Chapter 1

An Introduction to Evaluation of Social Programs

1.1 Introduction

A typical description of the phenomenon of *Randomness* can be stated as follows:

> *A phenomenon can be called as random if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.*

The existence of randomness is a crucial feature on making valid statistical inference and actually constitutes the basis in applying conventional statistical methods for research purposes. Most of the literature in distribution, sampling and statistical analysis theory assumes *random samples* of individuals in order to study particular phenomena. A random sample has a specific meaning in statistical theory. One does not automatically get a random sample. A random sample of size n is defined as the result of a sampling process from a population of size N where each unit 1, 2, …, N has the same probability of being chosen.

In several circumstances, though, cost or other external factors do not allow the analyst to design the necessary "randomized background" for the analysis so that data collection is not any longer based on random sampling procedures. In such cases, where only nonrandom samples may be available, conventional statistical methods of data analysis, e.g. simple Ordinary Least Squares (OLS), yield biased estimates of the parameters of interest. Even then however, *inference from nonrandom samples* can be extrapolated by more sophisticated methods.

1.2 The Research Framework

The present thesis is focussed on describing various methods of making inference from nonrandom samples under the *model of Causal inference*, which is of primary interest. *Causal inference* is a field of juxtapositions among several authors who all, at least, recognize its principal difference from the so-called *Associational inference model*.

Associational inference consists of making statistical inference (estimates, tests, posterior distributions etc) about the associational parameters relating variable Y (a particular outcome) and X (individuals' attributes) on the basis of data gathered about Y and X for a number of units i ($i = 1, \ldots, N$). Population of units exists within a time frame of some sort, and the measurements of characteristics of units that represent variables must also be made at particular times. In this sense, associational inference is simply descriptive statistics where the role of time is simply to affect the definition of the population of units or to specify the operational meaning of a particular variable.

The model of *causal inference* also begins with a population of units i. Units in this model are the objects of study on which causes (treatments) may act. Exposure to a treatment has always some effect to the unit, which is potentially exposable to anyone of the causes. This effect is always relative to another treatment (cause). For example, Holland (1986) discusses that the phrase "A causes B" almost always means that A causes B relative to some other cause that includes the condition "not A". Glymour (1986) interprets the above phrase rather differently as "A is a cause of B".

To cast the discussion in a more practical base, let us consider a specific example. Suppose that there are two causes: treatment, denoted by 1, and control, denoted by 0. Variable S indicates the cause to which each unit is exposed and is analogous to variable X (attributes), but with the essential difference that $S(i)$ indicates exposure of i to a specific cause, whereas $X(i)$ indicate a property or a characteristic of i. Holland (1986) indicates that an attribute cannot be a cause in an experiment since an individual cannot be exposed to the attribute. Rubin (1986) criticizes Holland's opinion to consider causal statements based on attributes as meaningless. Holland (1986, rejoinder) corrects his previous statements by explaining that in the case of attributable characteristics, causal effects are not well defined.

Granger (1986) and Glymour (1986) disagree with Holland's (1986) opinion on the difference between causes and attributes. They comment that questions such as "race affects crime rates" and "death sentence causes decreases in murder rates" are on the same causal footing. Nevertheless, Holland (1986, rejoinder) indicates that "race" is an attribute and is not manipulated by anyone, while "death sentence" is clearly a causal statement of great public policy interest and thus can be manipulated by governors and legislators. Thus, the first must be viewed as associational statement and analyzed differently from the second.

Upon these findings, Freedman (1997) states the difficulties the analysts are usually faced up with when trying to define causality in several applications. By performing *path analysis techniques* to establish causality, he concludes that "…the models tend to neglect the difficulties in establishing causal relations and the mathematical complexities tend to obscure rather than clarify the assumptions on which the analysis is based…" and "…Even so, causality is assumed into the picture at the beginning, not proved at the end…".

The role of time in causal inference is very important because, when a unit is exposed to a cause, this must occur within a specific time period. Under this perspective, variables can be divided in two classes: pre-exposure (prior to exposure to a cause) and post-exposure (after exposure to a cause). Variable Y_i falls into the post exposure class and measures the effect of the cause. For example, we say that:

$$\textit{Status 1 causes the effect } Y_{1i} - Y_{0i}$$

Y_{Di}*'s*, $D = 0, 1$, are perceived as specific post-exposure outcomes, while their difference indicates the related *benefit*.

Lindley (2002) explains the basic concepts of causality, including some comments on their relevance to inference and decision-making. Dawid (2002) considers a variety of ways in which causal models can be represented in graphical form. By adding nodes to the graphs to represent parameters, he supports that meaningful causal modeling and inference can be obtained.

In the chapters that follow, we mainly analyze various aspects connected with causal inference from nonrandom samples and the problems that it entails. Several methods of statistics and econometrics that deal with these problems are reviewed. A critical feature in the analysis is the *Stable–Unit–Treatment–Value Assumption (SUTVA)*, as labeled by Rubin (1986). Situations where this assumption does not hold cannot be analyzed using the methods described in this thesis. This assumption states that:

ASSUMPTION 1.1: Stable – Unit – Treatment – Value Assumption (SUTVA)

Consider the situation with N units indexed by i = 1, …, N; d treatments indexed by D =1, …, d; and outcome variable Y, whose possible values are represented by Y_{Di}. *SUTVA is simply the a priori assumption that the value of Y for unit i when exposed to treatment D will be the same no matter what mechanism is used to assign treatment D to unit i and no matter what treatments the other units receive, and this holds for all i = 1, …, N and all D = 1, …, d.*

Assumption (1.1) establishes the absence of interaction among the units. In particular, the values of the outcomes of the N participants are assumed to be stochastically independent across *i*. For this reason it is also known as "no interference among units" and as "absence of spillover effects".

1.3 Evaluating Social Programs with Nonrandom Samples

A modern field where this kind of analysis often takes place sets about the *evaluation of social programs*. We use the term "social program" with a rather general meaning. As a social program can be regarded any activity that is related to one of the following:

- *Participation in training programs for workers*: Assume that there are two potential states in the world of each individual, say 0 and 1, defined by the dummy variable D_i as:

$$D_i = \begin{cases} 0 & \text{if the individual participates in the training program} \\ 1 & \text{if the individual does not participate} \end{cases}$$

The workers are free to choose between the states (*self-selectivity*). To this extend, we treat Y_{1i} as the participants' outcomes (e.g. wages) and Y_{0i} as the non-participants' outcomes. In this framework the samples of persons that belong to states 0 and 1 respectively, cannot be thought to come from a random sampling design. In fact, this is a formal description of the process that generates the nonrandom (selective) samples and necessitates specific methods for causal inference. Equivalent statements can be considered for any of the following examples.

- *Participation in medical programs*, e.g. a new drug evaluation study.
- *College Educational Programs*, e.g. evaluation of college education.
- *Migration decision*, i.e. how migration affects the outcomes of migrants.
- *Labor Force Participation for a sample of individuals,* i.e. the effect of participation on individual outcomes.

The term "evaluation" entails the estimation of the benefits from participation to a program. In the simplest case these benefits are expressed by the difference in the above outcomes

$$\Delta_i = Y_{1i} - Y_{0i}$$

In this setting, participation can be thought as an exposure to a specific cause (e.g. training program) for a sample of N units. Thus, the analysis is closely related to the general framework of causal inference we have already set.

Several authors have worked on evaluation studies on the above fields. Among them are Heckman (1974), Gronau (1974), Lewis (1974), Heckman (1979), Heckman and Robb (1985a), Heckman (1990a, b), Maddala (1978, 1994), Heckman, Itcimura and Todd (1997, 1998), Heckman and Smith (1999) and Heckman, Tobias and Vytlacil (2000). A crucial aspect in these studies is the self-selective nature of the participation decision. Specifically, it is assumed that all studied units are free to participate in a particular program. Also important, though, is that for a single period an individual can be observed either in status 0 or in status 1 but never in both of them simultaneously. These two facts are carried through the present thesis as the key assumptions of the analysis.

Until the late 1960's, the availability of data sources for related empirical studies was very limited. However, over the last decades, significant breakthroughs in empirical research have been triggered by innovations statistical and econometric methods and by greater availability of new types of data.

Nowadays, the raw material in evaluation studies research is microdata, i.e. economic information about large groups of individuals, households or firms. They usually appear as cross-section data and, to an increasing degree, as longitudinal (panel) data. As it has been reported by the Royal Swedish Academy of Sciences (2000), a few years ago there was much of this kind of information only for groups of firms. It is only recently that this remarkable growth of microdata databases appeared also on individuals and households. As a result, microeconomic applications have largely been dominated by studies of individual and household behavior.

These kinds of data are not systematically collected on databases in Greece. Neither the National Bureau of Statistics of Greece nor any other institution could provide relevant data in order to be analyzed in this thesis. For this reason, the present work is limited to an extended review of the alternative methods that can be performed to evaluate a social program.

1.4 The Important Problems in Conducting Evaluation Studies

In any of the above evaluation applications, two important problems arise. The impossibility to estimate $\Delta_i = Y_{1i} - Y_{0i}$ for each person i in a single period, since each sampled individual can be perceived as either status 1 "receiver" (participant) or status 0 "receiver" (non-participant), constitutes the *evaluation problem*.

One may attempt to overcome this problem in various ways. Since Y_{1i} and Y_{0i} outcomes are observed for different persons, calculation of a simple mean estimates, $\overline{Y}_1$ and $\overline{Y}_0$, respectively, leads to an estimate of the mean benefit from participating to the program as:

$$\overline{\Delta} = \overline{Y}_1 - \overline{Y}_0 \tag{1.1}$$

One can do more than this. Assume the linear model

$$Y_i = X_i\beta + D_i\theta + u_i$$

where X_i is a $N \times J$ vector of individuals' attributes; D_i is the $N \times 1$ vector of dummy variables indicating the status for each person; β is a $J \times 1$ vector of the effect of attributes on Y_i while θ is the effect of participation on Y_i; and u_i is a $N \times 1$ vector of random components. Estimating the $N \times 1$ vector of outcomes Y_i with Ordinary Least Squares (OLS) and calculating $\overline{Y}_1$ and $\overline{Y}_0$, the gross impact $\overline{\Delta}$ can be evaluated as in equation (1.1). However, the evaluation problem cannot be approached by such simple procedures.

Due to the self-selective nature of individuals' decisions, the samples of participants and non-participants cannot be thought as random. Indeed, it is possible that persons with certain observable or/and unobservable characteristics tend to participate in the program. In such circumstances the conventional statistical inference collapses and simple methods (e.g. OLS regression and mean comparisons) yield biased estimates for the parameters and the mean outcomes. This problem is generally known as the *selection bias problem* and requires special methods of analysis.

Maddala (1978) discusses three types of program selection patterns that bias an evaluation process. Below we provide a brief description of them.

1. *Participation selectivity or self – selectivity:* This is the type of selectivity discussed in the previous paragraph. The problem is that some persons choose to be in the labor force and others choose to be out; some persons choose to go to college, other do not etc. The absence of randomness, when individuals with specific characteristics participate in a program, biases the evaluation study because the results are not the true that would have been obtained under participation from a representative population.

2. *Sample selectivity:* This bias is also sometimes referred as *Truncation bias.* It concerns a special case of self-selection bias that occurs in the case when the analyst observes only participants or non-participants data, but never both of them. Heckman

(1974) considers such an application. Evaluation of a program under sample selection bias can be performed only by applying econometric (model – based) approaches.

3. Program administrator selectivity: In many situations, participation in a social program depends partly on the administrator of the program itself. In other words, it is possible that participants are being chosen by a third person. This kind of sampling has been discussed by Manski and Lerman (1977) and Manski and McFadden (1981) and is known as *Choice-Based sampling*. Again, this sampling scheme requires specific econometric methods to produce free from bias estimates for a social program. Due to the limited literature on this selection pattern, this thesis does not study this case extensively. Only specific references are made in a subsequent paragraph.

The present thesis contains eight chapters. After the introductory comments (Chapter 1), in Chapter 2 we describe the theoretical and mathematical background of the evaluation and the selectivity bias problem. Three solutions are then suggested for cross-sectional studies. The first two concern distribution free methods, namely the *randomized social experiments*, discussed in Chapter 3, and the *method of Matching*, discussed in Chapter 4. In Chapter 5, the use of *behavioral models* is proposed as an appealing solution to the problems of interest. In Chapter 6, we recast the evaluation discussion in a panel data framework. In Chapter 7, we discuss the family of *Discrete Choice Models*, which are useful tools in modern evaluation studies. Finally, Chapter 8 summarizes the thesis and suggests topics for further research.