

# Decision-oriented statistical evaluation

*Valutazione orientata alle decisioni statistiche*

**Bruno Chiandotto, Fabrizia Mealli**

Dipartimento di Statistica “G. Parenti”

Viale Morgagni, 59 Firenze

chiandot@ds.unifi.it, mealli@ds.unifi.it

**Riassunto:** Il lavoro tratta dei possibili diversi obiettivi di un’analisi valutativa, evidenziando i momenti in cui è necessario l’esplicito coinvolgimento del decisore (pubblico) con la propria struttura di preferenze. In tale contesto, si illustra il contributo che l’analisi statistica può fornire nella risoluzione di problemi decisionali basati su processi di valutazione (ex-ante ed ex-post) facendo specifici riferimenti ad alcune scelte proprie del sistema formativo universitario.

**Keywords:** decision analysis, program evaluation, utility theory, effectiveness.

## 1. Introduction

Over the last decades Evaluation has been the focus of many papers in various fields, ranging from Economics and Econometrics (Heckman et al., 1999), Statistics (Holland, 1986), to Psychology, Political Science (Meny and Thoenig, 1991), and Law. In a very broad sense, Evaluation is seen as a multidisciplinary and integrated approach to produce evidence on systems that can serve to a) understand how the systems work; b) to assess whether certain objectives have been met; c) to promote and support evidence-based policy choices; d) to draw “scenarios” to help choosing among alternative new policies.

So, a fundamental objective of evaluation is to provide policy makers with useful information for choosing policies. In this paper we are going to restrict the attention to the so-called impact Evaluation of policies from a statistical point of view, bringing to the discussion Evaluation examples regarding Educational systems.

We will focus on the role of the decision maker in guiding the ex-post analysis by defining the subject of the evaluation as well as in analyzing the result of the evaluation. Formal decision analysis is also, and especially, needed in the ex-ante evaluation of alternative interventions, because evaluation has to take into account the policymaker’s attitude towards inequality and uncertainty in an appropriate way.

We will show that a proper decision-oriented statistical evaluation can be achieved not only if appropriate statistical tools are implemented for the production of evidence from available data, but also if those are used together with tools that enable to elicit individuals and policy makers preferences, in an interdisciplinary approach.

In particular, in the paper we would like to focus on specific questions which characterize impact evaluation analysis. Subject to impact evaluation are usually specific policies or interventions aimed at modifying in a certain desired way the condition or behavior of participants.

Within Educational system there are a variety of aspects that can be subject to evaluation in a very wide sense. Examples include the evaluation of the quality of

teaching, the evaluation of Educational systems' (internal and external) efficiency and effectiveness (Bini and Chiandotto, 2003). Some typical examples of more specific policies subject to impact evaluation are student aid policies, such as University grants, whose main objective is to guarantee higher education to motivated students irrespective of their income; the comparison of different teaching methods, whose main objective is to identify the method that is most effective in helping the students perform well. Another problem that can be cast in this framework is that of the evaluation of student orientation policies, whose objective is to provide students with information that can help them choose the career that mostly matches their preferences and attitudes, thus reducing drop out rates and duration of the studies and so improving the efficiency of the whole system.

All these problems of empirical research can be described as problems of causal inference, because they seek to estimate the effects of certain policies/interventions. It is usually a very difficult and often controversial analysis (Holland and Rubin, 1983).

Problems of causal inference involve "what if" statements and thus counterfactual outcomes and are usually motivated by policy concerns; they can be "translated" into a treatment-control situation typical of the experimental framework.

The fact the treatment is not assigned at random<sup>1</sup> prevents the possibility of comparing "treated" and "untreated" individuals, as such comparison is very unlikely to have a causal interpretation, because the two groups are different irrespective of their treatment status.

A growing list of papers in the statistical and economic literature (e.g., Rosenbaum and Rubin, 1983; Manski, 1990; Angrist et al., 1996; Heckman and Smith, 1998; Pearl, 2000) have tried to identify causal effects of interventions from observational, i.e. non experimental, studies, although using the conceptual framework of randomised experiments and the so-called *potential outcomes* approach, that allows to translate causal problems into a statistical model. Some identification strategies for causal effects can be found even in non experimental settings: as we will see, data alone do not suffice to identify treatment effects, but suitable assumptions, possibly containing prior information available to the researchers, are always needed.

The potential outcomes approach to causal inference, that we will use in the sequel, is rooted in the statistical work on randomized experiments by Fisher and Neyman, as extended by Rubin (1974, 1990) and subsequently by others. This perspective was called *Rubin's Causal Model* (RCM) by Holland (1986) because it views causal inference as a problem of missing data with explicit mathematical modeling of the assignment mechanism as a process for revealing the observed data<sup>2</sup>.

---

<sup>1</sup> This is the usual case in the social sciences where only observational data are available. Social experiments are still relatively rare, although they have been designed and implemented especially in North America: despite the ethical problems they may arise, they could be easily used for example to assess the effectiveness of different orientation policies, by selecting at random small groups of students.

<sup>2</sup> Another approach to causal modeling is based on the theory of causal graphs. Recent development in this field, which has many links with structural equation modeling, has provided a formal theory for evaluating causal effects (Pearl, 2000). The use of graphs provides a tool for integrating statistical and subject-matter information: it facilitates to explicate the assumptions underlying the causal model and to decide whether the assumptions are sufficient for obtaining estimates of causal effects from observed quantities. The two formalizations have many links (see for example Pearl (2000) for a discussion on this).

The starting essential feature of this approach is to define a causal effect as the comparison of the potential outcomes on the same unit measured at the same time. Consider a binary treatment:  $Y(0)$  is the value of the outcome variable  $Y$  if the unit is exposed to treatment  $T = 0$ , and  $Y(1)$  is the value of  $Y$  if exposed to treatment  $T = 1$ ; this same notation can be extended to the case treatment levels are more than two. Only one of these potential outcomes can be observed, the one corresponding to the treatment the unit received, yet causal effects are defined by their comparison, e.g.,  $Y(1) - Y(0)$  or  $Y(1)/Y(0)$ . Thus, causal inference becomes a problem of inference with missing data, because only one of the two potential outcomes is observed, the other one becoming a counterfactual. Also, causal effects defined as comparisons of potential outcomes force to think of a causal effect of a treatment only if the treatment status can be thought of as having been manipulated in some way. The focus of impact analysis is usually that of estimating a feature of the distribution of  $Y(1)-Y(0)$ , e.g., the average treatment effect  $ATE = E(Y(1) - Y(0))$  or the average treatment effect for subpopulations of individuals defined by the value of some variable, most notably the subpopulation of the treated individuals:  $ATT = E(Y(1) - Y(0) \mid T = 1)$ . Alternatively, the focus can involve only comparisons of the two marginal distributions of  $Y(0)$  and  $Y(1)$ .

In the following sections we will show how an Evaluation process can be thought of as a decision problem, where the results of a causal analysis are used to judge the effects of some interventions. We will distinguish ex-post (section 2) from ex-ante evaluation of policy interventions that have not yet been implemented (section 3).

## 2. Evaluation as a decision problem

Even in the preliminary step of the analysis, formal decision making tools are needed to guide the Evaluation process. First of all, the decision maker should inform the analyst on which outcome variable to choose, i.e., he should clearly identify what is the condition or behavior that the intervention aimed at modifying in a desired direction. This is usually not an easy task; the analyst usually has to cope with laws, where the policy objectives are only vaguely specified. More than formal methods for preferences elicitation, such as those for eliciting utility functions, there is the need of questions and explicit answers by decision makers. With this respect, policy makers can be aided, as usually done for the elicitation (often hierarchical) of criteria to be included in multi-attribute utility functions and multi-criteria decision making (Bana e Costa, 2001, Keeney et al., 1990), where an informative discussion between the analyst and the decision maker, can help both find the most appropriate outcome variables to focus on.

Secondly, restricting the analysis to average treatment effects (ATE), even if sometimes considered also for subsets of individuals defined on observed pre-treatment characteristics, does not usually reflect what the decision maker would like to know.

In order to illustrate this point, let us consider the following Higher Education related example.

The agency that administrates students' aids for the University of Florence (Azienda per il Diritto allo Studio) offers every year some scholarships to eligible freshmen, where eligibility is based on merit and economic needs. The main objective of this intervention is to give equal opportunity to achieve higher education to motivated students irrespective of their income. As grants' award is usually constrained by a fixed budget, not all eligible students who apply for a grant can receive one: so the applicants are

ranked on the basis of an economic indicator  $S$ , that depends in a deterministic way on family income, real and personal assets and family structure, and only a varying percentage of high ranking applicants with  $S$  lower than  $s_0$  receive a grant. In agreement with the objective of the intervention, the decision maker should first be interested in knowing whether those receiving a grant can reach the same education level as students that are equally motivated but belong to richer families. Motivation is usually a variable that is difficult to measure, so the analysis can be diverted to explore whether the grant is an effective tool to prevent good students from lower-income families from dropping out of higher education and to achieve a degree in a reasonable time period: indeed, low income students are characterized by a high drop out rate, as well as a long duration of their studies, because of their usual need to work while studying.

The analyst must first answer the following question: under which conditions are the distributions of  $Y(1)$  and  $Y(0)$  identified? In this case  $Y(0)$  and  $Y(1)$  are either the binary variable taking value 1 if the student drops out and 0 otherwise or the continuous variable representing time to degree achievement,  $T$  is the treatment indicator being 1 if the student receives the grant and 0 otherwise. In order to answer the question one must make assumptions on the assignment mechanism. The assignment mechanism is a stochastic rule for assigning treatments to units and thereby for revealing  $Y(0)$  or  $Y(1)$  for each unit: it can depend on other measurements, i.e.  $P(T = 1|Y(0), Y(1), X)$ ; if these other measurements are observed values, then the assignment mechanism is ignorable; if, given observed values, it involves missing values, possibly even missing  $Y$ 's, then it is non ignorable. Unconfoundedness is a special case of ignorable missing mechanism and holds when  $P(T = 1|Y(0), Y(1), X) = P(T = 1|X)$ , where  $X$  is a vector of observed pre-treatment variables; basically this assumption says that exposure to treatment is random within cells defined by the variables  $X$ ; matching methods rely on this assumption. Although very strong, the plausibility of this assumptions heavily relies on the amount and on the quality of the information on the individuals contained in  $X$ .

In the case of the evaluation of grants' effect, the study is a so-called sharp regression-discontinuity design (Thistlethwaite and Campbell, 1960), where  $T$  is known to depend in a deterministic way on some observable variable  $S$  (the economic indicator in our case)  $T = f(S)$ , where  $S$  takes on a continuum of values and the points  $s_0$  where the function  $f$  is discontinuous are assumed to be known. In this case the following identifying condition can be exploited:  $P(T = 1|Y(0), Y(1), X, S = s_0) = P(T = 1|X, S = s_0)$ . The intuition is that, in the neighborhood of  $s_0$ , the grant status  $T$  and the potential outcomes are independent, so that the sample of individuals within a very small interval around each cutoff point, who have essentially the same  $S$  value, are very similar to a randomized experiment at the cutoff value (Mealli and Rampichini, 2001). This condition, if accepted to be plausible, allows to identify the distribution of the two potential outcomes for students having an economic indicator value equal to the cutoff point  $s_0$  (Hahn et al., 2000). Note that, whatever model we apply to the data, evidence on the causal effects is contained only around the cut-off: if we extend the evidence for other subgroups of people (for example having different values of  $S$ ), we make an improper use of extrapolation, and the reliability of such causal effects' estimates cannot be assessed<sup>3</sup>. Usually, only the ATE at the cut-off point is estimated, i.e., in this case,  $E(Y(1)-Y(0)|X, S = s_0)$ ; as pointed out by Dehejia (2004) and others, focusing the

---

<sup>3</sup> The same argument can be applied when regression models for the estimation of causal effects under unconfoundedness are used, without previously checking if the distribution of the covariates in the two groups of treatments overlap to a large extent or not.

attention on ATE does not always solve the decision problem of assessing the two treatments.

An appropriate alternative is to compare the distribution of the potential outcomes by formal decision theory methods. The distributions in the outcome variables space should contain all the uncertainty from the model as well as from the parameter estimation: according to the Bayesian approach, a distribution with these characteristics is the posterior predictive distribution of  $Y(0)$  and  $Y(1)$ , i.e.,  $f(Y(0)|data)$  and  $f(Y(1)|data)$ . In order to derive the posterior predictive distribution one has to specify a statistical model for the outcome variables (see for a very simple example Gelman et al., 1995, ch. 7)  $f(Y(t) | \theta)$ , and elicit a priori distributions of the parameters characterizing the model,  $\pi(\theta)$ , so that  $f(Y(t) | data) = \int f(Y(t) | \theta) \pi(\theta | data) d\theta$ . The model must be consistent with the assumptions on the assignment mechanism that allow identification of causal parameters, because the model will be able to capture uncertainty correctly only if correctly specified<sup>4</sup>.  $Y$  can be multivariate (e.g., it may include drop-out and academic performance) the analyst must be very careful in specifying the model, as some variables in  $Y$  might be “intermediate” with respect to others so that a simple multivariate model might not be appropriated, as will be more clear in subsequent examples.

In order to judge whether treatment is effective for individuals with certain characteristics  $X$ , one should first elicit decision makers preferences with respect to  $Y$ , i.e., specify a utility function, that expresses policy makers risk attitudes and preferences over uncertain values of  $Y$ . Recent developments in methods for utility elicitation can be used and applied (Fennema and Van Assen, 1999). When  $Y$  is multivariate, multi-attribute utility functions should be elicited, with explicit consideration of the trade-offs between attributes (see French, 1988, for a deep discussion on these issues, and on the implicit assumptions of commonly used additive utility functions).

Once a utility function is elicited, treatments should then be judged comparing  $E(U(Y(1)))$  into  $E(U(Y(0)))$ . As it is well known, and shown in numerous examples, in cases where ATE cannot “distinguish” between the two treatments, expected utility comparison can lead to clearcut decisions.

Another more informative alternative to ATE is looking at other feature of the distribution of potential outcome differences, for example, the decision maker might be interested in knowing the proportion of individuals who have benefited from the intervention, i.e.,  $P(Y(1)-Y(0)>0)$ ; equal values of ATE can be generated by very different values of this proportion. The estimation of this proportion requires the knowledge of the *joint* distribution of potential outcomes, that depends on the correlation structure of potential outcomes on which we do not have direct evidence.

The data contain only limited information on the joint distribution but, under suitable assumptions, bounds on this quantity can be derived, as explained with the next example.

Consider again the problem of evaluating the impact of university grants on time to graduation; this variable is not observed on everybody. Indeed duration is only observed and defined for students who do not drop-out: thus a properly defined causal effect on

---

<sup>4</sup> Recent advances in Bayesian model choice and model checking have also been applied in the specific context of causal inference; see, among others, Hirano et al. (2000) and Mealli et al. (2004). Posterior predictive distributions can be easily approximated by simulations if parameters’ posterior distributions are obtained via MCMC methods.

duration is the effect of the grant on the subset of students who would not drop-out irrespective of receiving a grant or not. The analysis should first assess the effect on drop-out and then, once the group of students who do not drop-out irrespective of receiving the grant is *identified*, the effect on duration can be assessed. In order to simultaneously consider the two phenomena, we can adopt the principal strata framework (Frangakis and Rubin, 1999), that allows to properly account for intermediate outcomes, such as drop-out in this case, and unobservables.

The *intermediate variable*  $D_i(T_i)$  is 1 or 0 if student  $i$  drops out or does not drop out, when receiving treatment  $T_i$ . The *response variable*  $Y_i(T_i)$  the duration of the studies of student  $i$  when receiving treatment  $T_i$ . Since, again, for each individual the treatment variable assumes a single value, for every post-treatment variable only one of the two potential versions can be observed; because both the treatment variable and the intermediate variable are dichotomous, four *principal strata* can be defined in the following way through the latent variable  $L_i$ :  $L_i = \text{'DD'}$  (Dropout, Dropout) if  $D_i(1)=1$  and  $D_i(0)=1$ ;  $L_i = \text{'DN'}$  (Dropout, No dropout) if  $D_i(1)=1$  and  $D_i(0)=0$ ;  $L_i = \text{'ND'}$  (No dropout, Dropout) if  $D_i(1)=0$  and  $D_i(0)=1$ ;  $L_i = \text{'NN'}$  (No dropout, No dropout) if  $D_i(1)=0$  and  $D_i(0)=0$ . The principal strata are latent classes and the data allow to put bounds or even estimate, under certain assumptions, the probability that a given student belongs to a certain latent class. If the purpose is to evaluate the effect of the grant on drop-out, note that the probability of belonging to a principal strata answers the question about the proportion of individuals who have benefited from the intervention, that is  $P(L_i = \text{'ND'}) = \pi_{ND:i}$ . Also note that the probabilities of the principal strata,  $\pi_{ND:i}$ ,  $\pi_{DD:i}$ ,  $\pi_{DN:i}$ ,  $\pi_{NN:i}$ , are interesting per se, as they throw light on how the average effect has been achieved. In fact, the average causal effect on dropout is

$$P(D_i(1) = 1) - P(D_i(0) = 1) = (\pi_{DD:i} + \pi_{DN:i}) - (\pi_{ND:i} + \pi_{DD:i}) = \pi_{DN:i} - \pi_{ND:i}$$

Therefore, similar values of ATE can be produced by very different proportions of students who benefit or are “harmed” by the interventions. Decision maker preferences on these proportions can be elicited in order to judge an intervention.

We can put bounds on the probabilities for the principal strata following the procedure used in Zhang and Rubin (2004). The bounds can be sharpened by making some suitable assumptions (see also Grilli and Mealli, 2004).

If the purpose is to evaluate the impact of the grant on the duration of the studies, the response variable  $Y$  is defined only for the non dropouts, so the causal effect  $Y_i(1)-Y_i(0)$  is properly defined only for the NN stratum, i.e. the students who graduate with or without the grant. Again, bounds on this causal effect can be derived; more informative results, can be obtained by specifying suitable parametric models that include some identifying conditions, to be fitted with likelihood or Bayesian methods.

Note that these issues just presented are usually neglected in the literature of measuring the amount of human capital, by means of multilevel models. Also, the same approach can be applied to jointly analyze internal and external effectiveness (with respect to the labor market) of different degree programs as in Grilli and Mealli (2004).

### 3. Decision-oriented (ex-ante) evaluation

The usual goal of ex-ante evaluation is to perform an evaluations of policy interventions that have not yet been implemented. Such policies may well be variations of previously adopted interventions. For example, as far as the problem of distributing university grants, the policy maker might be interested in changing the criteria that identify the eligible students, or change the amount of the grants, or change the tool of student aid policies, for example offering *goods* (accommodation, meals, books) instead of a money transfer.

Interventions should be evaluated using social welfare functions reflecting the individual preferences, embedded in their utility functions, as well as the policy maker preferences, e.g. his degree of inequality aversion.

Such issues have been pointed out and analyzed by economists who raised the issue of making informed public choices, but are usually neglected in the evaluation literature. Exceptions include Manski (1990) who develops nonparametric bounds for the expected welfare from different post-evaluation assignment rules; Mealli and Pudney (1994) that analyze the cost-effectiveness of youth training funding for a wide range of the social evaluation of successful employment, also deriving bounds on optimal policy welfare; and Heckman and Smith (1998) who consider the evaluation of various welfare functions, some of which requiring information on the joint distribution of potential outcomes, as seen in the previous section.

Consider the following decision problem faced by a policy maker, which is a relatively standard model of welfare Economics: let the potential outcome(s) in the presence of treatment (policy)  $k$  for individual  $i$  be  $Y_{ki}$ , and let the preference of person  $i$  for outcome  $Y$  be denoted by  $U_i(Y)$ . It is postulated that a social welfare function exists<sup>5</sup> that is defined over utilities of the  $N$  individuals of a society or a target population (e.g. all the students enrolled in a given University):

$$W(k) = W(U_1(Y_{k1}), U_2(Y_{k2}), \dots, U_N(Y_{kN})) \quad (1)$$

A choice based on such a function would choose the treatment  $k$  with the highest value of  $W$  or, in case of uncertainty, the highest expected welfare<sup>6</sup>.

There are several issues that need to be solved before the above criterion can become operational; because of the difficulty and subjectivity of some of them, social welfare functions usually do not govern public decision making. Nevertheless, we believe that in more restricted problems (such as that of the distribution of grants among students) they can help making informed decision. These issues are the topic of next subsections.

### 3.1 How do we elicit preferences?

In principle the same procedures and tools, quoted in the previous section, could be applied to elicit individuals preferences  $U_i$ . These include also recent advances in Multi

---

<sup>5</sup> Here we do not discuss the issue related to Arrow's impossibility theorem, as well as to the problem of making interpersonal comparison. For a formal treatment of the issue, where assumptions on the decision maker preferences are made explicit, see Deaton and Muellbauer (1980) and French (1988).

<sup>6</sup> As in the ex-post analysis, we could consider here also criteria that involve the knowledge of the joint distribution of potential outcome; for example policy  $k$  could be preferred to policy  $j$  if the proportion of people benefiting from  $k$  is greater than 0.5. We limit our attention on criteria that requires, ex-ante, only the marginal distributions of outcomes under different policies (Dehejia, 2004).

Criteria Decision Making such as in Bana e Costa and Chagas (2004), with interactive tools as aid for preferences elicitations, that would very well match the needs of some program participants, such as students. Despite that, we believe that in most cases it would be almost impossible to elicit every individual  $i$ 's preferences, because of timing and costs. We can still think at possible solutions to this problem: for example we could elicit the preference of the modal, median or average individual, and use only his utility function. Alternatively we could assume preferences depend on observed characteristics, and use the estimation of discrete choice models from revealed preferences and *translate* them into a multiattribute utility function  $U_i(Y)$ .

The choice of a functional form for  $W$  and the specification of its arguments is usually rather complicated. Let us first assume (1) is indeed describing the policy maker decision problem; then, once the individual utility functions have been quantified, the choice of how to aggregate them depends on the decision-maker attitudes toward inequality. The following simple functional form can be used, that encompasses, depending on the value of  $\varepsilon$ , the inequality-neutral utilitarian preferences ( $\varepsilon = 0$ ), different degrees of inequality aversion ( $\varepsilon > 0$ ) to the leximin or Rawlsian preferences ( $\varepsilon = \infty$ ) that embodies an extreme concerns with the worst-off individual in the group (Deaton and Muelbauer, 1980):

$$W(k) = W(U_1(Y_{k1}), U_2(Y_{k2}), \dots, U_N(Y_{kN})) = \frac{\sum_h (U_h(Y_{kh}))^{1-\varepsilon}}{1-\varepsilon}.$$

Yet, policy makers preferences do not always match those of the individuals: decision makers may want to exclude (include) some variables in the  $Y$  vector that might be (not) relevant for the students. Also, even if the  $Y$  may be the same, the overall objective of the decision maker might not involve individual preferences and risk attitudes. If individual tastes are not explicitly taken into account but only the opportunity sets faced by different individuals with whom we are concerned, then to call  $W$  a social welfare function is misleading, since it is only an aggregate weighting functions:

$$W(k) = W(Y_{k1}, Y_{k2}, \dots, Y_{kN})$$

What to include in  $Y$  and how to aggregate the  $Y$ 's depends on the explicit objective of the decision maker. In the University grant example, the objective can be that of maximizing the number of students from low income families who get a degree, or maximizing the number of students who get a degree ( $L$ ) *and* reducing the time to get a degree ( $D$ ); in this second case  $W$  could be specified as follows  $Y_{kh} = \{D_{kh}, L_{kh}\}$  :

$$W(k) = \sum_h (\alpha D_{kh} + L_{kh})$$

where  $\alpha$  represents the trade-off between the two attributes.

In the presence of limited resources, the aim might be, not only that of ex-ante evaluating different interventions, but that of finding the “best” way of distributing resources. In this case the decision maker might be interested in exploiting treatment

effect heterogeneity, i.e., one kind of intervention may be effective for some people but not for others. This approach is usually called *profiling*, and is not always accepted, in particular it is not obvious to what extent observed characteristics can be used to choose between alternatives: the analyst may be precluded from using certain covariates (say gender, previous merit). Note that maximization of  $W(k)$  with respect to alternative distribution of resources is a problem that would require appropriate operational research tools. In Compagnino and Gori (1992) a similar approach was used in order to find an *optimal* allocation of university grants.

### 3.2 How can we derive predictions of the $Y_{ki}$ ?

The estimation of marginal distributions of the different  $Y_{ki}$  is a rather difficult task, especially if some of the alternative interventions under consideration have never been implemented before. A structural behavioral model must be specified and estimated from available data; the model should be able to predict the modifications of behavior or conditions induced by the different interventions or the different ways of making the intervention available in the future. For example, in the university grant problem, suppose grants have never been assigned to students before; thus we would need a model where students behavior (for example their academic performance as well as their decision to continue their study and achieve a university degree) are linked to some target variables, in this case their income, or the income of their family, because the grant can indeed be thought of as a money transfer that is changing their endowment.

The model can be specified at different degree of complexity, also depending on the availability of data. A simple model would be one for the probability of getting a degree, or the duration of the studies, where this depends on the student's characteristics, in particular merit and income (or some equivalent scale for it), and student choices, for example the faculty of enrolment.

A more complex structural model would be a dynamic behavioral model (Wolpin, 1996) in which each student maximizes his/her lifetime earnings and takes schooling decision, given an outside wage and some intertemporal discounting factor, where the disutility of loss income today is higher for the poor than for the rich. This model would be a lot richer than the previous one, but would also require more assumptions and more information (for example those on wages and on the returns to schooling) that are not always readily available.

If grants have already been assigned in the past, one can use the results of the ex-post analysis as a benchmark for the ex-ante analysis. If the structural model is able to reproduce the ex-post results up to a reasonable approximation, there is hope it to be appropriate for simulating the effects of alternative programs: Attanasio and Meghir (2001), Borghignon et al. (2002), Todd and Wolpin (2003) are examples on how this comparison can be made.

An implicit assumption that is usually made here is that results obtained for one population can be extrapolate to other populations and/or to other time period; recently this issue has been discussed in Hotz et al. (2004). Also general equilibrium effect are usually neglected, i.e., we can define the potential outcomes of an individual irrespective of the (level of the) treatments the others receive. This assumption is usually called SUTVA (Stable Unit Treatment Value Assignment) in the statistical literature.

Note that the output of such models, can be used also to inform individual decisions. For example, we could provide students with predictions on their academic performance and labor market performance, given their observed characteristics, with different degree programs, to help them choose the one they mostly prefer. This information can form the basis for a formal multi-criteria career choice problem (Bana e Costa and Chagas, 2004). This issue was discussed in Goracci (2000) where an optimal allocation of students among the various degree programs was found, optimal in the sense of maximizing a welfare function that depends on the estimated duration of the studies and on the labor market opportunities of each student given his/her characteristics.

As said in section 2, we believe that taking a Bayesian perspective to decision analysis, rather than frequentist one (as in Manski, 2003) allows to take account of all the uncertainty in a simple way by using the predictive distribution, now of the single  $Y_{ki}$ 's. If  $W(k)$  includes individual preference  $U_i$ , then we can first substitute the  $U_i$  with their certain equivalents, and apply the chosen  $W(k)$  in order to derive the expected welfare from policy  $k$ . If instead  $W(k)$  does not include individual preferences, then we can compute the posterior predictive distribution for  $W$ , by drawing from each individual predictive distribution of  $Y_{ki}$ , and then summarize it by its expected value.

#### 4. Concluding remarks

The main objective of Evaluation is to provide policy makers with useful information for choosing among policies. In this paper, attention was restricted to the so-called impact Evaluation of policies. We have shown the role of the decision maker in conducting both ex-post and ex-ante analysis. Formal decision analysis is needed to properly account for individuals' and policymakers' attitude towards uncertainty and inequality. Despite the fact that a rigorous Evaluation can only be implemented if appropriate statistical tools are used, the paper shows that Evaluation will benefit also from tools made available from other disciplines, such as Economics and Operational Research, that allow to elicit individuals and policy makers preferences and judge interventions.

#### References

- Angrist J., Imbens G., Rubin D.B (1996) Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, 91, 444-472.
- Attanasio O. , Meghir C. & Santiago A. (2001) Education Choices in Mexico: Using a structural Model and a Randomized Experiment to Evaluate PROGRESA, mimeo.
- Bana e Costa C.A. (2001) The use of multi-criteria decision analysis to support the search for less conflicting policy options in a multi-actor context: case study, *Journal of Multi-Criteria Decision Analysis*, 10, 2, 111-125.
- Bana e Costa C.A. & Chagas M.P. (2004) A career choice problem: an example of how to use MACBETH to build a quantitative value model based on qualitative value judgements, *European Journal of Operational Research*, 153, 2, 323-331.
- Bini M. & Chiandotto B. (2003) La valutazione del sistema universitario italiano alla luce della riforma dei cicli e degli ordinamenti didattici, *Studi e Note di Economia*, 2, 29-61.

- Bourguignon F., Ferriera F.H. & Leite P. (2002) Ex-ante evaluation of conditional cash transfer programs: the case of Bolsa Escola, World Bank Policy Research Working Paper n. 2916, The World Bank, Washington D.C.
- Bleichrodt H. Pinto J.L. & Wakker P. (2001) Making descriptive use of prospect theory to improve the prescriptive use of expected utility in [www.fee.uva.nl/creed/wakker/pcf/corrsg.pdf](http://www.fee.uva.nl/creed/wakker/pcf/corrsg.pdf).
- Chiandotto B. (2002) Valutazione dei processi formativi: cosa, come e perché, in Valutazione della didattica e dei servizi nel sistema Università, *Atti della giornata di Studio, Fisciano*, CUSL, Salerno.
- Compagnino A. & Gori E. (1992) *Il controllo di gestione degli enti per il diritto allo studio universitario. Efficienza ed efficacia*, Franco Angeli, Milano.
- Deaton A. & Muelbauer J. (1980) *Economics and consumer behavior*, Cambridge University Press, Cambridge.
- Dehejia R.H. & Wahba S. (1999) Causal effects in nonexperimental studies: reevaluating the evaluation of training programs, *Journal of the American Statistical Association*, 94, 1053-1062.
- Dehejia R.H., (2004) Program Evaluation as a Decision Problem, forthcoming, *Journal of Econometrics*.
- Fennema H. & Van Assen M. (1999) Measuring the utility of losses by means of the tradeoff method in *Journal of Risk and Uncertainty*, Vol. 17.
- Frankgakis C. E. & Rubin, D. B. (2002) Principal Stratification in Causal Inference, *Biometrics*, 58, 21-29.
- French S. (1988) *Decision Theory*, Ellis Horwood Limited, UK.
- Gelman A., Carlin J.B., Stern H.S. & Rubin D.B. (1995) *Bayesian Data Analysis*, Chapman & Hall, London, UK.
- Goracci L. (2000) *Scelte educative e sbocchi occupazionali degli studenti universitari: un'applicazione ai dati di Almalaurea*, Tesi di laurea, Corso di Laurea in Scienze Statistiche ed Economiche, Università di Firenze.
- Grilli L. & Mealli F. (2004) Analysis of the Effectiveness of Degree Programmes by means of Principal Stratification, mimeo.
- Hahn J., Todd P. & Van der Klaauw W. (2001) Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica*, 69, 1, 201-209.
- Heckman J.J. (1989) Causal Inference and nonrandom samples, *Journal of Educational Statistics*, 14, 159-168.
- Heckman J.J. & Hotz V.J. (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training, *Journal of the American Statistical Association*, 84, 408, 862-874.
- Heckman J. & Smith J. (1998) Evaluating the Welfare State. In Strom, S. (Ed.), *Econometrics in the 20th Century: The Ragnar Frisch Centenary*, Cambridge University Press, Econometric Society Monograph Series, Cambridge.
- Heckman J.J., Lalonde R. & Smith, J. (1999) The economics and econometrics of active labour market programs, in Ashenfelter O. e Card D. (eds), *Handbook of Labor Economics*, vol. 3a, North Holland, Amsterdam.
- Hirano K., Imbens G.W., Rubin D.B. & Zhou X. (2000) Assessing the effect of an influenza vaccine in an encouragement design, *Biostatistics*, 1, 69-88.
- Holland P. (1986) Statistics and Causal Inference, *Journal of the American Statistical Association*, 81, 945-970.

- Holland P.W. & Rubin D.B. (1983) On Lord's paradox. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement*, Hillsdale, NJ: Erlbaum.
- Hotz J., Imbens G.W. & Mortimer J. (2004) Predicting the Efficacy of Future Training Programs Using Past Experiences, *Journal of Econometrics*, forthcoming.
- Manski C.F. (1990) Nonparametric Bounds on Treatment Effects, *American Economic Review Papers and Proceedings*, 80, 319-323.
- Manski C.F. (2000) Identification Problems and Decisions under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice, *Journal of Econometrics*, 95, 415-442.
- Manski C.F. (2004) Statistical Treatment Rules for Heterogeneous Populations: with Application to Randomized Experiments, *Econometrica*, forthcoming.
- Mealli F., Pudney S. & Thomas J. (1994) Youth employment and the optimal structure of Youth Training: an econometric analysis, Discussion Papers in Economics n. 94/14, University of Leicester.
- Mealli F. & Rampichini C. (2001) Impact evaluation of University grants, *JSM, New York, Proceedings*.
- Mealli F. & Rubin D.B. (2003) Commentary: Assumptions allowing the estimation of direct causal effects, *Journal of Econometrics*, 112, 79-87.
- Mealli F., Imbens G.W., Ferro S. & Biggeri A. (2004) Analyzing a randomized experiments for breast self-examination with noncompliance and missing outcomes, *Biostatistics*, forthcoming.
- Meny Y. & Thoenig J.C. (1991), *Le politiche pubbliche*, Il Mulino, Bologna.
- Keefer D.L., Kirkwood C.W. & Corner J.L. (2004) Perspective on Decision Analysis, Applications, 1990-2001, *Decision Analysis*, 1, 5-24.
- Keeney R.L., von Winterfeldt D. & Eppel T. (1990), Eliciting values for complex policy decisions, *Management Science*, 36, 9, 1011-1030.
- Pearl J. (2000), *Causality*, Cambridge University Press.
- Rosenbaum P. & Rubin D.B. (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55.
- Rubin D.B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 668-701.
- Rubin D.B. (1990) Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science* 5, 472-480.
- Thistlethwaite D. & Campbell D. (1960) Regression-discontinuity analysis: an alternative to the ex-post facto experiment, *Journal of Educational Psychology*, 51, 309-317.
- Todd P. & Wolpin K.I. (2003) Using a social experiment to validate a dynamic behavioral model of child schooling and fertility: assessing the impact of school subsidy in Mexico, mimeo.
- Tsoukias A. (2003) From decision theory to decision aiding methodology, DIMACS Technical report 2003-21.
- Wolpin K.I. (1996) Public-Policy Uses of Discrete-Choice Dynamic Programming Models, *American Economic Review, Papers and Proceedings*, 86,2, 427-432.
- Zhang J. & Rubin D. B. (2004) Censoring due to death via principal stratification, *Journal of Educational and Behavioral Statistics*, forthcoming.
- Zhang J., Rubin D. B. & Mealli F. (2004) Evaluating the Effects of Training Programs on Wages with Experimental Data, mimeo.