Cuadernos de Economía, Año 40, Nº 121, pp. 589-598 (diciembre 2003)

COUNTERFACTUAL TO A MEASURE OF ITS IMPACT

Arístides Torche*

1. Introduction

The Department of Economics of the Catholic University of Chile has a long tradition of evaluating social projects, in other words, projects involving investment in human capital. It started) in 1977, with the first evaluation of the National Complementary Feeding Program (PNAC) and continued with the study of Full Care Centers for Minors (2-6 years old) in Extreme Poverty (Centro de Atención Integral para Menores en Pobreza Extrema, (CAI) and the study of the Centros de Atención Diurna (CAD), where primary and secondary education students in poverty would attend during the half day they are not in school, to name a few of the several social programs evaluated by the Inter American Course on Project Evaluation (CIAPEP). This effort continued with an analysis of housing subsidy systems conducted in the late 80s, and it continues now in 2003 with the evaluation of the labor and economic reconversion program which was put into operation when the coal mines were closed in Lota in the late 90s.

2. Project Evaluation: Methodological Elements

The purpose of this article is to describe the specific forms adopted by general project evaluation methodology in the case of social programs. Particularly, the definition of the "without project situation" and the mechanisms to identify the benefits that can be attributed to the project and only to the project.

But, which are the main elements of so-called project evaluation methodology?

In project evaluation terminology, a project may be understood to be a series of activities to assign resources in order to achieve certain objectives. The principle of separability establishes the convenience of limiting objectives to only one, if possible, in order to accurately identify the activities that generate the profitability in the project.

It is also necessary to identify and value the resources used to obtain the cost stream. Two principles have been used extensively in determining costs: the benefits in the best alternative use of the resources, and the possibility of truly

* Department of Economics, Catholic University of Chile. Email: atorche@faceapuc.cl
The author would like to thank Professor Rodrigo Cerda and Ernesto Fontaine for their valuable comments.

freeing them for other alternatives uses if the project is not carried out (principle that use of the resource may be avoided if the project is not carried out).

Determination of the project's benefits implies expressing its objectives in operative terms so they can be identified, measured, and valued in terms comparable with costs (in monetary units). In this case it is necessary to have a point of reference to make comparisons. This is the so-called "without project situation", which is the baseline against which the benefits are going to be measured. In general, the "without project situation" implies maintaining the status quo, projecting its normal evolution in certain cases. The benefits of the project are measured and valued by what the project adds beyond that state. For example, if the construction of a new port is evaluated in an area that has other ports, the "without project state" should include the normal growth of existing ports, so that the new port's benefits are only the shorter waiting lines over what they would have been with the existing ports, considering those modifications.

The "without project situation" normally coincides with the state in which the project does not operate. However, in the case of social programs that intend to apply their activities to people it is not possible to determine their behavior in the "without project situation", after they received the effects of the project. This is what is called the missing data problem. People who have not received the benefits of the project are not necessarily the same as the ones that participated in it, and therefore considering them as part of the "without-project state" might be inappropriate. In summary, if one wants to identify, measure and value the benefits, it would be necessary to locate at least one "clone" for each one of the participants and make the comparison between the beneficiary and his "clone". Or, in broader terms, a population that can be considered similar to the one that has received the benefits (control group).

The second point of interest consists of establishing whether the project is actually responsible for the changes observed in the beneficiaries. To make these ideas clear, let us suppose that a training project that has given specific training to a group of people is being evaluated. Suppose we want to measure the impact of the project by the change in the beneficiaries' access to jobs compared to their "clones".

Is it possible to assure that income differentials between trained people and their "clones" can be validly attributed to the training program? And, if the answer is affirmative, can one be sure that the results will be the same if the program is extended to other people?

These two problems: the identification of "clones" that define the "without project situation" and the determination of the benefits that can be attributed to the project, constitute two key points in the analysis of evaluation of social programs that will be studied below in more detail.

3. Construction of the Counterfactual Scenario

The idea of a counterfactual state has a long tradition in economics, and it has been used extensively in economic history. It consists of determining what would have happened if the subject being studied had not been present. An outstanding example is Fogel's work on the effect of railroads on the growth of the American economy. In that case, Fogel asks what would have happened to the American economy if railroads had not existed. There would have been more investment in roads and navigable waterways; cities might be closer to navigable waterways, with a city like Saint Louis being more developed and another city like Denver being smaller. (Mc Closkey D, 1987). The counterfactual make possible to construct the "clones" and to define the relevant statistical estimators.

In the case of social programs, two counterfactual lines have been proposed: 1) The first line is related to experimental analysis where chance is left to determine the control group and the treatment group, 2) The second line refers to non-experimental designs, where the following counterfactuals have been considered (i) The beneficiaries themselves before receiving the program's activities (ii) the group of non-beneficiaries and (iii) for each beneficiary, a non-beneficiary that is considered his clone and is matched up with him. In any case, the control group defined by each one of the counterfactuals is expressed as the group of people that possess an X vector with similar characteristics to the beneficiaries.

Associated with the first two counterfactuals presented above for the case of the non- experimental design (i and ii), the following estimators are used most frequently: the before/after estimator, the cross section estimator, and the differences to differences estimator.

If it is assumed that project outcome can be validly measured by working income, the before/after estimator compares the beneficiaries' working income before the program with their income after the program has been completed. The cross section estimator compares the program beneficiaries' income with other people similar to them who have not received the program. In this case, they are different people who are evaluated in the same period. Finally, the differences to differences estimator compares beneficiaries and non-beneficiaries before and after the program in order to establish the difference between the change in the beneficiaries' income before the project and afterwards, with the difference in non-beneficiaries' income before and after the project. The objective of using this procedure is to eliminate the effect of economic trends between the initial state and the post-project state.

The "before/after" estimator refers to the same people, and therefore it is not affected by problems of personal heterogeneity. Nevertheless, changes in income due to changes in the level of economic activity between the time when the program starts and the end of the program (after the program) may be erroneously considered as an effect that is attributed to the program.

The cross section indicator does not have the problem of change in trend, because the incomes of the beneficiaries and the control group are measured at the same time. Nevertheless, they are different people, and therefore there may be a selection bias; in other words, the people who chose to enter the program are different from the ones who did not. There was a process of self-selection that may affect the income level, in addition to those from the project itself. The selection bias resulting from the fact that people willing to participate in the program have some characteristics that make them different from non- participant, is one of the main problems in project evaluation. In this case differences in the outcome variable, i.e working income, are not only due to project impact but also to the personal bias of the beneficiaries.

The differences to differences estimator tries to remedy the selection bias and the effect of the growing economy through time. On the one hand, it compares the benefits of the beneficiaries, before and after the project and, on the other hand, it compares people in the control group also before and after. To the extent the control group's income has changed in a different way from the beneficiary group's income, the difference in the differences is indicative of the impact of the project.

Sometimes, it is not possible to construct the difference to difference estimator because there is no data on the outcome variable for participants prior to the program. There is a technique for dealing with the selection bias in those cases; it consists of selecting a set of variables X so that the outcome variable "y" will be independent of the project participation variable P, when controlling by X, i.e. (y II P/ X) . When those variables have been obtained, it is called ignorability of treatment (Rosenbaum & Rubin 1983)[1] because given X, the selection bias of project beneficiaries (treatment) do not exist any more i., e. The bias have been captured by the persons X characteristics.

On technical grounds, it can be said that it is enough for ignorability a some weaker assumption namely that $E(y/X; P=1) = E(y_1/ X)$ and $E(y/X; P=0) = E(y_0 / X)$, where P =1 indicates participation in the project, and P = 0 otherwise, and y_1 is the outcome variable for participants after participating in the project and y_0 is the outcome variable for non-participants.

Normally, the counterfactuals are reserved for the three estimators mentioned above; but there are non-parametric methods like matching that are procedures to build the clone or clones of each one of the beneficiaries, which should also be considered, by extension, among the counterfactuals.

The matching techniques consist of using the set of X variables to locate the "clone" of each beneficiary as the person, or people, who is not a beneficiary of the project and has the closest structure of X characteristics to the beneficiary he will be the "clone" of. This procedure is very hard to operate if the X matrix has more than one variable of characteristics, because of the difficulties in measuring the X_p characteristics that are closest to the beneficiary's X_b. Rubin's theorem establishes that if the outcome variable, which in this case is the participants' income y, is independent of variable P, which indicates participation in the program (P=1 if he participates and P=0 otherwise), when it is controlled by person X's

1 Ignorability holds trivially if P is a deterministic function of X which is called selection of observables (Heckman and Robb 1985)

characteristics like age, sex, education, among others, then there is a function p(X) so variable “y” is independent of P conditioned by p(X). In formal terms: if y is independent of P given X, i.e. (y II P/ X), then there is a function p(X) so y is independent of P given p(X), i.e. (y II P/ p(X)) where the function p(X) can be interpreted as the probability that a person with X characteristics participates in the program. (Dehejia, Rajeev and Sadek Wahba 1998)

In summary, the procedure consists of the following steps: a definition is made of people’s X characteristics that may be supposed to make income y independent from attributions to the program (P=1), after it has been controlled by them. Then, probability p of participating in the program in view of those X characteristics is calculated, and the clone, or clones, of each beneficiary with X_b characteristics is chosen as the person that minimizes the difference of his or her probability of being accepted in the program $p(X_c)$ and the probability $p(X_b)$ of the beneficiary, in other words $p(X_c) = \min \{p(X_r) - p(X_b)\}$, where r refers to any other person in the control group.

The study of siblings and twings provide another way to construct “clones”. In these cases, analysis of panel data has been used in order to follow the activities of beneficiaries and control groups over several years. (Griliches 1979), (Ashenfelter and Krueger , 1994), (Ashenfelter and Rouse, 1998).

4. The Determination of the True Benefits of the Program

A second problem that program evaluation poses refers to the method followed in order to make sure that the benefits are due to the project. This is a matter of determining that the program is the “cause” of the increase in working income that has been observed among its beneficiaries, compared to the participants in the control group.

The Dictionary of the Royal Academy defines causality as a “law under which effects are produced” and the Salvat encyclopedia adds “and also relationship between cause and effect.” In this perspective, the causality relationship arises from a theory, and what empirical analysis does is establish procedures to recover the value of parameters with the information that is available. For example, geometry teaches us that the perimeter of a rectangle is equal to twice one side (l_1) plus twice the other side (l_2). In other words: $P = 2\, l_1 + 2\, l_2$. The problem arises because the available information may be incomplete. For example, no information may be available on the length of the second side. In this case the estimable model will be $y = \alpha\, l_1 + u$. The purpose of empirical analysis consists of designing estimation mechanisms that make it possible, with the available information, to recover the parameters’ values; in this case, $\alpha = 2$.

These procedures become more relevant because often theory does not provide precise information about the parameters’ value, which is added to the problem posed for unobservable variables. For example, the theory of demand establishes a direct and inverse relationship (when one variable increases, the other decreases) between the quantity demanded and the sale price, but it does not delve into further indications about the value of that slope parameter. It is very important to have estimation methods that make it possible to recover the true but unknown value of the parameters.

In regard to the determination of a program's impact, the causality problem can be expressed in a double perspective. On the one hand, by determination of an equation that reflects the theoretical relationship that links the program to the objective variable that is to be modified. To clarify, suppose that a social program aims to improve the nutritional state of the pregnant woman and to monitor her health during pregnancy. Suppose that a positive relationship has been established at a theoretical level between the nutritional state of the pregnant woman and the birth weight of her child. The empirical model can emphasize the specific channels whereby the program would affect the birth weight (structural model), or a quicker path can be taken and a relationship can be established between the result variable (birth weight) and attribution to the program, in other words, the equation can be set forth directly: $Pn = f(X) + \alpha_p P + u$ (reduced model) where matrix X is indicative of other variables considered important to determine birth weight, like the mother's age, other children born live, etc., which have been selected to make Pn independent of P when controlling for them, i.e there is ignorability of treatment. In this case, P is the mute variable of attribution to the program that has only two values: 1 if the person is a beneficiary of the program, and 0 if he is not. But in a deeper sense it summarizes the channels through which the effect of the program occurs; for example, better nutritional state during pregnancy, control of iron and other micronutrients, and health checkups, among others, and u is indicative of all the other unobservable variables that also affect birth weight, such as hereditary genetic conditions.

An "adequate" estimate of the unknown parameter α_p permits a precise evaluation of the program. But what do we mean by an "adequate" estimate of the parameter? There are two properties that have been required of estimators in order to be "adequate." They are Unbiasedness and Consistency. Unbiasedness refers to the fact that if many samples are taken, the average of the estimator's values that result from the different samples coincides with the value of the parameter. Consistency establishes that the estimator's value approaches the value of the parameter if the size of the sample approaches the size of the population.

Given the base econometric model, $y = f(X) + u$, where "u" is an error term, the classical assumption that the expected value of u given X is null i.e. that $E(u/X) = 0$ play a crucial role. First, it permits to interpret f(X) as the expected value of y given X. In other words, $f(X) = E(y/X)$ which relies the econometric model to the substantive one. Second, it is not hard to show that $E(u/X) = 0$ is also the key assumption in order to comply with unbiasedness and consistency. In fact, given model $y = X\beta + u$, one method to estimate the parameter or vector of β parameters, in an unbiased and consistent way, consists of using the information contained in X and constructing the expression $X'y = X'X\beta + X'u$, where X' indicates the transposed matrix. If $X'X$ can be inverted, then: $(X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$ and $(X'X)^{-1}X'y$ can be considered an "adequate" estimator of β if $E(X'u) = 0$ provide $E(X'X)$ is finite.

E(u/ X) = 0 implies that E(xi u) = 0 and every xi variable which does not satisfy the basic assumption i.e., where $E(xi\,u) \neq 0$ is called endogenous. The main causes of endogenousness are: omitted variables, errors of measurement in variables and reversion of causality, also called simultaneousness. The case of omitted variables and simultaneousness will be further studied in this article. In fact, the two most serious difficulties that prevent customary estimation methods like ordinary least square and maximum likelihood from producing consistent estimators are the omitted variable and reversion of causality.

The problem of the omitted variable arises when, for example, in a study that measures the effect of education on income ($y = \alpha_0 + \alpha_1 e + u$), there is a non-observable variable like skill (H) which affects educational level and income received. In that case, the relationship between income and education ($y = \alpha_0 + \alpha_1 e + u$) would be spurious, because it would be affected by the skill variable (H). In particular, the α_1 parameter does not indicate the true impact of e on Y. It indicates the more complex impact of H on y through e and u. The omitted variable generates problems when it simultaneously affects y (the dependent variable) and e (the independent variable). When those omitted variables exist, then $E(e \cdot u) \neq 0$. In that case, it is said that the education variable is related to the error term, in other words, that it is endogenous.

An interesting case of omitted variables is associated with the problem of selection bias that has been mentioned previously. This problem can be presented for example, when there are two alternatives to reach a certain objective, and one of them is to be used as the control group for the other. This situation arises in the evaluation of subsidized private schools, when municipal schools are used as the control group. If an outcome variable is chosen, such as the Simce test[2] for example and the difference in that test among students of subsidized private schools and municipal schools is considered a measurement of the incremental impact of the “subsidized private school” program, an error is made, because the students are not distributed randomly between the two kinds of establishments. There is a self-selection bias, which is the case in several non-experimental project designs . It is necessary to model the selection mechanism that leads certain students to prefer the subsidized private school and others to prefer the municipal school. The modeling of this selection process makes it possible to define a new variable that has the characteristics of an omitted variable, which must be added to the other ones in the model in order to complete it (Heckman, J., H. Ichimura and P. Todd, 1997).

In the description of a model that sets forth the impact of subsidized private schools, one may consider that the expected performance depends on the characteristics of the students and the characteristics of the schools. Nevertheless, the characteristics of subsidized private schools compared to the characteristics of municipal schools constitute the differences between the “treatment” group

2 SIMCE is a national wide test taken every year in Chile in order to assess learning achievement in primary and secondary school.

and the control group, and it is not necessary to break them down further. It is not necessary either to further break down the characteristics of students if they are distributed randomly in the different establishments, because then they are balanced through the population and systematic biases are not generated.

If the assignment of students is not random, then certain students with specific characteristics will go to subsidized private schools and others with different skills will go to municipal schools. The systematic effect attributable to students must be eliminated in order to have the pure effect of the project. In that case it is necessary to model the selection procedure in terms of the particular characteristics of the students that went to one of the groups of establishments or the other.

There are at least three procedures for eliminating the problem of the omitted variable:

i) Look for information about the unobservable variables that could be related simultaneously to the dependent variable and to some of the independent variables.

ii) Replace the endogenous variable x with another one that is related to x but not to the unobservable variable (H in this case). This is the choice of an instrumental variable that only affects the dependent variable "y" through variable X

iii) Model the correlation between variable u that is the error term in the initial model and the level of education (e). These are the sample selection models that explicitly incorporate the effect of E(u e) when it is not equal to zero, i.e when e is an endogenous variable

The other broad problem is causality reversion that arises from a joint determination of the dependent and independent variables of a model. In fact, suppose a model is used where the health status, ES, would be a function of the use of the health insurance ($ES = \alpha_0 + \alpha_1 SS + uss$) chosen, among other variables. However, the choice of the kind of health insurance also depends on the health status ($SS = \beta_0 + \beta_1 ES + v$). In that case, $E(SS\ u) \neq 0$, which expresses that the SS variable is endogenous. It is interesting to note that although the problem of causality reversion is very different from the problem of omitted variables, both are expressed as a situation of endogenousness of some dependent variable. (Dowd and Town 2002).

When causality reversion exists, the only mechanism for solving the problem consists of manipulating the SS variable by identifying an instrumental variable that is related to SS, but not to the variable u.

5. Average Measures of Outcome Commonly Used in Micro-econometrics

Econometric work on social projects has dealt with three average measures of outcome. The average treatment effect (ATE), which is the expected effect of the program over and above the non-treated, measured in a person drawn randomly from the population i.e. $E(y_1 - y_0)$ (Wooldridge, 2002). The ATE is a special case of the average partial effect for the binary P variable: beneficiaries of the program and others. The second measure is the average treatment effect on the treated (ATT), which is the average effect of the program on those who actually participate in it. It is the impact of the treatment over and above the treated if they had not participated. Finally, the local average treatment effect (LATE) introduced by Inbens and Angrist in 1994, measures the impact of the program on people that satisfy some condition expressed through an instrumental variable or a set of instrumental variables. Assume that attendance to subsidized schools is being studied, and a variable that measures whether or not the mother is a high school graduate is chosen. (Angrist, Joshua 1998) LATE measures the average effect of the program (attendance at subsidized school) for those students in the program whose mothers are high school graduates. These statistics are interesting in themselves, but also because of their relationship to the traditional measures of benefit employed in net present values, which is the main measure of project evaluation.

6. Conclusions

This article presents two basic elements for the evaluation of the impact of a project: the without project situation and the methodology for determining the benefits attributable to the project. It has been shown in several articles that those principles acquire a particular connotation in the case of social programs, because the without project situation acquires the form of a control group and the people participating in the program cannot be part of that control group, which makes it necessary to design procedures to find people similar to the beneficiaries: i.e, their clones. On the other hand, the choice of those clones creates limitations on the definition of the methodology used to determine the benefits of the program. Problems of the omitted variable and causality reversion arise, which are some of the most frequent reasons why “adequate” estimators of project impact are not obtained. The article ends with a brief description of procedures to eliminate those problems.

REFERENCES

Angrist, J. (1998), “Estimating the Labor Market Impact of Voluntary Military Service using Social Security Data on Military Applicants”. *Econometrica* 66, 249-288.

Ashenfelter, O. and A. Krueger (1994), “Estimates of the Economic Returns to Schooling from a New Sample of Twins”, *American Economic Review*, 84, 5, 1157-1173.

Ashenfelter, O. and C.E. Rouse (1998), “Income, Schooling and Ability: Evidence from a New Sample of Identical Turns”. *Quarterly Journal of Economics* 113, 253-284.

Dowd B. y R. Town (2002), “Does X Really Cause Y? in Academy Health, The Robert Wood Johnson Foundation. 1-21.

Griliches (1979), “Siblings models and data in Economics: Beginning of a survey”, *Journal of Political Economy* 87(5), 537-564.

Dehejia, R. R. and S. Wahba (1998), “Propensity Score Matching Methods for Nonexperimental Causal Studies”, NBER Working Paper Nº 6829.

Imbens, G.W. and J.D. Angrist (1994), “Identification and Estimation of Local Average Treatment Effects”. *Econometrica* 62, 467-476.

Heckman, J.J. and R. Robb (1985), “Alternative Methods for Evaluating the Impact of Interventions” in *Longitudinal analysis of Labor Market Data*, ed. J.J. Heckman and B. Singer, New York. Cambridge University Press, 156-245.

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998), “Characterizing Selection Bias Using Experimental Data”. *Econometrica*, 66(5), 1017-1098.

Heckman, J., H. Ichimura and P. Todd (1997), “Matching as an Econometric Evaluation Estimator”. *Review of Economics Studies*, 65(2), 261-294.

Mc Closkey, D. (1987), “Counterfactuals” in Eatwell J. *et al.* Editors *The New Palgrave. A Dictionary of Economics.* 701-703.

Rosenbaum, P.R. and D.R. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects". *Biometrica* 70, 41-45.

Salvat ed. (1976), *Enciclopedia Salvat Diccionario*. Barcelona.

Wooldridge, J. (2002), *Econometric analysis of Cross Section and Panel Data*. The MIT Press Cambridge, Massachusetts. 604-606.