ACTIVE LABOUR MARKET POLICIES
FOR THE EUROPE 2020 STRATEGY
WAYS TO MOVE FORWARD

Annex: Methodological paper

1. The importance of evidence based ALMPs

In spite of the fact that considerable amounts of money are devoted to active labour market policies in several EU-member states, there is still a lot of uncertainty as to the relative effectiveness (and cost effectiveness) of various active instruments. During its EU-presidency Belgium will be the first country charged with the implementation of the new European Employment Strategy in the framework of Europe 2020. Therefore, this is an excellent time to reflect, more in general, on the transition from the Lisbon strategy towards Europe 2020, and more in particular, on what constitutes effective ALMP's and on how they can be identified.

It would, however, be a bit naive to believe that knowing which programmes are effective and which aren't, will have a direct and unequivocal impact on actual labour market policy practices. Policy choices can be made on the basis of several grounds.

For one, there are political grounds. As regards e.g. the effectiveness of training for the unemployed, policy makers tend to have some firm and strong beliefs which are not necessarily substantiated by the scientific literature.

Another important determinant of actual policy choices is the bias towards maintaining the status quo: there will e.g. often be strong resistance againstabolishing or even merely changing programmes that are already in place. Yet, while acknowledging that policy choices can be a complex process, one of the recurring themes in this conference will be that, when deciding about the composition and the magnitude of the active labour market policies basket, there has to be at least some room for evidence based policy choice.

Evidence based policy, is, in contrast to opinion based policy, a policy approach that *"helps people make well informed decisions about policies, programmes and projects by putting the best available evidence from research at the heart of policy development and implementation"* (Davies 1999).

2. ALMP evaluation: what and why?

Why evaluate active labour market programmes? ALMP's are believed to have some beneficial effects, that is why they were installed in the first place. From the scientific literature, we however know that this belief is not always corroborated by the facts: some programmes indeed have positive effects, but some have no effects or even have negative effects.

It is hard to defend letting the taxpayer pay (sometimes huge amounts of money) for running programmes that have no or negative effects. Therefore, (impact) evaluation is necessary from a simple accountability perspective. Often, there are a lot of different active programmes in one country. Even if they all have positive effects, it is unlikely that these are all equally effective. Here, impact evaluations can reveal where savings can be made.

There are different, often mutually reinforcing, types of evaluation.. In a *process evaluation*, typically questions regarding the actual functioning of the programme are tackled : how is the programme implemented? What is the quality of programme management? Is the target group reached or are there specific selection processes?, etc. In an *impact evaluation*, one seeks to estimate the effectiveness of the programme. But even an effective programme is not necessarily efficient. Here is a role for a *cost effectiveness evaluation*, which seeks to identify how much it costs to obtain the found effect. This is rather important, as e.g. two equally effective programmes may have a quite different cost per participant. Unfortunately, it is not yet customary to include cost effectiveness evaluation in the evaluations of ALMP's, although there are some examples.

Labour market programmes can have several objectives, implying that there is a multitude of possible outcomes (and outcome indicators). Examples of common outcome indicators are e.g. the % of participants that find a job, % of participants that leave unemployment, % that find a stable job, or a stable employment situation. Occasionally, some more qualitative outcomes can be encountered: % that find a decent job (e.g. in terms of wage), effects on health, effects on well-being.

Generally, the observed outcome is not only due to programme participation, but can also be caused by factors outside the programme. Suppose that we observe that 6 months after finishing a training programme 60% of the participants do have a job. This can not entirely be attributed to programme participation: even without participating in the programme, some unemployed would have found a job within 6 months

This implies that in order to find the proper impact of the programme (the "value added", the "impact", the "net effectiveness"), the observed outcome ("gross effectiveness") has to be corrected. In order to obtain the so-called net effect in this case, one has to subtract from the gross effect (60%) *the % of participants that would have found a job even without participating*. Since participants can not at the same time be non-participants, the second quantity (in italics) cannot be observed (and is therefore called a "counterfactual") and must be estimated one way or another.

3. ALMP evaluation: how? (some advice)

On the one hand, there are the so-called non-experimental approaches to estimate a counterfactual (including quasi-experiments). The basic idea of a non-experimental design is to compose a comparison group of persons who are comparable to the participants group, except for the fact that they did not participate. There are several, more or less sophisticated approaches: matching, difference-in-differences, regression discontinuity, instrumental variables, "natural" experiments, etc. In the parallel sessions several will be presented. A potential weakness of the non-experimental approach is that the comparability between the participants and the comparison group may not be perfect, e.g. due to (self-) selection effects. For example, more motivated unemployed are more likely to participate in a training programme, but, because of their higher motivation, would have been more likely to find a job even in the absence of the training. When the participants have a better outcome than the comparison group, one can never be sure whether this is due to the training, or due to their (on average) higher motivation.

On the other hand, there is the experimental approach to estimate a counterfactual. The basic idea of an experimental approach is to take the group of persons who are willing to participate in a programme, and then randomly assign half of them to a experimental group, and half of them to a control group. The experimental group is allowed to participate; the control group is excluded from participation. The results of the control group serve as counterfactual. An obvious advantage of this approach (when compared to non-experimental designs) is that here there is a better guarantee for comparability, factors like e.g. motivation will on average be the same in both groups. In a lot of countries, there is however a strong resistance to social experiments, as they are perceived as a form of unequal treatment.

There is a lot to be said to reconsider this widespread resistance against social experiments. While a social experiment is no guarantee for the absence of methodological problems, a sound experimental design is generally thought to be superior to a non experimental design (in addition, the experimental results are often far more easy to communicate and to understand). Social experiments can detect the proper impact of the programme (such as the "value added", the "impact", the "net effectiveness"), correcting the observed outcomes ("gross effectiveness"). Considering the intrinsic uncertainty about the effects of an active programme (positive?, negative?, none?), the often heard "unequal treatment"- argument does not necessarily hold: the experimental group is indeed treated differently from the control group, but it is not clear who will benefit most beforehand. If, however, experiments are not acceptable to policy makers, the use of an encouragement design could be a middle course: instead of randomly assigning participation, an encouragement to participate is randomly assigned and nobody is formally excluded from participation.

Policy evaluation should be regarded as an integral part of policy making and thus has to be integrated in policy planning. Ideally, the evaluation should be planned before the introduction of a new programme: which evaluation design will be used, what kind of data will be necessary and how to make sure that these data will be available, what funding will be necessary, etc. It is a good practice to involve the evaluator in this process. Although ideally the evaluation is planned before the introduction of a new programme, it is not a very good idea to evaluate a new programme in its starting months. On the contrary, it is far better to wait until things have settled down and some routine is installed. In general, radically changing (or abolishing) a programme before the end of its evaluation, makes the evaluation results somewhat irrelevant.

Finally, ALMPs should also be judged on their macro-economic effects. Thus far, we have argued that it is important to evaluate whether an active labour market programme has a net effect on its participants. However, this is only one part of the story. One should be aware of the fact that even a programme with a clear positive impact on its participants can be undesirable from a societal point of view. A programme typically will not only affect participants, but also non-participants. Well-known effects are the displacement effect (e.g. through a wage subsidy some unemployed persons do find a job, but as a result some of the current employees are dismissed), the substitution effect and the deadweight loss effect. Another effect has to do with the cost of ALMPs: these are typically paid by the government, who will have to collect taxes for this goal. Taxes and more specifically wage taxation probably will have an effect on the supply of, and demand for labour, and thus will affect the employment level. These and other mechanisms will mean that ALMPs not only have an effect on participants, but rather will have an effect on macro-quantities such as the employment level, the unemployment level, productivity and the matching effectiveness to name but a few. In the end, these macro- economic effects can be considered as the final yardstick to assess whether ALMPs are successful or not. As measuring these macro effects obviously is (even) more difficult than estimating effects on the participants, it won't come as a surprise that there is still a lot of uncertainty with regard to the size and the magnitude of macro-effects of ALMPs.

4. Potential conflicts in ALMP evaluation

In the relation between the evaluator on the one hand, and policy makers/ evaluation sponsors/ programme administrators on the other hand, there are some potential conflicts.

Conflict 1. Different time perspective. Policy makers, evaluation sponsors, programme administrators etc. often immediately want evaluation results, whereas the evaluator will insist that a thorough evaluation takes time. Impact evaluation results necessarily will only be available some time after participation, since it is necessary to follow up former participants' labour market results at least for some time after participation. The resulting "this is old stuff"-argument is not per se valid.

Conflict 2. Different expectations as to the expected output of the evaluation. Policy makers, evaluation sponsors, programme administrators etc. desire "usable information" (e.g. what can be used to adjust and manage the programme), whereas scientific evaluators often are (somewhat myopically?) in the first place interested in the validity of their impact estimates. Without a doubt, more attention should be devoted to black box issues: we know that some programmes do (not) work, but we often don't know enough about why they (don't) work. By uncovering the relationship between the effectiveness on the one hand, and specific design aspects of the programme on the other hand, the evaluation results can be made more useful, both from a policy and a scientific point of view.

Conflict 3. Different expectations as to what should be the consequences of the evaluation results. Evaluators, in a rather objective-instrumental view on how policy making works, tend to expect that their evaluation results will have consequences for the running programme: depending on the direction of the impact estimates, the programme will either be continued, expanded, restructured, redesigned or even abolished. Policy makers, on the other hand, only seem to be interested in impact estimates when these are positive, while negative results often are downplayed or outright neglected. This state of affairs is not particularly helped by the fact that a lot of evaluation studies come to contradictory conclusions. For example, with respect to training for the unemployed, one can find widely differing results. However, one should be aware that one training programme can be radically different from another one (short versus long, comprehensive or vocational, etc.). Here meta-analysis, where all available studies are bundled in order to see what will rise to the surface, definitely can help.

5. Knowledge gaps with regard to effectiveness research

To conclude, we provide a list of **effectiveness-related problems and issues that remain partly unanswered to date**, and thus are interesting when commissioning an evaluation or when attending the presentation of an actual evaluation study.

- Is the net effectiveness related to specific groups (age, sex, skill level, ...) ? What does (doesn't) work for whom and why (not)? An intense classroom training for recent high school drop outs might e.g. not be the most fruitful approach;
- Is the net effectiveness related to the combination of several policies? Is e.g. the combination of vocational training and job search training more effective than providing vocational training on its own? Is the order of the distinct policies relevant, e.g. first the vocational training and then the job training, or the other way round?
- Is the net effectiveness related to the timing of intervention : e.g. at the beginning of unemployment spell or later?
- Is the net effectiveness related to labour market institutions? Is there e.g. an interaction between the effectiveness of active policies and the characteristics of the unemployment benefit system?
- Is the net effectiveness related to the intensity or "dose" or duration of the programme?
- Is the net-effectiveness different between public and private providers?
- Is the net-effectiveness different depending on whether the programme is local or nationwide?
- Is the net-effectiveness different between favourable and unfavourable business cycle conditions?
- Is there a difference between the short run and the long run effects? Are the effects lasting? Is a locking-in effect (less job search during programme participation) possibly offset by more favourable long term effects?
- What is the external validity of the results? The evaluation results apply, strictly speaking, only to the participants' sample that was used in the evaluation study. Can one be sure that future participants will be comparable, will behave identically, will face an identical context?