What training for the unemployed?

An impact evaluation for targeting training courses

Silvia Duranti*
Valentina Patacchini*
Maria Luisa Maitino*
Carla Rampichini**
Nicola Sciclone*

** Irpet   ** University of Florence*

Objective and contribution of the paper

In the last years, the trend towards activation has been one of the major issues in welfare and labour market reforms in Europe. Italy has lacked for a long time a strong net of activation policies for the unemployed, but the latest reforms have placed great emphasis on the need to invest in Public Employment Services to make labour market more inclusive. At the same time the European Union, through the European Social Fund, has made available the necessary financial resources to promote active policies, especially those related to training, considering them crucial for the development of human resources in a knowledge-based economy.

Italian Regions play a major role in the planning and managing of training activities for the unemployed and this role became strategic with the upsurge of the economic crisis and the increasing needs of reskilling jobseeker.

The increasing relevance placed on training and activation policies calls for a development of a culture of able to identify the effects of the interventions and provide guidance on how to target them relying on appropriate statistical methods. In particular, in the case of training courses for the unemployed, an impact evaluation by type of users can provide guidance for the development of users segmentation systems in order to steer every unemployed to the most effective training program for his (re-) employment

This paper proposes an impact evaluation of training courses for the unemployed financed in Tuscany (Italy) by the European Social Fund. In particular, the evaluation is carried out for four different types of users, identified on the basis of a profiling system which summarizes each jobseeker's distance from the labour market. The aim is to identify, for every profiling group, if the attendance of different types of training courses improves the chances of re-employment.

Methodology

We make use of counterfactual impact evaluation to identify the impact of training activities on unemployed people (Imbens and Rubin, 2015).

Let Y(1) be the potential outcome that would result from attending a training course and Y(0) the potential outcome if deciding not to participate in the course. For a given subject, the casual effect of participation in the training programme would be the difference between Y(1) and Y(0). The average impact of the programme for participants, is given by the average treatment effect on the treated (ATT):

$$ATT = E[Y(1) - Y(0) \mid W = 1]$$

where W is the treatment status. However, the outcome Y(0) is not observable for a participant. To overcome this problem, in an impact evaluation the comparison group plays the role of providing a good approximation to the *counterfactual* outcome of treated units. Indeed, the comparison group should include units as similar as possible to those exposed to the treatment with respect to all the characteristics relevant for the choice of participating in the programme.

In observational studies, the identification of the control group is the most critical part of counterfactual analysis, unless in experimental contexts, in which the allocation to the treatment is random. In a non-experimental context, information on the characteristics of participants are normally collected for monitoring purposes, but information on the control group are difficult to find.

In our analysis the "treated" are the unemployed who have started a course in 2011, 2012, 2013 or 2014[1]. Trainees are grouped by two-month periods on the basis of the starting date of the course, in order to match them with unemployed in the same period not participating in any training course, who represent the control group.

The administrative database on jobseekers signed up by Public Employment Services (PES) can be used to select the control group. In this database, subjects never enrolled in a training course are probably those with better employment chances, and thus positive labour market outcomes at t+1, t+2, t+3, …, t+n. To avoid the potential endogeneity problem arising in choosing such subjects, we followed the approach of Sianesi (2004), i.e. the controls for a subject participating in a course starting in period *t* is constituted by jobseekers not starting a training course at the same period *t*. Therefore, the comparison group for participants in a course starting at time *t* might include people taking part in any course starting at a subsequent period[2]. The proposed approach allows to reproduce a sort of experimental context, where, in each period, an unemployed subject can decide whether to sign up to a training course or to search for work autonomously, conscious that it will be possible to start a course in the next period. For all two-month periods identified on the basis of the courses' starting dates, the control group is thus represented by all unemployed people being registered by Tuscan PES in the middle of the period. Given the huge number of controls, a stratified sample of 283,896 controls was drawn before matching. The strata are proportional to the distribution of the following characteristics of the treated: sex, nationality, area, education level, period and sector of activity in the last two years (before treatment).

At the end, the database used for evaluation has 14,258 treated subjects and 283,896 controls.

The analysis is based on matching, a methodology frequently used in impact evaluation studies because it does not need to specify a particular parametric relation between the outcome and the covariates. Moreover, a matching procedure reduces the number of non-treated to a sub-sample (selected controls) with characteristics similar to the treated individuals. There are mainly two kind of matching estimators: matching on covariates and propensity score matching (see Stuart, 2010 for a review).

In the first case, the matching among subjects is done using all the covariates; in particular the procedure makes use of a distance measure between the *x* of the treated individuals and those of the control individuals to define, for each treated unit, one or more similar non treated units. In our case, the matching on covariate has been implemented on continuous covariates (age, years of education, length of unemployment spell, days worked in the last 2 years, previous occupation[3]), after an exact matching on categorical ones (sex, nationality, area, period, and sector of activity in the last two years). The advantage of this procedure is that matching is exact on some characteristics, leaving the matching on covariates distance only for some continuous variables.

In the second case, individuals are matched using a single index, the propensity score, *p(X)*, that is, the probability of participating in training courses with respect to all continuous and categorical covariates available in the dataset. In this case, matching is done on the basis of the distance of propensity scores of treated and controls individuals. In this case, no exact matching is done.

In this paper, both types of matching procedure have been used, in order to provide robustness to results. In both cases, matching is done only with the nearest neighbor according to the distance measure used; thus, not all controls are used, but only the subsample of non treated individuals more similar to the treated ones.

[1] In the case of multiple treatment (eg. people who benefited from two or more training courses), it was decided to consider the last one; in case of course of the same duration, the longest one was chosen as treatment.

[2] Only 1,5% of controls start a training course in the observation period.

[3] Despite being categorical, this variable was used as a continuous one, giving it an ordinal value starting from the intellectual professions down to the unskilled occupations.

Outcome variables are represented by dichotomous variables indicating whether the unemployed has been hired at least once after 9, 12 and 18 months since the start of the course.

Data

Data used in the impact evaluation of training activities come from three different administrative databases.

The primary source of data for evaluating the training system is the database of the European Social Fund, which represents the main source of funding for the regional training system. The dataset contains information on the trainees (our "treated" individuals), specifically concerning sex, age, nationality, education level, previous work experience and duration of unemployment. Information on courses are also very rich and concern duration, thematic content, class size, cost to public finance, data of beginning and end.

A second database contains information on jobseekers signed up by a Public Employment Service, from which we select the "control group" for the counterfactual analysis.

The third source of data is represented by the Compulsory Communications System of administrative data on employment dynamics, which record all the activations, transformations, fixed-term extensions and anticipated terminations of employment relationships between any worker and employer since the beginning of 2008.

Merging labour market administrative data with the first two aforementioned databases it is possible to check the employment outcomes of both treated and controls and to reconstruct their previous work history. A limitation of the joined data set is the lack of information on self-employment: placement rates are therefore net of activation as self-employed.

Preliminary results

Preliminary results show that training courses have a positive effect on the employment probability of unemployed people. The effect increases with the time spent after the beginning of the course, being 5% after 9 months and 8% after 18 months.

However, the effect is highly heterogeneous between groups of users, as identified on the basis of their distance from the labour market (profiling score). Indeed, our paper will show that the training course does not have any effect on employment probabilities for some kind of users, while for others the effect is positive and statistically significant. However, only for a subgroup the effect is remarkable. Moreover, the analysis revealed that different types of users take the most advantage from different types of courses; those nearer to the labour market benefit more from short courses, while only long courses have an impact on those hardest to place into employment.

References

Imbens G. and Rubin D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86(1), 133-155.

Stuart E.A (2010). Matching methods for causal inference: A review and a look forward, *Stat Sci*. 2010; 25(1): 1–21. doi:10.1214/09-STS313