The Case for Randomization in the Evaluation of Active Labor Market Policy

Mauro Sylos Labini

University of Alicante and IMT Lucca

ABSTRACT

The present essay examines the case for social experiments in evaluating active labor market policy. First, basic advantages and disadvantages of controlled experimentation are assessed. Second, the reasons why social experimentation may turn out to be important in shaping new form of work are discussed.

Keywords: Social Experiments, Labor Market Policies, Program Evaluation.

March 31, 2007

---

1Prepared for the "Impulses from Salzburg 2007" Workshop at the Center for Ethics and Poverty Research of the University of Salzburg.
1. Introduction

The impressive pace of institutional and technological change occurred in the last century has to some extent reinforced what the Nobel laureate in physics Nils Bohr once said: "Prediction is very difficult, especially if it’s about the future". The present essay, rather than trying to forecast general and complex trends concerning labor characteristics, concentrates on the future of labor market policy. More specifically, it examines the case for social experiments in evaluating active labor market policies. Even if the topic might seem peripheral to the present cluster (The Future of Work), in the last section of the essay we shall try to argue that initiatives likewise the New Work Project directed by Frithjof Bergmann, especially if financed with public money, may be envisaged as social experiments and could be evaluated using randomization.

For the layperson an experiment is any major deviation from past policy or practices. However, the scientific notion of experiment is much narrower. In particular, it requires that the researcher controls both the variable under investigation and the environment in which this variable is observed. Of course, accurate measures of the effect require reliable basis of comparison and outside the laboratory this is far for being trivial.

A natural basis of comparison is the state of the variable under investigation before the experiment takes place. However, the so-called before-and-after comparison is not always accurate. The impact of minimum wages on employment, for example, is not detectable simply measuring the change in employment levels occurred in a given area after that a raise of minimum wages has taken place. In fact, taken for granted that many additional time-variant factors influence employment, it would be inappropriate to attribute the whole change detected to a single policy measure. To repeat, the basic problem is to establish a credible basis of comparison, i.e. what one may call the control group has to be as similar as possible to the treatment one. In the minimum wage example a reasonable procedure would be to confront the change occurred in an area where minimum wages have been raised with changes occurred in the same period in a very similar area where minimum wages level has remained constant.

Fisher (1929) has probably been the first to argue that the only satisfactory method of achieving reliable groups is to assign subjects (in the above example geographical areas) to the treatment group at random. This kind of procedure is often referred as a randomized experiment.

Running experiments happens to be much easier for natural scientists than social ones. One may even argue this is one of the major epistemological differences between the natural and social sciences. Nevertheless, contrary to the above position social experimentation is feasible and dates back more than thirty years. Greenberg and Shroder (1991) identify more than 90 field trials in social science research areas including social insurance, labor supply, worker training, and housing subsidies. In the present essay we examine the rationale for field experimentation in evaluating active labor market policies (ALMP).

Active labor market policies differ from others policies affecting the labor market in two basic
Fig. 1.— Spending in ALMP as percentage of GDP 2002


ways: first, they are targeted towards the unemployed or towards individuals with low skills or little work experience. Second, they are aimed at promoting employment or wage growth among disadvantaged groups, rather than transferring income (e.g. unemployment benefits).

As showed in Figure 1 expenditure for ALMP are not negligible in many European countries and are higher than in the US. Ironically, much of what we know about the effects of such policies and many of the methodological developments used to evaluate them come form US based studies. This is indeed one of the reasons why a more careful and detailed assessment by European policy makers is needed. Interestingly enough, Kluve (2006) has found in his recent survey that rather than country specific factors such as labor market institutions, it is almost exclusively the program type that matters for program effectiveness. In what follows, five broad set of policies are consideres as ALMP:2 (i) classroom training consisting of education to remedy lack in general or vocational skills; (ii) subsidized employment, which includes both public service employment and wage supplement to private firms for hiring new workers; (iii) subsidies to workers and private firms for the provision of on-the-job training; (iv) training on job searching and how to obtain a job; (v) non-monetary subsidies to job search. Incidentally, New Work initiative may be envisaged as a mix of the above measures.

2This classification borrows from Heckman et al. (1999).
The rest of the paper is organized as follows. Section 2 points out the basic advantages of controlled experimentation over other methods. Section 3 presents some of its main limitations and the criticisms leveled in recent years. Finally, section 4 tries to pinpoint some of the reasons why social experimentation may be relevant for the future of work.

2. Causal Inference and Randomization

Politicians may be skeptical of a research methods involving experimentation with human objects. It is therefore important to be very clear in highlighting strengths and limitation of this research tool. The basic advantage of experimental methods over other non-experimental ones is that they allow to measure the causal effect of a given policy with high reliability. To fix ideas, let us use the notion of potential outcomes, introduced by Rubin (1974). Suppose we are interested in measuring the impact of a given training program on individual wages. Let us define $Y^T_i$ as the wage of individual $i$ if she enrolled in the training program under study and $Y^C_i$ as the wage of the same individual in case she did not. Let us also assume that we are interested in $Y^T_i - Y^C_i$, which is the effect of the training program on individual $i$. Of course, it is not possible to observe both outcomes for the same individual. This problem, sometimes called the fundamental problem of causality, makes explicit the impossibility of estimating individual treatment effects. However, we can hope to learn something about the problem under scrutiny estimating the expected average effect that training program have on a given population of workers:

$$E[Y^T_i - Y^C_i].$$

Note that we do not assume that the effect is the same across workers, rather we aim at calculating an average allowing for heterogeneity. Let us imagine we have access to data on a large number of workers of a given population, and some of them are actually enrolled in a training program ($G_i = 1$) and some did not ($G_i = 0$). A possible approach is calculating wage averages of both groups and computing the difference between the two. In large sample this will converge to:

$$D = E[Y^T_i | G_i = 1] - E[Y^C_i | G_i = 0].$$

If one adds and subtracts $E[Y^C_i | G_i = 1]$, i.e. the expected wage of an individual actually enrolled in the program had she not enrolled (a quantity not observable, but logically well defined), one obtains:

$$D = E[Y^T_i - Y^C_i | G_i = 1] + E[Y^C_i | G_i = 1] - E[Y^C_i | G_i = 0].$$

The first term $E[Y^T_i - Y^C_i | G_i = 1]$ is the so called average treatment effect on the treated. In our case, is the average effect of the programs on those who participated. The second term $E[Y^C_i | G_i = 1] - E[Y^C_i | G_i = 0]$ capture the difference in terms of expected wages between the two groups and is commonly dubbed the selection bias. There are good reasons to believe that the latter term is different form zero. For example, those who enroll in training programs may be more motivated and therefore more productive than those who did not enrol. Alternatively, those who participate...
may be the most needy and therefore less skilled. Both factors obviously have an impact on wage. The general point is that simply calculating the average of groups and computing the difference may generate biased estimates, given that there might be systematic difference between individuals who participate in training programs and those who do not. Moreover, given that $E[Y^C_i|G_i = 1]$ is not observable, it is very hard to assess the magnitude and even the sign of the bias.

Most non-experimental methods try to correct for the above bias or to identify situations where it does not exist. In both cases it is necessary to rest on untestable assumption about either individual unobservables or the decision to participate in a program. On the other hand, if we could assign randomly part of the population to training program, treatment and control groups would differ in expectation only through their exposure to the treatment. This implies that the selection bias, $E[Y^C_i|G_i = 1] - E[Y^C_i|G_i = 0]$, would be equal to zero. Moreover, if the potential outcomes of an individual are unrelated to the treatment status of any other individual, we have that:

$$E[Y^T_i|G_i = 1] - E[Y^C_i|G_i = 0] = E[Y^T_i - Y^C_i|G_i = 1] = E[Y^T_i - Y^C_i],$$

which is exactly what we wanted to estimate in equation 1. Therefore, when a randomized evaluation is implemented it provides an unbiased estimate of the impact of a given policy on the sample under study.

To conclude this section, is worth mentioning a second set of "less technical" and to a certain extent more controversial advantages of experiments over more traditional methods. First, in many cases experiments allow to measure the effect of policies that have not previously been implemented. This is obviously impossible for empirical studies who has to rely on traditional sources of data. Second, the simplicity of experiments offers clear advantages in making results convincing for politicians and public opinion in general. In fact, contrary to most claims derived from nonexperimental investigations, experiments permit to drive analytical findings that are not subject to the complicated qualifications of most standard econometric methods.

### 3. Randomization and Its Discontents

Notwithstanding the advantages discussed in the previous section, randomized field trials face numerous problems. A preliminary aspect concerns the ethical issues raised by experimentation with human beings. In fact, random assignment is often considered as an unfair way to ration public resources. Why, for example, individuals who are eligible for a training program should not participate because of a lottery? However, similar ethical concerns are also present in studies of new medicines and medical procedures, where stakes for participants are often much grater than in social experiments. Still such trials are in many cases required to prove efficacy of new medical treatments. Moreover, good experimental design can reduce ethical objection: first, randomization

---

3See for example the different views expressed by Burtless (1995) and Heckman and Smith (1995).
may be undertaken only among those subjects who are willing to participate to a given program. In this way one would avoid forced compliance. Second, in certain circumstances it is possible to compensate individuals who are offered potentially harmful treatments or denied beneficial services. More generally, the basic and more convincing ethical argument in favor of experimentation is that it is better to inflict possible harm to a limited number of individuals through a small scale experiment rather than on a much larger scale as a result of unsound public policy.

The most popular criticism to experiments is that they do not reflect the general equilibrium effect of a particular policy. The basic intuition here is that small scale random evaluations compare the difference between treatment and control groups in a given region without assessing the effect that a policy may have on other subjects and/or other geographical areas. For example, in the evaluation of a training program that randomly subsidize a few workers, their advantages may be magnified by the fact that few other people in the same local market receive additional training. Moreover, little is known on how the experiments itself affects other important aspects of the local economy like the price of training courses, workers job search activity, and firms decision to implement training programs.

It is certainly true that randomized experiments are able to measure only partial equilibrium responses to policy changes. Nevertheless, sometimes this does not seem a real problem, given that the parameter of interest is exactly the partial effect. For example, if a small scale program whose targets are disadvantaged workers is found to have an impact on their labor market outcomes and from a policy perspective we are interested in the welfare of those particular workers, the general equilibrium effects are probably small and irrelevant. Moreover, one may argue that even if the objective is general equilibrium effect, knowing the partial equilibrium one is better than nothing. More generally, researchers may try to overcome the problem in a number of ways: first, one wants to be extremely clear on the population of interest; second, if the possible treats posed by non measurable general equilibrium effects are explicitly addressed, they can be tested using additional control groups.

A last problem we want to explicitly address is the ability of randomized experiments to capture structural parameters. This is a subtle but important point: experiments allow to estimate the overall impact of a particular policy, such as a subsidized training program, on an outcome, such as worker wages or satisfaction, allowing other inputs to change in response to the program. This might well differ from the impact of a training program on wages keeping everything else constant.

To see the difference between the two effects, let us assume that it exists a function $Y = f(I)$, where $Y$ is the outcome of interest and $I$ is a vector of inputs, one of which, let us call it $p$, is a given policy. Of course $p$ may also have an impact also on other inputs. The relation between $Y$ and $I$ is structural in the sense that it holds regardless of the actions of individuals or institutions affected by the policy changes.

Consider now an exogenous change in the policy $p$. One interesting estimate is how a variation in $p$ affects $Y$ when all other input are held constant. This is the partial derivative of $Y$ with
respect to \( p \). Another interesting estimate is the total derivative of \( Y \) with respect to \( p \), which includes changes in \( Y \) caused by variation in the other elements of \( I \) as a result of changes in \( p \). Both derivatives may be interesting for policymakers. The total derivative tells us what happens to outcomes after an input is exogenously provided and agents react. In some sense this is the true impact of the policy. Nevertheless, it may not provide a reliable measure of overall welfare effects. Consider for example a policy of providing subsidies to job search activity. Workers may respond to the policy by decreasing their search effort in favor of some leisure activity. The total derivative of wage or other occupational outcomes will not capture the welfare increases due to more leisure. On the other hand, partial derivative may provide an appropriate guide to the welfare impact of the policy.

To be sure, results from experiments provide reduced form estimates of the impacts of a given policy and therefore what one ultimately obtains are total derivatives. Partial derivatives can only be inferred if one is able to specify a valid model that links inputs to the outcomes of interest and collects data on these intermediate inputs. This underscores that experiments are not substitutes for economic models. Rather, to estimate welfare impact of a policy, randomization needs to be combined with theory.

4. Discussion

To which extent the above assessment of social experimentation is relevant for the Future of Work? Does the research tool assessed in this essay contribute to predict which forms of work would become determining in the next decades? These are open questions and this section is meant to argue that social experimentation might well be important in shaping new forms of work.

Since Adam Smith, economists acknowledge that many of the improvements in the productive power of labor have been the effects of the division of labor. At the same time some of its drawbacks have also been underscored. In particular, the social and moral consequences of a system that if pushed too far may deprive the workman of his independence and weaken intellectual skills have been studied. To be sure, technological and institutional change have often interacted with workforce specialization and standardization in unpredictable ways changing the basic characteristics of gainful employment. In the introduction we acknowledged that it is very difficult to foreseen which forms of work will prevail in the balancing between the benefits of the division of labor and its harmful consequences. We then concentrated a research tool that is able to measure the impact of active labor market programs on different kinds of labor market outcomes. Sharp and well functioning policies might well shape the future of labor.

New Work organization could be a good example here. Founded in the early 80s also as a consequence to the massive layoffs of General Motors workers, it was originally intended to give people practical guidance in exploring new employment opportunities. Nowadays New Work has come to embrace a more ambitious program dedicated to reformulating the way people conceive their work
activity. Even if it is an open question if such programs are suitable for randomization, in some broad sense New Work may be consider a sort of social experiment. Together with detailed accounts on how New Work initiatives have so far worked (i.e. case studies), randomization over distinct locations may represent an important source of information on the impact of New Work program. Moreover, for the reasons presented in Section 2, it could facilitate public opinion mobilization and politicians’ awareness. An important issue concerns the selection of a few outcome variables. In other words it is important to be clear on which dimensions one expects New Work offices do have an impact. To the best of my knowledge, given the objectives of the initiative, a meaningful option, together with standard labor market outcomes, is collecting an array of indicators on individuals job satisfaction.

REFERENCES


