Chapter 2

Mathematical and Theoretical Background

2.1 Introduction

In this chapter we provide a description of the evaluation problem in a mathematical notation. After presenting the basic features, we describe theoretically the evaluation process as well as its different types mentioned in the literature. Then, we discuss the selection bias problem and introduce some possible solutions.

2.2 A Mathematical Description of the Evaluation Process

In its simplest form, the evaluation model consists of three pairs of discrete time real valued stochastic processes:

$$\begin{aligned} &\{\mathrm{Y}_{0\mathrm{i}}(t), \mathrm{X}_{0i}(t)\}_{t=-\infty}^{\infty} \\ &\{\mathrm{Y}_{1\mathrm{i}}(t), \mathrm{X}_{1i}(t)\}_{t=-\infty}^{\infty} \\ &\{D_i(t), Z_i(t)\}_{t=-\infty}^{\infty} \end{aligned} \tag{2.1}$$

$Y_{0i}(t)$ is a baseline stochastic process, with associated vector of explanatory variables $X_{0i}(t)$; $Y_{1i}(t)$ is a process that arises when an agent had participated in a social program with associated vector of explanatory variables $X_{1i}(t)$; $D_i(t)$ is a stochastic process

denoting whether or not a person participates in the program by time $T = t$ with associated vector of explanatory variables $Z_i(t)$.

The probability space generating each pair of stochastic process is $(\Omega^{(I)}, F^{(I)}, P^{(I)})$ with a family of sub-sigma algebras $(F_t^{(\mathrm{I})})$ such that:

$$F_{t-1}^{(I)} \subseteq F_t^{(I)} \subseteq F_{t+1}^{(I)} \qquad \mathrm{I} = 0,1,2.$$

where "2" denotes the indicator $D_i(t)$ process. $\Omega^{(I)}$ is understood to be a $\Re^{NI}$ sample space, $P^{(I)}$ is a common measure and $F^{(I)}$ is the associated system of Borel sets.

Definition1: σ-algebra (or Borel set or B-measurable set)

A collection F of subsets of a set X is called a σ – algebra on X, denoted by σ(F), iff:

1. $\varnothing$, X $\in F$;
2. if A $\in F$, then X \ A = A′ $\in F$;
3. if $A_1, A_2, \ldots \in F$, then $\cup_{i \in N} A_i \in F$.

Let S be a metric space. A subset of S is called a Borel set if and only if it belongs to the σ – algebra on S generating by the open sets.

Moreover, it is easy to prove the following Theorem:

Theorem 2.1: Let X be a set, and let D be any set of subsets of X. Then there is a set F of subsets of X such that:

1. F is a σ – algebra on X;
2. $F \supseteq$ D;
3. if G is any σ – algebra on X with $G \supseteq$ D, then $G \supseteq F$.

For the proof see Anderson, de Palma, Thisse (1992).

The pairs in braces defined in equation (2.1) are $F_t^{(0)}$, $F_t^{(1)}$, $F_t^{(2)}$ measurable, respectively. The three probability spaces are not necessarily disjoint and are elements of a common probability space:

$$(\Omega, F, P) = \left(\Omega^{(0)} \times \Omega^{(1)} \times \Omega^{(2)}, F^{(0)} \times F^{(1)} \times F^{(2)}, P\right)$$

The associated regressor processes are such that the σ – algebras generating $Y_i(t)$ and $Z_i(t)$ lie in the σ – algebras generating the regressors. Thus:

$$\begin{aligned} \sigma\left(\mathrm{Y}_{0i}(t)\right) &\subseteq \sigma\left(\mathrm{X}_{0i}(t)\right) \\ \sigma\left(\mathrm{Y}_{1i}(t)\right) &\subseteq \sigma\left(\mathrm{X}_{1i}(t)\right) \\ \sigma\left(D_i(t)\right) &\subseteq \sigma\left(Z_i(t)\right) \end{aligned} \tag{2.2}$$

As Heckman (1990a) states: "*this framework captures the essential idea that the dependent variables are perfectly predicted (or predictable) with respect to the regressors. Any unpredictable component of the dependent variables arises because of a failure to measure all of the relevant data generating sets*".

After setting the essential mathematical background we may proceed to a more practical description of the evaluation process without complex mathematical or probabilistic theory. In the next paragraph, the description begins by introducing some useful notation.

2.3 Useful Notation

In the simplest form, persons are imagined as being able to occupy one of two mutually exclusive states: "0" for the untreated state (i.e. non-participants) and "1" for the treated state (i.e. participants). Associated with each state is an outcome, or a $N \times 1$ vector of outcomes $Y_{1i}(t)$ and $Y_{0i}(t)$. For simplicity, we assume access to cross-sectional data so that outcomes are observed at time $t = 0$, after the completion of the program. Regarding this, we may drop out the t script from the analysis. Later, we will cast the discussion in a panel framework and t script is used again.

Let Y_{1i} be the outcome obtained given participation in a program being evaluated and Y_{0i} be the outcome in the benchmark state of non-participation. It is easiest to think of

each state as consisting of only a single outcome measure, such as earnings, but just as easily, this framework can be applied to model vectors of earnings, employment and participation to a social program.

Define, as described before, a dummy variable denoting participation:

$$D_i = \begin{cases} 1, & \text{if a person participates} \\ 0, & \text{otherwise} \end{cases}$$

Note that the choice of the non-participation (base) state "0" is arbitrary. Clearly the roles of "0" and "1" can be reversed. However, in many applications it is convenient to think of "0" as a benchmark "non-participation" state. Assumptions appropriate for one choice of "0" and "1" need not carry over to the opposite choice. With this cautionary note in mind, we proceed as a well-defined base state exists.

The outcome Y_i, observed for an individual is modeled as:

$$Y_i = D_i \times (Y_{1i}^{(c)} - C_i) + (1 - D_i) \times Y_{0i}$$
$$D_i \times Y_{1i} + (1 - D_i) \times Y_{0i} \qquad (2.3)$$

Model (2.3) is the Roy model (1951) or the switching regression model of Quandt (1972), also described in Heckman and Sedlacek (1985) and Heckman (1990a, b). C_i represents the cost of participating in the treated state. For the first time here $Y_{1i}^{(c)}$, instead of Y_{1i}, indicates the pure outcome from the treated state. From this outcome the cost of participation C_i have to be subtracted in order to take the actual outcome from participating in the social program. Intuitively, individuals choose to participate if the gain from participation minus the cost is non-negative. Throughout this thesis we assume, for reasons of simplicity, that the element or the vector of C_i represents a known constant. Therefore, in any of the following formulas C_i has already been subtracted from the pure outcome $Y_{1i}^{(c)}$ and Y_{1i} is observed directly.

2.4 The Evaluation Process

An important question that has to be answered is what exactly an analyst seeks to evaluate in terms of a program. In evaluating a social program, many different comparisons can be considered. For example, one might like to compare the individuals' outcomes of an already organized program with the individuals' outcomes of

1. The same program if it was operated in a different way,
2. The non-participation to a program or the non – existence of a program at all,
3. An alternative program etc.

In all cases, the analyst has to define the outcomes being compared. For example, in terms of a training program, these outcomes include *direct benefits* received like post-program payments and *indirect benefits* for both participants and non-participants like an increase in payment due to specialization to a job sector. Relative to the indirect benefits, Lewis (1963) states that in modern economy a training program also affects the persons with whom the participants compete in the labor market and the firms that hire them. A discussion on indirect benefits is found in Chapter 3 of the thesis.

Heckman (1990a) distinguishes three types (*branches*) of evaluation studies. It is important to recognize that different scholars of evaluation have different levels of interest in the corresponding evaluation questions. Below we attempt an introductory approach to them.

(E-1): Gross Benefit Evaluation Study

a) Y_{1i} observed only when $D_i = 1$
   Y_{0i} observed only when $D_i = 0$
   Δ_i never observed

Aim: We seek some feature of the distribution of Δ_i conditioned or not conditioned on X_{0i}, X_{1i}, Z_i.

This type receives the most attention in the literature. The exclusive focus on (E-1) by biometricians, psychometricians and statisticians is a consequence of the perceived unimportance of purposive self-selection decisions by individuals in the programs considered by those analysts.

Under (E-1) there are usually two questions of interest for an evaluator. We represent them ranked by the degree of attention each question receives by the program evaluators. Here we limit the discussion to the simple theoretical description of those questions. In another paragraph we will see more extensively how they can be answered.

Question 1: What is the effect of training on the trained.

After the subtraction of the program costs, the net benefit $(\Delta_i | D_i) = Y_{1i} - Y_{0i}$ (or the equivalent expression given also X_{0i}, X_{1i}, Z_i), for each participant $(D_i = 1)$ is sought to be calculated. In other words, the analyst attempts to evaluate *the effect of Treatment on the Treated persons (TT)*. This is the "bottom line" stressed in many evaluations.

Question 2: What is the effect of training on randomly assigned trainees.

The net benefit $\Delta_i = Y_{1i} - Y_{0i}$ (or equivalently given X_{0i} and X_{1i}) of a group of trainees independently of the status of D_i is sought to be calculated. The answer to this question, referred formally as the *Average Treatment Effect (ATE)*, would be of great interest if training were mandated for an entire population, as in workfare programs that force welfare recipients to take training.

Answering these questions is of great interest in modern econometric literature. A huge bibliography exists on this subject. The most representative references are these of Heckman (1978, 1979, 1991b), Manski (1989), Newey, Powell and Walker (1990), Heckman and Smith (1998), Heckman, Ichimura and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998), Heckman, Lalonde and Smith (1999), Heckman and Vytlacil (1999).

Except for these two mean effects, analysts are also interested in estimating the Local Average Treatment Effect (LATE) and the Local Instrumental Variable Effect (LIV). These parameters, which developed recently, are best to be described in a later paragraph of the thesis, under the *instrumental variable* framework.

Although formulation of these parameters is simple, estimation is much more difficult that it seems due to the evaluation problem. In fact, calculation of Δ_i cannot be performed

because an analyst may observe either Y_{0i} or Y_{1i} for a person in a single period, but never both of them. Only by considering particular econometric procedures or by adopting specific assumptions, one can deal with this feature that constitutes the so-called *Evaluation Problem* or the *Fundamental Problem of Causal Inference* (Holland, 1986).

(E-2): Participation Evaluation Study

a) Which variables of $\sigma(Z_i)$ affects participation decision given that new variables not in $\sigma(Z_i)$ may become relevant or that some components of Z_i may be missing.

Aim: Assessing the effect of alterations in Z (variables in conditioning set and associated probability measures) on program participation.

Question 3: What is the effect of various individuals' characteristics, job subsidies, advertising, local labor market conditions, family income race and sex on application decisions.

Question 4: What is the effect of center performance standards, profit rates, local labor market structure and governmental monitoring on training center acceptance of applicant decisions and placement in specific programs.

This is a special case of (E-1) evaluation study. In evaluating alternative policies, it is essential to know not only the gross benefits to participants but also how the socio-economic environment determines participation.

An important factor that affects participation to a social program has to do with the possibility of a choice between the program itself and a subsidized job. Subsidized jobs have been proposed as a flexible, efficient and, in contrast to alternative employment policies such as job training programs, relatively low-cost method for reducing unemployment. Persons offered a subsidized job might take it or opt for their best, unsubsidized alternative. The option may be conferred simply by eligibility or it may be conferred only on participants.

Though in principle an attractive alternative to conventional unemployment policies, subsidized jobs have been tried only rarely in the United States. Recent empirical studies have raised questions about whether these jobs are a practical policy for reducing unemployment. Burtless (1985) speculates that being identified as a subsidy recipient has a stigmatic effect on job seekers that outweighs the value of the subsidy but he also recognizes the existence of other reports that have yielded more positive results though they are statistically significant only for specific social groups.

(E-3): Prediction Evaluation Study

a) Combination of both features of evaluation problems (E-1) and (E-2).

Aim: Assessing the effect of alterations in $\sigma(Z_i)$, $\sigma(X_{0i})$ and $\sigma(X_{1i})$ and probability measures defined on these algebras on D_i, Y_{0i} and Y_{1i} respectively.

Question 5: What are the effects of individuals' attributes, family background, center profit rates, subsidies and local labor market conditions on the decision to participate in a program on the length of time taken to complete the program and on the amount of outcome Y_i for every individual.

Heckman (1991b), discriminates economists from statisticians by noting that the latter are mainly focused with the mean comparison problem of (E-1) while the former are typically interested in estimating causal effects rather than mere associations between variables. Therefore, they are more often preoccupied with (E-3) and the problem of recovering parameters of structural models in order to conduct counterfactual policy analysis. The present thesis intends to cover both evaluation approaches. Chapters 3 and 4 are mainly concerned with estimation of mean impacts. Then, in Chapter 5, we recast the discussion to account for the economists approach.

2.5 Other Evaluation Questions

Apart from the above issues, several interesting evaluation questions require knowledge of the feature of the distribution of outcomes. As referred in Heckman, Smith and Clements (1997) from the standpoint of the analyst that denotes as "0" the state of absence from the program, it is of interest to know, among other things:

❖ The proportion of people taking the program who benefit from it:

$$\Pr\left(\mathrm{Y}_{1i} \succ \mathrm{Y}_{0i} \middle| \mathrm{D_i} = 1, \mathrm{X}_{0\mathrm{i}}, X_{1i}, Z_i\right) = \Pr\left(\Delta_i \succ 0 \middle| \mathrm{D_i} = 1, \mathrm{X}_{0\mathrm{i}}, X_{1i}, Z_i\right)$$

This is a measure of how widely program gains are distributed among participants. Participants in the political process with preferences over distributions of program outcomes would be assign more weight to the program that distributes the favorable outcomes more broadly than another, even though the two programs have the same mean outcome.

❖ The proportion of the total population that benefits from the program:

$$\begin{aligned} &\Pr\left(\mathrm{Y}_{1i} \succ \mathrm{Y}_{0i} \middle| \mathrm{D_i} = 1, \mathrm{X}_{0\mathrm{i}}, X_{1i}, Z_i\right) \times \Pr\left(\mathrm{D_i} = 1 \middle| Z_i\right) \\ = &\Pr\left(\Delta_i \succ 0 \middle| \mathrm{D_i} = 1, \mathrm{X}_{0\mathrm{i}}, X_{1i}, Z_i\right) \times \Pr\left(\mathrm{D_i} = 1 \middle| Z_i\right) \end{aligned}$$

The above formula measures the proportion of the entire population that benefits from the program, assuming costs of financing the program are broadly distributed without being related to the specific evaluated program.

❖ Selected quantiles of the impact distribution of participants:

$$\inf_{\Delta}\left\{\Delta_i : F\left(\Delta_i \middle| \mathrm{D_i} = 1, \mathrm{X}_{0\mathrm{i}}, X_{1i}, Z_i\right) \succ q\right\}$$

where q is a quantile of the distribution. This measure is of interest if one wants to study the distribution of program benefits.

❖ The distribution of gains of participants at selected non - participation state values:

$$F\left(\Delta_i \mid \mathrm{D_i} = 1, \mathrm{X_{0i}}, \mathrm{X_{1i}}, \mathrm{Z_i}, \mathrm{Y_{0i}} = y_{0i}\right)$$

Evaluators who take a special interest in the impact of a program on recipients in the lower tail of the no-participate distribution would find this measure interesting.

❖ The increase in the level of outcomes for participants above a certain threshold $\bar{y}$ due to a policy:

$$\Pr\left(\mathrm{Y_{1i}} \succ \bar{y} \mid \mathrm{D_i} = 1, \mathrm{X_{0i}}, X_{1i}, Z_i\right) - \Pr\left(\mathrm{Y_{0i}} \succ \bar{y} \mid \mathrm{D_i} = 1, \mathrm{X_{1i}}, X_{0i}, Z_i\right)$$

Positiveness in the above expression indicates that the distribution of gains for the participants dominate the distribution of outcomes of the same persons if they did not participated.

Obviously, answering these questions presupposes knowledge on the joint distribution of outcomes, $F\left(Y_{0i}, Y_{1i}, D_i \middle| X_{0i}, X_{1i}, Z_i\right)$. From ordinary data on participants and non-participants estimation of the conditional distributions

$$F_1\left(Y_{1i} \middle| D_i = 1, X_{0i}, X_{1i}, Z_i\right) \quad \text{for participants}$$

$$F_0\left(Y_{0i} \middle| D_i = 0, X_{0i}, X_{1i}, Z_i\right) \quad \text{for non-participants}$$

is a simple task. However, due to the evaluation problem, it is not possible to estimate the counterfactual conditional distributions

$$F_1\left(Y_{0i} \middle| D_i = 1, X_{0i}, X_{1i}, Z_i\right) \quad \text{for participants had they not participated}$$

$$F_0\left(Y_{1i} \middle| D_i = 0, X_{0i}, X_{1i}, Z_i\right) \quad \text{for non-participants had they participated}$$

at least without additional information. As a consequence neither the joint distribution

$$F\left(Y_{0i}, Y_{1i}, D_i \middle| X_{0i}, X_{1i}, Z_i\right)$$

can be estimated, unless specific assumptions hold. The perceived assumptions are analyzed in a later paragraph of this thesis. Without their imposition none of the above questions can be answered.

2.6 Solutions to Evaluation Problem

The evaluation problem is a missing data problem. Either Y_{0i} or Y_{1i} may be observed for a person but never both of them. Thus Δ_i cannot be calculated individually, at least without further assumptions. Holland (1986) considers a set of alternative assumptions that when either is valid, evaluation can be performed individually.

Temporal Stability and Causal Transience

When the value of Y_{0i} does not depend on when the sequence "apply 0 to *i* then measure *Y* on *i*", constancy of response over time is asserted. In addition when the value of Y_{1i} is not affected by the prior exposure of *i* to the above sequence, the effect of cause 0 and the measurement process that results in Y_{0i} is transient and does not change *i* enough to affect Y_{1i} later. These two assumptions are called *Temporal Stability* and *Causal Transience*, respectively, and when are plausible it is a simple matter to measure Y_{1i} and Y_{0i} by sequential exposure of *i* to 0 and then 1, measuring Y_i after each exposure.

Unit Homogeneity

By simply assuming that Y_{1i} is the same across all participants and correspondingly Y_{0i} is the same across all non-participants it is a simple matter to calculate the difference $\Delta_i = Y_{1i} - Y_{0i}$ for each person. This assumption is often applicable in laboratory studies and is known as *Unit Homogeneity*.

Constant Effect

By making the assumption of Constant Effect, one assumes that the effect of 1 to each unit (Y_{1i}) can be measured by adding a constant amount Δ_i to its status "0" response (Y_{0i}). Thus

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

which is constant for all units. This assumption is called *additivity* and when holds Δ_i can be measured for each units separately. Cox (1986) comments that this assumption is not always plausible and cannot be considered without being tested first.

It is true that the above assumptions oversimplify the evaluation process. In fact, this constitutes the major reason for which they are considered as implausible in most practical situations. An alternative approach to evaluate a social program is obtained through carriage of the evaluation problem on a population level and work within this framework. Instead of estimating Δ_i for every person, the analyst can estimate the mean of Δ_i, $\overline{\Delta}$. This measure of *Location* or *Central Tendency* may offer a valuable compendious view of the benefits from participating into the program. Although the limitations of this measure (sensitiveness to outliers, not representative as typical value for skewed, unimodal distributions), often enact the analyst to consider a more robust estimator, like the median, the difficulties in making statistical inference with such measures indicate the most common reason of focusing on mean impacts to evaluate a social program. A relative difficulty in implementing median or other quantile inference is the estimation of the corresponding distribution function, at least without numerical methods like *bootstrap*.

Calculation of mean differences

In any case, simple computation of simple mean parameters, $\overline{Y}_0$ and $\overline{Y}_1$, from the dataset cannot be performed in the presence of selectivity bias. Regarding ATE and TT, such an approach would intimately lead to biased results. According to the econometric theory of *omitted variables* (see Griliches, 1957), "incorrect conceptual understanding of the phenomena under study or inability to collect data on all the relevant factors related to

the outcome under study can result in seriously biased estimates of mean effects". For the two mean parameters, it can be shown that selectivity results in $E^r(\Delta_i) \neq E(Y_1 - Y_0)$ and $E^r(\Delta_i | D_i = 1) \neq E(Y_1 - Y_0 | D_i = 1)$, respectively, where "*r*" superscript denotes mean impacts of the representative sample (no selectivity).

Alternatively, a model approach can be considered. Assume that elements of Y_{0i}, Y_{1i} and D_i are functions of X_{0i}, X_{1i} and Z_i, respectively, and use structural models to write:

$$\begin{aligned} \mathrm{Y}_{0i} &= g_0(\mathrm{X}_{0i}) + u_{0i} \\ \mathrm{Y}_{1i} &= g_1(\mathrm{X}_{1i}) + u_{1i} \\ D_i &= \begin{cases} 1 & \text{if } \mathrm{Z}_i \in \widetilde{Z}_i \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2.4}$$

where $\widetilde{Z}_i$ is a measurable subset of the sample space of Z_i and u_{0i} and u_{1i} are random errors associated with Y_{0i} and Y_{1i}.

Simple Ordinary Least Squares (OLS) regression would lead in biased estimates of program impacts due to the non-random design of the relative samples. For example, if persons elect to participate in a program precisely because of the poor alternatives outside the program, non-participants would have outcomes higher than those participants would have if they had not participated, implied a negative bias term. Denoting as "α" the information available to the analyst and with "*m*" the missing information (unrepresented information), the vectors of attributes is divided into two components:

$$\mathrm{X}_{1i} = \mathrm{X}_{1ai} + \mathrm{X}_{1mi}$$
$$\mathrm{X}_{0i} = \mathrm{X}_{0ai} + \mathrm{X}_{0mi}$$
$$Z_i = Z_{ai} + Z_{mi}$$

In terms of ATE(X) and TT(X), selection bias is represented as

$$E^r\left(\Delta_i \middle| X_{0i}, X_{1i}\right) \neq E\left(Y_1 - Y_0 \middle| X_{0i}, X_{1i}\right) \tag{2.5}$$

$$E^r\left(\Delta_i \middle| D_i = 1, X_{0i}, X_{1i}, Z_i\right) \neq E\left(Y_1 - Y_0 \middle| D_i = 1, X_{0i}, X_{1i}, Z_i\right) \tag{2.6}$$

respectively.

2.7 Discriminating Selection Bias

The inequalities represented by equations (2.5) and (2.6) define the population problem of *self selection bias* in making mean comparisons. Regarding equation (2.4), sample selection bias occurs because of the dependence between the outcomes Y_{Di} and the selection variable D_i. Consequently, individuals are not assigned randomly in either state but in terms of a specific rule that affects their outcomes. Generally, sample or self selection bias arises for one of two not necessarily mutually exclusive reasons. The first is termed as *selection on observables* while the second, and more burdensome to cope with, is known as *selection on unobservables*.

Selection on Observables

Selection on observables occurs when the dependence between Y_{Di} and D_i is purely through the observed variables, X_{0i} and X_{1i}, that influences selection into the program. In other words, for a "returns from college education" study, individuals that belong to a specific socio-economic class may compose the sample of college education receivers. Thus, the sample will not be representative of the target population of the study.

In terms of the model (2.4), selection on observables causes

$$E(u_i \middle| D_i, X_i) \neq 0 \tag{2.7}$$

$$E(u_i \middle| D_i, X_i, Z_i) \neq 0 \tag{2.8}$$

but $E(u_i|X_i,Z_i)=E(u_i|D_i,X_i,Z)$ and there are not any unobservable characteristics that determine participation. Assuming knowledge of the functional form of $E(u_i|X_i,Z_i)$, the corresponding term can be inserted in (2.4) and the resulting equation can be estimated by regression methods to obtain consistent estimates of the effect of participation. Such estimators are members of the class of *Control Function Estimators*, where $E(u_i|X_i,Z_i)$ is called a control function. This is an econometric approach to the evaluation problem and it is further discussed in Chapter 5. Alternatively, non-parametric *matching methods* or *randomized experimental procedures* can be conducted to contrast the earnings of participants and non-participants. A relative analysis is found on Chapters 3 and 4, respectively.

Selection on Unobservables

Selection on unobservables appears when the dependence between Y_{Di} and the indicator variable D_i is due to unobserved variables. Put differently, it is considered that unobserved characteristics, affecting the participation decision, are correlated with unobserved characteristics affecting the outcomes. In this case persons' recorded outcomes do not come from a randomly chosen population due to variables that do not appear in the analysis. This causes

$$E(u_i|D_i,X_i) \neq 0 \tag{2.9}$$

$$E(u_i|D_i,X_i,Z_i) \neq E(u_i|X_i,Z_i) \tag{2.10}$$

Failure to include an estimate of the unobservables as a correction term in the model leads in incorrect inference regarding the impact of participation to outcomes. As before, experimental designs are able to provide adequate answers to this problem. Other *parametric* and *semi-parametric* approaches are discussed in Chapter 5.