column 3 while the 2SLS estimates are reported in Column 4. Participating in the programme increases the probability of employment by 1.8 pp (Column 4). This effect is twice as large compared to the OLS estimates, and would point towards a negative selection story: that is, the least employable are more likely to enroll in this type of programmes. This is consistent with the treated group, for instance, to be characterised by lower

[21] According to the rule-of-thumb, in case of one instrument and one endogenous regressor, an F-statistics lower than 10 suggests that the instrument is weak.

educational attainment compared to the control group. However, like in the case of OLS, the effect is not statistically significant.

We have already said that eligibility to the priority rule at January 2014 may be only an imprecise proxy for the eligibility to priority across the whole 2014, where the precision depends on an individual's date of birth and the distribution of the courses' starting dates in 2014. In particular, using the FRDD and age measures at Janary 2014 we expect to have an underestimate of the effect of the priority rule on participation (first-stage effects), since eligibility will be loss by many individuals during the year (as 24-year olds turn 25). For this reason, we also report 2SLS estimates using a more precise measure of eligibility based on a continuous version of the instrument, that is the fraction of months an individual is under the priority rule in 2014. Results of this exercise are shown in Table 5. The estimates from the OLS regression in Column 1 are the same as in Table 4. Column 2 shows the first-stage results for the continuous instrument. Being under the priority rule for one additional month increases the probability to participate in the programme by 0.4 pp. The reduced form estimates are shown in Column 3. The effect of one additional month under the priority rule does not have a significant effect on the probability to be employed in June 2016. Column 4 shows the 2SLS estimates. The positive effect is consistent with the idea that participating in the treatment increases the probability of being employed in June 2016 by 30%. However, results are not statistically significant. This is in line with results reported in Table 4.

The difference observed for the results in the first stage are explained by the different nature of the instrument: by using the continuous variable we exploit additional information and assign higher weight to the individuals who are under the priority rule for for a higher number of months within the first wave (for instance, individuals who turn 25 years old after December 2014 are under the priority rule for 12 months, while individuals who turn 25 years old in February 2014 are exposed to the priority rule for only 2 months).

Table 4: The effect of vocational training on the probability of being employed in June 2016. Instrument: under age 25 on 1 January 2014.

Variables	(1) OLS	(2) 1st stage (2SLS)	(3) reduced form	(4) 2nd stage (2SLS)
Treated	0.00705			0.0187
	(0.0190)			(0.369)
$aget1\_notnorm < 25$		0.0373***	0.000700	
		(0.00504)	(0.0138)	
daily aget1 normalised at avg.sample	-0.00580***	-0.00377***	-0.00576**	-0.00569
	(0.00127)	(0.000877)	(0.00240)	(0.00366)
aget1 squared	-0.00127***	0.000524***	-0.00126***	-0.00127***
	(0.000366)	(0.000153)	(0.000418)	(0.000366)
Female	-0.0419***	0.00789***	-0.0419***	-0.0420***
	(0.00701)	(0.00256)	(0.00701)	(0.00762)
Foreign nationality	-0.0352***	0.00128	-0.0352***	-0.0352***
	(0.00748)	(0.00273)	(0.00748)	(0.00748)
Lower than primary	-0.194***	0.00601	-0.194***	-0.194***
	(0.0110)	(0.00401)	(0.0110)	(0.0111)
General secondary	-0.0694***	0.0101***	-0.0693***	-0.0695***
	(0.0105)	(0.00383)	(0.0105)	(0.0111)
Professional secondary	-0.0747***	-0.00530	-0.0747***	-0.0746***
	(0.0111)	(0.00407)	(0.0111)	(0.0113)
Not specified	-0.110	-0.0320	-0.111	-0.110
	(0.134)	(0.0492)	(0.134)	(0.135)
Local center	-0.00406	0.0124**	-0.00397	-0.00421
	(0.0159)	(0.00580)	(0.0159)	(0.0165)
National center	-0.0227**	-0.0226***	-0.0229**	-0.0224*
	(0.0104)	(0.00380)	(0.0104)	(0.0133)
National dev. center	-0.0324	0.936***	-0.0259	-0.0434
	(0.0382)	(0.0124)	(0.0338)	(0.348)
Regional center	-0.0289**	0.0477***	-0.0286**	-0.0295
	(0.0122)	(0.00446)	(0.0122)	(0.0215)
Rural area	-0.0137	0.0216***	-0.0135	-0.0139
	(0.00956)	(0.00349)	(0.00955)	(0.0124)
Constant	0.561***	-0.0437***	0.560***	0.561***
	(0.0211)	(0.00858)	(0.0235)	(0.0220)
Observations	20,781	20,781	20,781	20,781
R-squared	0.053	0.271	0.053	0.053

*** $P-val$ <0.01, ** $P-val$ <0.05, * $P-val$ <0.1.

7 Second wave of the programme

7.1 Descriptive statistics

In this section we will present descriptive statistics for the young NEETs who enrolled in the training courses in the second wave, between May and December 2015. Firstly, we restrict the sample used for the analysis according to the age criteria. In the second wave of the programme, the priority rule of Youth Guarantee age eligibility criteria is relaxed to allow the participation of young NEETs aged up to 30 years. Therefore for this second analysis the treated group consists of individuals who are aged between 15 and 30 years at the start of the participation period, i.e. in May 2015. Accordingly, we exclude from the control group 762 individuals who are older than 30 on 1 May 2015. We then impose common support in the date of entry into unemployment between treated and controls, and this results in selecting treated units and control units who entered unemployment between March 2010 and November 2011 We drop therefore those control units (1873) who enter unemployment before March 2010 or after December 2011. The selected sample consists of 889 treated units and 35.871 control units.

Table 6 and Table 7 report descriptive statistics on outcome variables and covariates by treatment status.

Table 6 clearly shows a drop of employment for treated at December 2015. However, this is a mechanical effect partly due to the fact that some of the treated individuals were still enrolled in the VT courses as of December 2015. For this reason, in what follows, we just focus on the probability of employment in June 2016.

Table 7 reports the descriptive statistics for the relevant pre-treatment variables separately for the two groups of NEETs participating in the second wave and the selected control units. As shown also by the results of the t-tests on the mean, the two groups are balanced in terms of nationality (Belorussian, Latvian, Lithuanian, Polish, Rome, Russian and Ukrainian) and residence area (local center and rural area), while they differ with respect to pre-treatment income, gender and education level. In particular, the frequency of NEETs with a positive income in the pre-treatment period (before 2014) is higher in the control group. The same can be observed for the average income in the pre-treatment period. This means that although unemployed at the start of the training, NEETs may have experienced some employment spells, associated with positive labour income, after registering in the employment office. This is true especially for the non-participants. The proportion of female NEETs is higher in the treated group rather than in the control one. Treated

Table 5: The effect of vocational training on the probability of being employed in June 2016. Continuous instrument: number of months below age 25 (in 2014).

Variables	(1)	(2)	(3)	(4)
	OLS	1st stage (2SLS)	reduced form	2nd stage (2SLS)
1=participated in 1/1/2014-31/12/2015	0.00705			0.286
	(0.0190)			(0.277)
Nr.months exposed to treatment		0.00460***	0.00132	
		(0.000463)	(0.00127)	
daily aget12 normalised at avg.sample	-0.00580***	-0.00114	-0.00353	-0.00321
	(0.00127)	(0.000942)	(0.00258)	(0.00287)
aget12 squared	-0.00127***	0.000531***	-0.00111***	-0.00127***
	(0.000366)	(0.000145)	(0.000397)	(0.000368)
Female	-0.0419***	0.00768***	-0.0420***	-0.0442***
	(0.00701)	(0.00256)	(0.00701)	(0.00739)
Foreign nationality	-0.0352***	0.00152	-0.0350***	-0.0354***
	(0.00748)	(0.00273)	(0.00748)	(0.00751)
Lower than primary	-0.194***	0.00672*	-0.193***	-0.195***
	(0.0110)	(0.00401)	(0.0110)	(0.0111)
General secondary	-0.0694***	0.00976**	-0.0693***	-0.0720***
	(0.0105)	(0.00383)	(0.0105)	(0.0109)
Professional secondary	-0.0747***	-0.00625	-0.0751***	-0.0733***
	(0.0111)	(0.00407)	(0.0111)	(0.0113)
Not specified	-0.110	-0.0326	-0.110	-0.101
	(0.134)	(0.0491)	(0.134)	(0.135)
Local center	-0.00406	0.0126**	-0.00389	-0.00750
	(0.0159)	(0.00580)	(0.0159)	(0.0163)
National center	-0.0227**	-0.0221***	-0.0227**	-0.0164
	(0.0104)	(0.00379)	(0.0104)	(0.0122)
National dev. center	-0.0324	0.933***	-0.0274	-0.294
	(0.0382)	(0.0123)	(0.0338)	(0.262)
Regional center	-0.0289**	0.0477***	-0.0286**	-0.0423**
	(0.0122)	(0.00446)	(0.0122)	(0.0181)
Rural area	-0.0137	0.0218***	-0.0135	-0.0197*
	(0.00956)	(0.00349)	(0.00955)	(0.0113)
Constant	0.561***	-0.0516***	0.551***	0.566***
	(0.0211)	(0.00850)	(0.0233)	(0.0217)
Observations	20,781	20,781	20,781	20,781
R-squared	0.053	0.273	0.053	0.043

*** $P-val$ <0.01, ** $P-val$ <0.05, * $P-val$ <0.1.

Table 6: Descriptive statistics of outcomes

Treated group						
	Variable	Obs	Mean	Std. Dev.	Min	Max
Income December 2015	889	113.2841	230.2781	0	1295.94	
Income June 2016	889	209.6556	307.0931	0	1820	
Employed December 2015	889	0.259843	0.438795	0	1	
Employed June 2016	889	0.424072	0.49448	0	1	
Control group						
	Variable	Obs	Mean	Std. Dev.	Min	Max
Income December 2015	35,871	249.4677	439.7438	0	16471.7	
Income June 2016	35,871	285.3207	468.6489	0	10175.85	
Employed December 2015	35,871	0.397062	0.489296	0	1	
Employed June 2016	35,871	0.431797	0.495334	0	1	

NEETs have on average lower level on education, as primary or generalist secondary, than the controls, where NEETs with professional secondary or higher level of education are predominant. Like for the first wave, also for this second wave of participants there is therefore evidence of negative self-selection, i.e. less employable individuals were more likely to participate in the training. As for the residence area, treated and controls differ with respect to their provenience from the capital city or national centres (this proportion is higher in the treated group) vs. national development center or regional center (this proportion is higher in the control group).

Figure 8 shows the participation starting date of the vocational training programme and Figure 9 shows the age of participants at the start of the treatment. Consistent with the relaxation of the priority rule in the second wave of implementation, also individuals above age 25 participated in vocational training programmes in 2015.

7.2 Counterfactual Impact evaluation analysis: Matching methods

As mentioned in previous section, starting from 2014 the age priority rule is relaxed: for VT courses starting in May 2015 also NEETs aged above 25 years were allowed to participate. This implies that, differently from the first wave of the programme, there is no clear discontinuity in the eligibility criteria based on age that could be exploited in the identification strategy. Therefore, for the analysis of the second wave of participants, we cannot apply the fuzzy Regression Discontinuity Design method and we rely on *matching* estimators. The availability of information on demographic and labour market characteristics allows to evaluate the similarity of treated and control units with

Table 7: Descriptive statistics of covariates

Controls			Treated				
Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Difference in means	$t-$statistics
35,871	172.3261	275.7235	889	115.5981	187.9707	56.73***	(6.10)
35,871	1.725238	1.707001	889	1.428571	1.641836	0.297***	(5.12)
35,871	0.5274456	0.4992531	889	0.6029246	0.4895673	-0.0755***	(-4.45)
35,871	0.298319	0.4575266	889	0.3397075	0.4738765	-0.0414**	(-2.66)
35,871	0.2888963	0.453256	889	0.3498313	0.477185	-0.0609***	(-3.95)
35,871	0.230409	0.4211005	889	0.208099	0.4061765	0.0223	(1.56)
35,871	0.1814279	0.3853777	889	0.1023622	0.3032947	0.0791***	(6.07)
35,871	0.0009478	0.0307729	889	0	0	0.000948	(0.92)
35,871	0.0101196	0.1000873	889	0.007874	0.0884354	0.00225	(0.66)
35,871	0.0005297	0.0230089	889	0	0	0.000530	(0.69)
35,871	0.6320705	0.4822488	889	0.6602925	0.4738765	-0.0282	(-1.72)
35,871	0.0078615	0.0883171	889	0.0044994	0.0669644	0.00336	(1.13)
35,871	0.0113183	0.1057854	889	0.0101237	0.1001625	0.00119	(0.33)
35,871	0.0074155	0.0857944	889	0.0056243	0.0748262	0.00179	(0.62)
35,871	0.177748	0.3823058	889	0.1653543	0.3717092	0.0124	(0.96)
35,871	0.0095899	0.0974588	889	0.0067492	0.0819216	0.00284	(0.86)
35,871	0.1373533	0.3442247	889	0.136108	0.3430962	0.00125	(0.11)
35,871	0.2378244	0.425757	889	0.1417323	0.3489717	0.0961***	(6.67)
35,871	0.0575395	0.2328739	889	0.056243	0.23052	0.00130	(0.16)
35,871	0.2065178	0.404812	889	0	0	0.207***	(15.21)
35,871	0	0	889	0.2193476	0.4140375	-0.219***	(-100.39)
35,871	0.1192328	0.3240668	889	0.1777278	0.3824986	-0.0585***	(-5.29)
35,871	0.3788855	0.4851163	889	0.4049494	0.4911586	-0.0261	(-1.58)

*** $P-val$ <0.01, ** $P-val$ <0.05, * $P-val$ <0.1.

Figure 8: Starting date of unemployment spell by treatment status

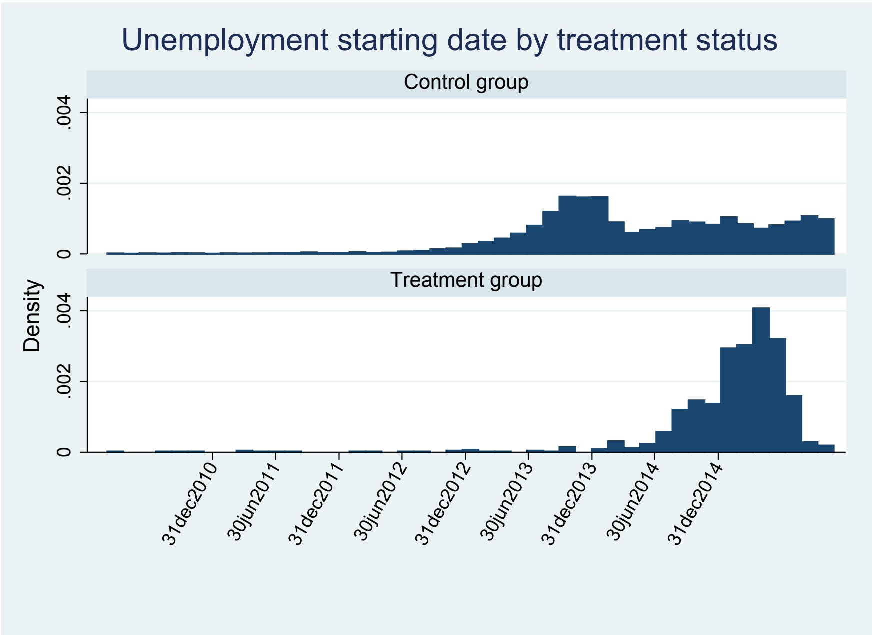
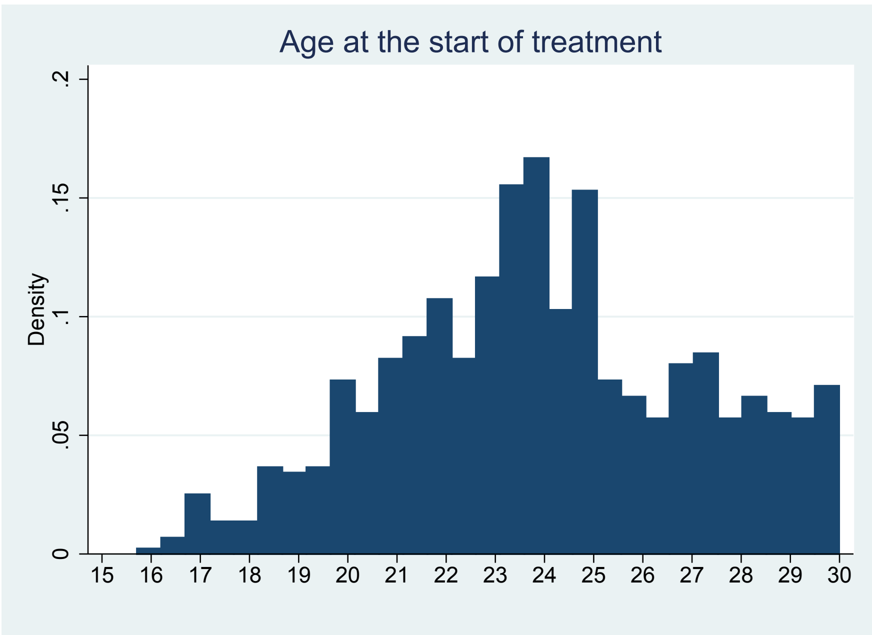


Figure 9: Age at participation in vocational training starting date



respect to relevant pre-intervention features and hence to identify those who are more comparable for the assessment of the treatment effect. In detail, we apply two *matching* procedures, i.e. the Propensity Score Matching (PSM) and the Coarsened Exact Matching (CEM). Both methods are based on the following assumptions.

- *Conditional Independence Assumption (CIA)*
  The potential outcomes are assumed to be independent of the assignment of treatment conditional on the observable characteristics **X**. Namely, controlling for all observable characteristics, the participation decision is uncorrelated with the potential outcomes. The extent to which this assumption is reasonable depends on the richness of the data in **X**.

- *Stable-Unit-Treatment-Value Assumption (SUTVA)*
  the effect of the treatment on the outcome of one unit does not affect the outcome of other units (no interference). This assumption rules out spill over effects or general equilibrium effects of the treatment that change the behaviour of the control units. For instance, it prevents crowding-out (displacement) effects in local labour markets: namely, if treated units are more likely to find a job due to the intervention, this should not deteriorate the likelihood to find a job of control units. In addition, the intervention should not affect the control units through changes in the general equilibrium of the wage in the labour market. Although the SUTVA may be a quite strong assumption in the context of very large-scale labour market policies, in our case it is a reasonable assumption since the intervention's target group is quite small (treated units are 889 whereas control units are 35.871) and hence unlikely to drive general equilibrium or displacement effects.

- *Common Support Assumption:* $0 < P(D = 1|X = x) < 1$
  This assumption implies that for any given value of the observable characteristics **X**, the probability of being treated should vary between 0 and 1 and hence not be certain. Therefore, for each value of the confounding variables **X** an individual could be potentially observed as treated or not. This assumption ensures that for each treated individual (with given realisations of variables **X**), it is possible to find a sufficiently similar individual in the control group, i.e. a control unit that is similar (or even identical) to the treated one in terms of the variables X.

The purpose of the matching procedures is to estimate the treatment effect by comparing treated units with control units who are similar in terms of observable characteristics, that affect both the treatment participation and the outcome variables. The CIA, if valid, will also ensure that after matching treated and control units are also similar with respect to characteristics which are unobservable to the analyst. The outcome value of each treated unit i should be compared with the outcome value of a control unit j that is identical to i in terms of a number of characteristics contained in $\mathbf{X}$. Finding an exact match for each individual i becomes more and more difficult as the number of $\mathbf{X}$ characteristics increases. This is defined as a curse of dimensionality problem. The matching procedures used here differ in the way they solve it.

The PSM is based on the estimation of an indicator which summarises all information contained in the $\mathbf{X}$. This is the probability of being assigned to the treatment conditional on the observed characteristics, i.e. the propensity score. Matching on the propensity score has in fact been proved to be equivalent to matching units on the characteristics $\mathbf{X}$ (see, for instance, Bryson, Dorsett and Purdon, 2002). Once the propensity score is estimated for each individual, treated and control units with similar values of the propensity score are compared. If the propensity score is correctly estimated, individuals with similar values in the propensity score are also similar in terms of unobservable confounding factors. This also means that one is comparing treated and control units which are similar in terms of potential counterfactual outcomes.

Analogously with the PSM, the CEM is aimed at achieving balance between the treated and control units in terms of characteristics $\mathbf{X}$. It differs from the PSM, because the balance between the treated and the control groups is chosen ex-ante, hence avoiding the need to check it after the matching and eventually to repeat the procedure. The check on the validity of the common support assumption is not needed since the CEM automatically restricts the matched data to areas of common empirical support (King and Zeng, 2006).

The key property of CEM is that it belongs to the class of matching methods known as Monotonic Imbalance Bounding (MIB). These MIB methods bound the maximum imbalance in some feature of the empirical distributions through an *ex-ante* choice by the user (Iacus, King and Porro, 2008). The idea of CEM is to temporarily coarsen each variable into substantively meaningful groups, exact match on these coarsened data and then only retain the original (uncoarsened) values of the matched data. In this way the CEM refines the standard exact matching procedure, by creating strata for the $\mathbf{X}$ variables, and avoiding the limitation of few matches due to curse-of-dimensionality issues.

These two procedures, PSM and CEM, can also be combined, by applying the PSM on the common support of units identified by the CEM. The application of the two methods together will be shown in the last part of this session.

The aim of matching procedures is to estimate the Average Treatment Effect on the Treated (ATT), i.e. the impact of the intervention for the group of participants (Angrist and Pischke, 2008). The ATT is calculated as the difference between the average outcome of the treated group given the treatment and the average outcome of the treated group in the counterfactual situation where the treatment did not take place:

$$ATT = E(Y_i^1|D_i = 1) - E(Y_i^0|D_i = 1) \tag{3}$$

where D_i is an indicator equal to one if the treatment takes place and zero otherwise, Y_i^1 is the individual potential outcome given treatment and Y_i^0 is the individual potential outcome in the absence of the treatment. Note that for the ATT both potential outcomes refer to the treated group since they are conditioned upon $D_i = 1$.

While the PSM provides estimates for the ATT, the CEM allows to estimate the ATT for the well-matched subsample of treated units, that is the Sample Average Treatment Effect on the Treated (SATT). This is in fact calculated discarding all observations within any stratum that do not have at least one observation for each unique value of the treatment variable.

7.3 Results

In the following tables the results of the two matching procedures are shown.

Table 8: PSM results on the probability of employment at June 2016.

		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
	No replacement, RM[†]									
1	att_norepl	0.01	0.01	0.00	0.00	0.01	0.00	0.03	0.02	0.03
	ttest_norepl	0.34	0.34	0.20	0.20	0.61	.	1.76	.	.
2	att_norepl_c001 ($c = 0.001$)[§]	0.01	0.01	0.00	0.00	0.01	0.00	0.03	0.02	0.03
	ttest_norepl_c001	0.25	0.25	0.14	0.14	0.44	0.15	1.41	1.01	1.41
3	att_norepl_c01 ($c = 0.01$)	0.01	0.01	0.00	0.00	0.01	0.00	0.03	0.02	0.03
	ttest_norepl_c01	0.26	0.27	0.16	0.16	0.51	0.17	1.54	1.08	1.53
	Replacement, RM[†]									
4	att_repl	0.07	0.07	0.42	0.42	-0.03	-0.07	-0.12	0.02	0.05
	ttest_repl	0.65	0.65	2.51	2.51	-0.14	-0.48	-1.32	0.36	0.95
5	att_repl_c001 ($c = 0.001$)	0.07	0.07	0.42	0.42	-0.03	-0.07	-0.12	0.02	0.05
	ttest_repl_c001	0.65	0.65	2.51	2.51	-0.14	-0.48	-1.32	0.36	0.92
6	att_repl_c01 ($c = 0.01$)	0.07	0.07	0.42	0.42	-0.03	-0.07	-0.12	0.02	0.05
	ttest_repl_c01	0.65	0.65	2.51	2.51	-0.14	-0.48	-1.32	0.36	0.95
	Replacement, NNM[‡]									
7	att_repl_n5 ($n = 5$)[§§]	-0.03	-0.03	0.20	0.20	-0.03	-0.07	0.03	0.02	0.02
	ttest_repl_n5	-0.56	-0.56	1.63	1.63	-0.27	-0.98	0.59	0.51	0.58
8	att_repl_n10 ($n = 10$)	-0.02	-0.02	0.06	0.06	0.01	-0.08	0.01	0.02	0.02
	ttest_repl_n10	-0.49	-0.51	0.73	0.73	0.21	-1.48	0.32	0.95	0.91
9	att_repl_n20 ($n = 20$)	0.00	0.00	-0.05	-0.05	-0.07	-0.01	0.01	0.01	0.02
	ttest_repl_n20	-0.05	-0.02	-0.76	-0.76	-1.44	-0.21	0.51	0.39	0.80
	Matching variables									
	age linear	X	X							
	age quadratic		X							
	age dummy			X						
	age classes				X	X	X	X	X	X
	unemployment years	X	X	X	X					
	unemployment starting date					X	X	X	X	X
	gender						X	X	X	X
	area							X	X	X
	education								X	X
	nationality									X

† *RM* is Radius Matching. § c is caliper used to define the radius in the *RM*. ‡ *NNM* is Nearest-Neighbor Matching. §§ n is the number of observations used to define the Neighborhood in the *NNM*.

Table 8 contains the ATT values estimated by PSM, using different sets of control variables for the estimation of the Propensity Score. These are reported in the different columns. The rows contain the estimates (and the corresponding p-values) obtained with different matching algorithms. Rows 1-3 report the results from matching algorithms where replacement is allowed. This means that each control unit may be used as a match for treated units more than once. In rows 2 and 3 different values of caliper is used to perform radius matching within the specified radius. Starting from row fourth no replacement of control units is used. Algorithms based on different number of control units to be used in each match are shown in rows 7-9. Increasing the number of control units to be assigned in each matching pair (5, 10 and 20) allows to increase the precision of the estimate although it may increase the bias (since each treated unit is compared with more control units, that may be progressively have less similar values of the Propensity Score).

Using each control unit as a match only once through no replacement decreases the comparability of the groups. The graphical inspection of the common support shows that this is undermined. Therefore only estimates obtained with replacement can be taken into consideration. To start only two variables are taken into account to perform the matching: age and a proxy for duration in unemployment[22] For age, linear and quadratic values, dichotomic variable (below 25) and classes are used. As regards the unemployment duration, two variables are considered: i) the number of years passed since the registration at the SEA; ii) (a dummy variable indicating) the exact year when unemployed registered at the SEA. Matching results in terms of balance and t-tests improve when age classes and the year of unemployment starting date are considered. Therefore, these variables are used as baseline for the estimation of the Propensity Score. Subsequently a step-wise approach is used for the estimation, adding progressively the following (dummy) variables: gender, area of residence, level of education and nationality.

As shown in the different columns, results are mixed. ATT estimates obtained in most cases are positive. However, when only few variables are used in the estimation of the Propensity Score, i.e. age classes, unemployment starting date and gender (as shown in columns 5 and 6), the estimated values for the ATT are negative. This is consistent with PSM-ATT not performing well in the absence of a large enough set of matching variables. Indeed, a poor match is likely to be reflected in negative estimates of the treatment effects, which mainly capture the negative selection in program participation that we described above. In all cases, results are not statistically

[22]Note, as explained in the text, due to data limitation we cannot exactly measure the duration in unemployment. With "duration in unemployment" we mean time since registration as unemployment at the SEA.

significant. In order to check the robustness of our estimates, these are compared with the results obtained with the CEM procedure.

Table 9: CEM results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	CEM	CEM	CEM	CEM	CEM	CEM
treated	0.0206	0.0205	0.0205	0.0180	0.0182	0.0177	0.0260
	(0.0188)	(0.0187)	(0.0187)	(0.0188)	(0.0189)	(0.0195)	(0.0197)
Observations	36,760	28,870	28,870	25,710	16,316	11,166	8,700
R-squared	0.032	0.038	0.036	0.036	0.036	0.001	0.000
Matching variables							
age continuous	X	X	X	X	X	X	X
unemployment starting date	X	X	X	X	X	X	X
gender	X		X	X	X	X	X
area	X			X	X	X	X
education	X				X	X	X
nationality	X						X

The table reports CEM estimates. The dependent variable is a dummy for being employed in June 2016.

Table 9 reports the SATT estimates resulting from the CEM. In this case too, for robustness, estimates obtained using different matching variables are reported. As for the PSM procedure, age and unemployment duration are used as baseline to perform the coarsened matching. Then, the following variables are added progressively according to a step-wise approach: gender, area of residence, level of education and nationality. Values presented in the different columns show that in all cases the treatment effect is positive but not statistically significant. Results do not change if different age variables are used: age in classes or a dummy variable indicating whether the unemployed is aged less than 25. In this case, the age classes are defined automatically by the CEM algorithm.

Although the CEM estimates refer only to a subsample of treated units, i.e. the ones well-matched with the control units, the obtained results are in line with the PSM ones, referring to the whole treated group. Both matching procedures indicate that for the second wave of the programme the treatment effects are positive but not statistically significant. This is in line with the results illustrated in the previous section for the first wave of the programme and the findings from the literature that investigates the effects of similar programmes in the Nordic countries and the UK.

8 Conclusion

In this report we analyse the impact of a vocational training programme that is part of the Youth Guarantee package implemented in Latvia starting from 2014. Youth Guarantee is targeted to the NEETs aged 15–29 and aims to increase the employment rate of this group in the country.

This evaluation relies on an administrative database obtained from merging information from the State Employment Agency and the State Revenue Service (i.e. the State Tax Authority). Since the training programme is ongoing, the analysis is limited to the first two years of implementation of the programme, i.e. the period 2014-2015. The outcome variables, measured as the probability to be employed, are measured in June 2016. Within this two-year period, two waves of training courses have been offered, the first in 2014 and a second one in 2015. Since the two waves differ in terms of the eligibility conditions for participation, we carry out the analysis separately using two different empirical strategy. In the first wave, the unemployed youth aged less than 25 had the priority to participate in the training courses as opposed to their older peers aged 25–29. Therefore, within the target population, the former group had a higher probability to participate in the programme based only on the age at the start of the programme. We exploit this priority rule and compare the outcomes of interest between those who are just below age 25 and those who are just above it, since the former has a higher probability to participate in the programme, everything else being equal. This naturally leads to applying a strategy called Fuzzy Regression Discontinuity Design, which exploits an exogenous variation in the participation probability due to the age cut-off set at 25. However, the strength of the empirical strategy in terms of internal validity (i.e. the ability to estimate causal effects and not simple correlations) comes at the cost of reduced external validity, since the impact refers to the sub-group of unemployed youth close to the age 25 cut-off. The results point towards positive albeit statistically insignificant effects of the programme on employment measured as of June 2016. The priority rule set at age 25 increases the probability of participation in the programme for people below age 25. However, results from the reduced form equation (that is the effect of the priority rule on employment rate) do not show any significant difference in the probability of being employed between treated and control groups. As a consequence we do not find any significant effect in terms of employability due to the training course.

Differently from the first wave, in the second wave the age-eligibility condition was relaxed, and all individuals in the target group (15–29) were given the chance to participate in the programme. As a consequence, we cannot use the Fuzzy RDD as before, so we rely on matching methods and compare the employment rate of participants with the employment rate of similar non-participants based on a set of observable characteristics. This second exercise allows us to evaluate the impact of the measure on the whole target population (unemployed individuals aged 15–29), albeit at the cost of a strong identifying assumption (i.e. less internal validity) that all the relevant characteristics are controlled for (Conditional Independence Assumption). Also in this case we find positive but not statistically significant evidence that participation in the training course increased the probability of employment.

Overall we find no statistically significant positive effect of vocational training courses for the two waves of participants. Because the underlying assumptions in the two exercises are different the results are not comparable and should be read separately. There may be different explanations behind these results. First, the ALMPs that aim to increase the employability of the unemployed may not reach their ultimate objective if there is no labour demand that may high enough to absorb the targeted unemployed. This suggests that in order to increase programme effectiveness training measures may be ideally combined with measures fostering job creation, such as tax rebates for firms hiring unemployed individuals. In this specific case the programme we are evaluating was not accompanied by such measures to increase labour demand. Second, the programme started during an upturn period and hence both treated and control individuals may have found a job independently of this specific programme. According to a recent report (Escudero and López Mourelo 2015), it is important to implement such measures on time so as to increase their effectiveness. This hypothesis is supported by the increase in the employment rate in the 15-24 age group observed in Latvia since 2011 (see YG factsheet Latvia, 2016). Third, a higher increase in the number of vacancies does not automatically lead to a decrease of the unemployment rate if unemployed do not possess the skills required. The mismatch between job characteristics and skills should also be considered as a possible explanation for our findings.

9 References

Angrist, J. D. and Jrn-Steffen Pischke (2009). "Mostly Harmless Econometrics: An Empiricist's Companion," Economics Books, Princeton University Press, edition 1.

Blundell R., Costas M., Costa Dias M., and J. Van Reenen (2004), "Evaluating the employment impact of a mandatory job search program", *Journal of the European Economic Association*, 2(4): 569–606.

Bryson, A., Dorsett R. and S. Purdon (2002). "The use of propensity score matching in the evaluation of active labour market policies," LSE Research Online Documents on Economics 4993, London School of Economics and Political Science, LSE Library.

Cabasés Piqué A., Pardell Veá M., and T. Strecker (2016), "The EU youth guarantee ? a critical analysis of its implementation in Spain", *Journal of Youth Studies*, Vol. 19 (5), 684–704.

Carling K, and L. Larsson (2005), "Does early intervention help the unemployed youth?", *Labour Economics*, 12: 301–319.

European Commission, Country Report Latvia 2016.

Escudero, V., and E. L. Mourelo (2015), "The Youth Guarantee programme in Europe: Features, implementation and challenges", *International Labour Office*, Research Department Working Paper No. 4.

Hämäläinen, K., Hämäläinen U. and J. Tuomala (2014), "The Labour Market Impacts of a Youth Guarantee: Lessons for Europe?", *Government Institute for Economic Research Working Papers* No. 60.

Iacus, S., King, G. and G. Porro (2008), "Matching for Causal Inference Without Balance Checking," UNIMI - Research Papers in Economics, Business, and Statistics unimi-1073, Universit degli Studi di Milano.

Pastore, F. (2015), ?The European Youth Guarantee: labor market context, conditions and opportunities in Italy?, *IZA Journal of European Labor Studies*, 4: 11.

YG factsheet Latvia (2016), "Factsheet: Youth Guarantee & Youth Employment Initiative in Latvia", European Commission. link