Exploiting instrumental variables in causal inference with nonignorable outcome nonresponse using principal stratification [1]

Fabrizia Mealli, Università di Firenze, mealli@ds.unifi.it

Barbara Pacini, Università di Bologna, barbara.pacini@unibo.it

Preliminary draft - May 2008

Abstract: In this paper we consider a specific post-treatment complication that may arise in both randomized and observational studies, namely the problem of nonignorable nonresponse on an outcome variable. This is a typical topic usually known in the econometric literature as *endogenous selection*; here we tackle this problem specifically within a causal inference framework. By exploiting Principal Stratification, we analyze and propose identification strategies with and without the availability of an *instrumental* variable for nonresponse.

We focus on the different role and meaning of the instrumental variable, also by comparing our framework with a general nonseparable selection model setting. As a motivating example we consider a simplified evaluation study in the field of financial aids to firms, where typically missingness on the outcome variables, such as variables related to firms'performances, can rarely be assumed missing at random.

Keywords: causal inference, nonresponse, principal stratification, instrumental variables

[1]Financial support for this research was provided by Miur Cofin 2005 grant.

1 Introduction

When the goal of inference is estimating causal effects, we usually have to face problems related to how data are observed. In observational studies, the most relevant of such problems is the fact that assignment to treatment is not under the control of the investigator; in addition some studies, both observational and experimental, may be affected by different sorts of post-treatment selection of observations due to, e.g., non response, truncation or censoring "due to death". Moreover, one may be interested in decomposing the total effect of a treatment on an outcome into a direct effect and an indirect one mediated by another intermediate variable. All such complications require to somehow control for them, but the use of the standard statistical conditioning may be in general improper (Rubin, 1978; Heckman, 1974; Rosembaum, 1984; Rubin, 2004).

A relatively recent approach to deal with post-treatment complications is Principal Stratification (PS), as first defined by Frangakis and Rubin (2002) within the framework of the Rubin Causal model (Rubin, 1974; Holland, 1986) and applied mainly in experimental studies (Barnard et. al., 2003; Zhang et. al., 2006; Mattei and Mealli, 2007). In Frangakis and Rubin (2002), PS was introduced in order to give a formal definition of surrogate endpoints; it was then used to define direct and indirect effects (Mealli and Rubin, 2003). As Rubin (2004) points out, the PS framework can be viewed as having its seeds in the Instrumental Variables (IV) method of estimation of causal effects. Indeed, the approach to adjust for noncompliance applied in Angrist, Imbens and Rubin (1996) and in Imbens and Rubin (1997) is a special application of the PS framework, where the compliers are a principal stratum with respect to the post-treatment compliance behavior.

Despite the use of PS to solve particular issues, the framework appears to be a very general one, that can be applied in various contexts, also not originally related to causal inference (Frangakis et al., 2007). Furthermore, the framework may lead to both parametric and semi(non)parametric inference, depending on the set of assumptions that can be reasonably maintained, as well as whether point or partial identification is to be achieved.

In the econometric literature post-treatment complications are usually described as problems of endogenous selection and include treatment assignment in observational studies, self-selection, non response, censoring or truncation "due to death". They are usually represented by means of

selection models (SM, Heckman, 1974; Gronau, 1974). Since the seminal work of Heckman, various extentions of the model have been proposed, which include semi and nonparametric versions (Pagan and Ullah, 1997; Vella, 1998). While parametric selection models can be rather restrictive in terms of distributional assumptions, semi and nonparametric extensions usually require additional exclusion restrictions (instrumental variables) to maintain point-identification of (causal) functionals of interest (parameters, conditional expectations, partial derivatives, probability distributions). Identification issues in general nonseparable structural equation models is reviewed in Imbens (2006).

In this paper we consider a specific endogenous selection problem, namely a nonignorable nonresponse on an outcome variable. Other complications are not explicitly treated, such as endogenous treatmente assignment. measurement errors, censoring due to death. In particular, hypotheses on the exogeneity of the treatment are assumed to hold by design (experimental case) or by assumptions (e.g., unconfoundedness in observational studies). Using Principal Stratification, we show which causal estimands may be identified and estimated under different sets of assumptions, in particular depending on the existence of a variable that may serve as an instrument, in some sense, for nonresponse. Indeed, causal inference requires some assumptions about the population, the sampling process and the behavior of the subjects under study. The *credibility* of (causal) inference decreases with the strength of the assumptions maintained (Manski, 2003). Note that hypotheses are not all on the same ground, and they may have different nature, as well as a different degree of agreement.

We focus on the different role and meaning of the instrumental variable, also by comparing our framework with a general nonseparable selection model setting. We also argue that Principal Stratification is able to suggest alternative identification strategies not always easily translated into assumptions of a selection model.

In order to support our reasoning we consider, as a motivating example, a simplified evaluation study in the field of financial aids to firms, where typically nonresponse on the outcome variables, such as variables related to firms'performances, can rarely be assumed missing at random.

The paper is organized as follows: in section 2 principal stratification is presented together with

its main characteristics; section 3 is devoted to recall the basic features of simultaneous equation systems with a sample selection equation. In section 4, we describe our working example and show partial and point identification strategies in the absence of an instrumental variable.

Section 5 introduces an instrumental variable for nonresponse. The instrument can be interpreted as an additional intervention, i.e., a treatment variable, or as an additional post-treatment variable. We consider alternative hypotheses characterizing the instrument that may allow to point-identify specific causal estimands. Section 6 summarizes our findings and concludes.

2 Principal Stratification and its Role for Causal Inference in Experimental and Observational Studies

Principal stratification has been first introduced by Frangakis and Rubin (2002), in order to address post-treatment complications in an experimental setting. The framework can, however, be easily extended to an observational setting under specific hypotheses on the assignment mechanism. We first introduce "potential outcomes" (see Rubin, 1974) for one post-treatment variable, Y, and a binary treatment, T. If unit i in the study ($i = 1, \ldots, N$) is to be assigned to treatment $T = t$ ($t = 1$ for treatment and $t = 0$ for no treatment), we denote by $Y_i(1)$ and $Y_i(0)$ the two potential outcomes, either of which can be observed depending on the value of T. A causal effect of T on Y is defined, on a single unit, as a comparison between $Y_i(1)$ and $Y_i(0)$. The fact that only two potential outcomes for each unit are defined reflects the acceptance of the SUTVA (Stable Unit Treatment Value Assumption, Rubin, 1980) assumption, which rules out interference between units and difference versions of the treatment for the units. We also denote with $S_i(t)$ a post-treatment potential variable, which is, without loss of generality, assumed to be an indicator equal to 1 if a specific post-treatment event happens and 0 otherwise. For instance, S may represent a response indicator for a specific item in a questionnaire, or a survival indicator: in these examples, $S_i(t) = 0$ precludes the observation of $Y_i(t)$.

In an observational setting, various hypotheses can be posed on the assignment mechanism. In what follows, we will assume that treatment assignment is unconfounded given a vector X of

observed pre-treatment variables:

Assumption $PS\,1:\ T \perp\!\!\!\perp S(0), S(1), Y(0), Y(1)|X.$

In other words, we assume that within cells defined by the values of pre-treatment variables X, the treatment is randomly assigned or, at least, is assigned independently of the post-treatment variables considered relevant for the study. If we indicate with t_i the observed treatment assigned and received, the observed data are

$$\Big(t_i, S(t_i), Y(t_i), x_i\Big) \qquad i = 1, \ldots, N.$$

Consider now the potential post-treatment variables $S_i(0)$ and $S_i(1)$. Within each cell defined by specific values of the pretreatment variables, the units under study can be stratified into four latent groups, named Principal Strata, according to the joint value of the potential variables $(S_i(0), S_i(1))$; the strata are the following:

$$11 = \{i : S_i(1) = S_i(0) = 1\}$$

$$10 = \{i : S_i(1) = 1, S_i(0) = 0\}$$

$$01 = \{i : S_i(1) = 0, S_i(0) = 1\}$$

$$00 = \{i : S_i(1) = S_i(0) = 0\}.$$

Define the population proportions of units belonging to each stratum in the cell $X = x$ as $\pi_{11|x}$, $\pi_{10|x}$, $\pi_{01|x}$, and $\pi_{00|x} = 1 - \pi_{11|x} - \pi_{10|x} - \pi_{01|x}$. This stratification of units corresponds to the basic principal stratification, as defined in Frangakis and Rubin (2002): the basic principal stratification P_0 with respect to post-treatment variable S is the partition of units $i = 1, \cdots, n$ such that, within any set of P_0, all units have the same vector of $(S_i(0), S_i(1))$. More generally, a principal stratification P with respect to post-treatment variable S is a partition of the units whose sets are unions of sets in the basic principal stratification P_0. Let G_i represent the principal stratum to which subject i

belongs, $G_i = \{11, 10, 01, 00\}$. The principal stratum membership, G_i, is not affected by treatment assignment t_i, so it only reflects characteristics of subject i, and can be regarded as a covariate, which is only partially observed in the sample (Angrist et al., 1996); by unconfoundedness, however, it is guaranteed to have the same distribution in both treatment arms, within cells defined by pre-treatment variables. We usually need to adjust for the principal strata, which synthesize important unobservable characteristics of the subjects in the study. Note that Assumption PS1 implies the following: $Y(0), Y(1) \perp\!\!\!\perp T | S(0), S(1), X$: potential outcomes are, therefore, independent of the treatment given the principal strata. This implication confirms the idea that treated and control units can be compared conditional on a principal stratum. In some sense, we can state that principal strata play a similar role of control functions in deriving independence conditions, even if not derived from a model.

Usually, information on causal effects is contained in a particular principal stratum and, as a consequence, evidence on causal effects is sought only for this specific stratum. In the present exemplified context, direct information on the causal effect can be found in the 11 stratum (e.g., respondents or survivors under treatment and control), because only for units belonging to this stratum one can consistently compare $Y(1)$ and $Y(0)$: if S represents non response or death, in fact, in all strata but 11 only one potential outcome or none can be observed, so that the causal effect which can be estimated is an effect within stratum, e.g., the average effect $E(Y(1) - Y(0)|11)$. Indeed, this is a principal causal effect, which is a properly defined causal effect, because it is obtained as a comparison on a common set of units, as defined in Frangakis and Rubin (2002) (equation 2.1).

The purpose of inference is to estimate the probabilities of strata belonging ($\pi_{11|x}$, $\pi_{10|x}$,$\pi_{01|x}$) and the distribution of the potential outcomes within each stratum, under different identifying distributional and behavioral assumptions.

Note that the same framework can be easily extended to cases with non binary treatment, non binary post-treatment variable and more than one post-treatment variables. In particular, the framework will also be used to characterize an instrumental variable for nonresponse, where the instrument can be interpreted as both an additional treatment, and as an additional post-treatment

variable.

3 Representing endogenous selection problems by means of simultaneous equation systems with a sample selection equation

The potential outcome framework, on which Principal Stratification rests, is now widely used to represent causal inference problems. We can however find some connections with the traditional econometric literature, which relies on structural simultaneous equations to represent causal relations. In order to represent the causal problem formalized in previous section using PS, consider the following three-equation structural model:

$$\begin{aligned} Y_i &= g(t_i, x_i, \epsilon_i) \\ S_i &= h(t_i, x_i, \eta_i) \\ T_i &= l(x_i, \omega_i), \end{aligned} \tag{1}$$

where Y_i is observed only if $S_i = 1$. The three equations describe causal relationships between the left- and right-hand side variables; in particular the first equation is the one of primary interest and it is a primitive for causal inference purposes. Note that, if we do not impose any functional restrictions on g, h, and l and we do not restrict ϵ_i, η_i, and ω_i to be scalars, potential outcomes can be retrieved as:

$$\begin{aligned} Y_i(t) &= g(t, x_i, \epsilon_i) \\ S_i(t) &= h(t, x_i, \eta_i), \end{aligned} \tag{2}$$

for $t = \{0, 1\}$. We maintain the following hypothesis:

Assumption $SM1$: $\omega_i \perp\!\!\!\perp \eta_i, \epsilon_i$.

which holds by design if T is randomized and by assumption conditional on some pre-treatment covariates X under unconfoundedness (see Assumption PS1, section 2). Consequently, and in order to simplify notation, the third equation in model (1) will be omitted in the sequel, as well as the covariates X, because we will implicitly condition on them, assuming thus to be within cells defined by covariates.

Our setup is closely related to the one involving the estimation of the effect of a potentially endogenous regressor on an outcome, where the endogeneity is due to the association among unobservables entering the outcome and regressor equations (Imbens, 2006; Hoderlein and Mammen, 2007). As Imbens (2006) points out, in such a setting a possible strategy is to eliminate the endogeneity at least for some subpopulation (types of units), usually characterized by certain values of the unobservables that induce them to respond to regressors and instruments (if available) in a specific way[2]. These subpopulations, as we will see in the sequel, may find a correspondence in terms of principal strata; with this respect, the simultaneous equation framework provides us with an alternative viewpoint in the interpretation of identifying assumptions and of estimands that we will find within the PS approach. In addition, this general equation system offers us the opportunity to review the results developed in the econometric literature by imposing some functional and distributional restrictions on g, h, and ϵ_i, η_i.

Note that the choice of estimands plays a crucial role in our and related approaches, and it is sometimes questioned if these effects defined for subpopulations are of interest, because they generally do not correspond to specific policy parameters. We argue again that it is often difficult to infer on effects that affect the whole population, and it is much easier, namely requires fewer assumptions, to evaluate treatments that move individuals only locally, by eliminating endogeneity for some subgroups and not necessarily for others.

In order to avoid such criticism, other approaches follow either a global identification goal of g or some of its functionals (Imbens and Newey, 2003), or a local identification objective at a fixed

[2]Vytlacil (2002) shows an equivalence between PS and selection models in the specific linear IV setting. He proves that assumptions on the principal strata directly translate into assumptions on a selection latent index model and vice-versa: under the LATE (Latent Average Treatment Effect, Imbens and Angrist, 1994) independence assumption and the LATE monotonicity assumption the two models are shown to be observationally equivalent.

position of the regressors (Chesher, 2003, 2005). In these cases identifiability strongly depends on the nature of regressors (discrete or continuous) and on monotonicity assumptions with respect to the unobservables. For the identification of average marginal effects, Imbens and Newey (2003), Altonji and Matzkin (2005) and Hoderlein (2005) provide results without assuming monotonicity in unobservables, using derivatives of the mean regressions. Some related work can be found in Florens, Heckman, Meghir and Vytlacil (2005), Chernozhukov, Imbens and Newey (2005), Chernozhukov and Hansen (2005) and Heckman and Vytlacil (1999, 2001).

A closely related stream of the literature focuses on conditioning directly on the unobserved components in the selection equation in order to eliminate endogeneity; because these variables cannot be observed, they can be sometimes consistently estimated using different assumptions and estimation strategies and used as covariates in a generalized control function approach. Additivity with respect to the unobservables is usually required. This kind of approach includes parametric and semiparametric specifications of selection models (Heckman, 1974; Lee, 1982, 1983; Gallant and Nychka, 1987) and two-step estimation settings (Olsen, 1980; Wooldridge, 1994; Lee, 1984; Pagan and Vella, 1989; Honore *et al.*, 1997), that may or not include some single-index restrictions (Cosslet, 1991; Robinson, 1988; Newey, 1990; Lee, 1994; Ahn and Powell, 1993; Choi, 1990; Li and Wooldridge, 2002; Ichimura, 1993).

The approach we follow is nonparametric: without a priori introducing monotonicity and additivity of the error terms we use principal stratification to discuss identification issues. In particular, we show that some parameters can only be partially identified (Manski, 2003) and, more importantly, propose sets of assumptions to yield point identification of treatment effects for some subpopulations. We focus on the identification issues assuming that the distribution that are asymptotically revealed by the sampling process are known, so avoiding to take account of specific statistical inference problems related to estimation in finite samples.

4 Identifying causal effect with nonignorable nonresponse on the outcome

In order to focus on a specific post-treatment complication, which is the main goal of the paper, we consider, as a motivating example, the following simplified setting concerning the evaluation of financial aids to firms. Let T be a binary treatment which represents public financial assistance to firms ($T = 1$ for treatment and $T = 0$ for no treatment). We assume that T is unconfounded given a vector of pre-treatment covariate X; we will implicitly condition on X in what follows, i.e. we will assume to be within a cell defined by the values of X. The intermediate post-treatment variable S represents the response to a post-treatment questionnaire on firms' performances, the outcome variable of interest being the turnover (sales proceeds) Y. We are facing the post-treatment complication of a potentially nonignorable missing mechanism of the outcome variable. This artificial setting is consistent with evidence from the real world, where for Italian sole traders company accounts are not public and variables on firm performances must be gathered through an interview. Typically missingness on turnover variables can rarely be assumed missing at random[3].

This empirical setting can be formalized within the PS framework as presented in section 2, where the intermediate binary variable S is the response indicator. There are thus potentially four latent strata. Without any further restriction and avoiding any form of extrapolation, the only causal effect on which we can have evidence from the data is the effect within the 11 stratum, because only for individuals belonging to the stratum we have observations on both $Y(1)$ and $Y(0)$.

In order to highlight the identification issue, we assume to know the distribution that are asymptotically revealed by the sampling process without dwelling on issues of statistical inference related to finite samples. We point out that we will use sample analogs to nonparametrically estimate features of the distributions of the observed variables. With this respect it is useful to state the correspondence between observed groups, defined by T and S, and latent strata, as shown in Table 1. Note that all the four observed groups result from a mixture of two principal strata; without any fur-

[3]For example, Mattei and Mauro (2007), from a survey of Tuscan artisan enterprises, found evidence of nonignorability of nonresponse.

Observed subgroups $O(T,S)$	Turnover Y	Latent strata
$O(1,1) = \{i : T_i = 1, S_i = 1\}$	OBS	11 or 10
$O(1,0) = \{i : T_i = 1, S_i = 0\}$	·	00 or 01
$O(0,1) = \{i : T_i = 0, S_i = 1\}$	OBS	11 or 01
$O(0,0) = \{i : T_i = 0, S_i = 0\}$	·	10 or 00

Table 1: Correspondence between observed and latent strata

ther restriction it is not possible to point-identify the strata proportions, as well as the distribution of Y within the strata, that would allow to estimate causal effects.

A common assumption used to improve identification is a kind of monotonicity assumption, that here could be stated as the nonexistence of the 01 stratum: there is no firm which would not respond under treatment but would respond under control. Under monotonicity, the relationship between observed and latent groups allows to point-identify at least the strata proportions, because we are left with two unknown proportions ($\pi_{11|x}$ and $\pi_{10|x}$) and two pieces of sampling information (coming from the proportions of respondents among treated and control units), as shown in Table 2. It is not however possible to disentangle the distribution of Y in the treatment group between the

Observed subgroups $O(T,S)$	Turnover Y	Latent strata
$O(1,1) = \{i : T_i = 1, S_i = 1\}$	OBS	11 or 10
$O(1,0) = \{i : T_i = 1, S_i = 0\}$	·	00
$O(0,1) = \{i : T_i = 0, S_i = 1\}$	OBS	11
$O(0,0) = \{i : T_i = 0, S_i = 0\}$	·	10 or 00

Table 2: Correspondence between observed and latent strata under monotonicity

strata 11 and 10. In this case only nonparametric bounds can be derived, unless some parametric distributional assumptions are introduced. Parametric identification is thus achieved, within the PS framework, thank to the results on finite mixture distribution theory (see e.g., McLachlan and Peel, 2000), although unlike standard mixture models some units have zero probability of belonging to

some strata, which may facilitate disentangling the mixture. Indeed, the likelihood function results in a finite mixture of distributions and identification is straightforward, except when $\pi_{11|x} = \pi_{10|x}$.

In order to alternatively represent the post-treatment complication formalized above using PS, recall the structural model (1)[4]:

$$Y_i = g(t_i, \epsilon_i)$$

$$S_i = h(t_i, \eta_i)$$

where Y_i is observed only if $S_i = 1$. Note that the nonignorability (endogeneity) of S depends on the association between ϵ and η. If it were possible to condition on η, S and ϵ would become conditionally independent, eliminating the endogeneity problem. Because η cannot be observed, we can look for a function of η, $G(\eta)$, called *type of unit* (Imbens, 2006) such that

$$\epsilon \perp\!\!\!\perp S \,|G(\eta).$$

A natural choice for $G(.)$ is a function which is constant over values of η that, for all t, lead to the same value of S:

$$G(\eta) = G(\eta') \qquad \text{if h(t,}\eta\text{) = h(t,}\eta'\text{)} \;\; \forall \text{t}$$

$$G(\eta) \neq G(\eta') \quad \text{if h(t,}\eta\text{)} \neq \text{h(t,}\eta'\text{)} \;\; \text{for some t.}$$

In fact, in this case $\epsilon \perp\!\!\!\perp S|G(\eta)$ by construction. Nonresponse becomes ignorable conditional on the type of unit, so that, once we condition on the value of G, only respondents can be used. This corresponds to finding a common set of units within the Rubin Causal Model, on which proper causal estimands can be defined; with this respect, Principal Stratification represents the coarest choice of the type function. As shown within the PS approach, the comparison of Y between treated and controls is possible only for some values of G, in particular those with $h(0, \eta) = 1$ and $h(1, \eta) = 1$.

Note that, within the structural equation model, in the absence of exclusion restrictions, it would not be possible to nonparametrically point identify treatment effects and most of the semi-parametric versions of selection models would require instruments. In a parametric setting without

[4]The covariates X are omitted, because, as in previous section, we are implicitly conditioning on them, i.e., we are assuming to be within cells defined by the covariates.

IV, identification is achieved by joint normality (or by other parametric specification of the joint distribution), which is a stronger assumption than that required within the PS framework, where only the distributions of Y within strata must be parameterized. When specifying a mixture model, distributional assumptions do not usually involve the joint specification of a model for the outcome and the selection process, but rather refer to distributions for the outcome variables conditional on the strata. These distributions have explicit implications on the probability law of variables within observed groups, in terms of mixture distributions, so that the theory on mixture models can be exploited for both identification and specification testing[5].

5 Identifying causal effect with nonignorable nonresponse on the outcome and an instrumental variable

As said in sections 3 and 4, the lack of nonparametric identification in the presence of endogenous selection is at least partially solved by introducing some exclusion restriction, i.e., by introducing a variable that may serve as an instrument for nonresponse. In what follows, we will consider the availability of an instrumental variable Z, that we will characterize, together with the related assumptions, in terms of its nature, whether it can be regarded as a) an additional intervention or b) as an additional post-treatment variable. Note that in an observational setting all the identifying hypotheses are assumed to hold conditional on the covariates; indeed in the IV literature methods to control for covariates have been introduced because "instruments may require conditioning on a set of covariates to be valid" (Abadie, 2003).

We will use PS in order to define meaningful subpopulations of units, to present identifying assumptions and estimands, although they will be also *translated* into the simultaneous equations framework.

First suppose that, in addition to treatment T, whose causal effect on Y is still our primary inter-

[5]Some simulation results are shown in an extended version of the paper and in Mealli, Pacini (2007), when parametric specifications are exploited to achieve identification. Simulations are aimed at studying the different performances of PS and parametric selection models together with their robustness.

est, units are exposed to an additional treatment which is, in a broad sense to be further specified, related to nonresponse S but unrelated to the outcome Y. In our example, suppose that several persons with a different job task may respond to the phone interview, and define Z the indicator variable which assumes value 1 if an employee responds and 0 if the owner responds.

Because now we have two binary treatments, T and Z, four potential outcomes can be defined for each post-treatment variable, Y and S in our case: $S(t,z)$, $Y(t,z)$ for $t = 0, 1$ and $z = 0, 1$. Both treatments are assumed randomized conditional on a set of pre-treatment covariates (on which we implicitly condition) so that the following hypotheses hold:

$Assumption\ \ PS2:\ T \perp\!\!\!\perp S(0,0), S(0,1), S(1,0), S(1,1), Y(0,0), Y(0,1), Y(1,0), Y(1,1)$

$Assumption\ \ PS3:\ Z \perp\!\!\!\perp S(0,0), S(0,1), S(1,0), S(1,1), Y(0,0), Y(0,1), Y(1,0), Y(1,1)$

In order to characterize Z as an instrument, we impose the following exclusion-restriction type of assumption:

$Assumption\ \ PS4:\ Y(0,0) = Y(0,1)\ \text{ and }\ Y(1,0) = Y(1,1)$

which says that for beneficiaries, and for non beneficiaries, the value assumed by the instrument (i.e., the person who responds to the phone interview) is unrelated to the outcome (turnover).

Because the post-treatment response indicator is also binary, 16 principal strata can be defined, if no other restriction is imposed (see Table 3).

The specific selection problem with one instrument can be alternatively formalized in the simultaneous equations framework as follows:

$$Y_i = g(t_i, \epsilon_i) \tag{3}$$

$$S_i = h(Z_i, t_i, \eta_i)$$

where Y is observed only if $S = 1$ and $Y \not\perp\!\!\!\perp S$. The hypotheses concerning Z and T are stated as:

$Assumption\ \ SM2:\ T \perp\!\!\!\perp (\epsilon, \eta)$

G	S(0,0)	S(0,1)	S(1,0)	S(1,1)
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

Table 3: Principal strata with two binary treatments and a binary intermediate variable

Assumption $SM3-4:\ Z \perp\!\!\!\perp (\epsilon, \eta)$.

Note that Assumption SM3-4 includes both the hypothesis of randomization (Ass. PS3) and the exclusion restriction (Ass. PS4). The classification of units in 16 strata can be retrieved in this framework by defining the type of unit function, G, such that:

$$G(\eta) = G(\eta') \qquad \text{if } \mathrm{h(z,t,}\eta\mathrm{)} = \mathrm{h(z,t,}\eta'\mathrm{)} \ \ \forall \mathrm{z,t}$$

$$G(\eta) \neq G(\eta') \quad \text{if } \mathrm{h(z,t,}\eta\mathrm{)} \neq \mathrm{h(z,t,}\eta'\mathrm{)} \ \text{ for some z,t.}$$

Let's now analyze how the presence of an instrument can be exploited to achieve identification of some causal estimands. Using the potential outcome language, assume that the instrument is *perfect*, in the sense that, for certain level of the instrument Z, S is always equal to 1, i.e.:

Assumption PS5 : $S(0,1) = 1$ and $S(1,1) = 1$.

This additional assumption, which can be easily verified from the data[6], removes some of the strata ($1-5$,7,$9-13$,15); the remaining 4 strata (see Table 4) allows to point identify the marginal distribution of $Y(T=0)$ and $Y(T=1)$, so that all estimands involving only these marginals, as the ATE (Average Treatment Effect), can be identified.

G	S(0,0)	S(0,1)	S(1,0)	S(1,1)
6	0	1	0	1
8	0	1	1	1
14	1	1	0	1
16	1	1	1	1

Table 4: Principal strata with two binary treatments and a binary intermediate variable, under Assumption PS5

Indeed, Assumption PS4 implies that $f(Y(0,1)) = f(Y(0,0)) = f(Y(T=0))$ and $f(Y(1,0)) = f(Y(1,1)) = f(Y(T=1))$. As we can see from Table 4, $Y(0,1)$ and $Y(1,1)$ can always be observed in all the four strata due to the *perfect* instrument assumption, so that $f(Y(T=0))$ and $f(Y(T=1))$ can be identified and estimate using observations where $Z=1$.

The same result can be obtained following Manski (2003); the goal there is how to identify the marginal distribution of a variable which is subject to a nonignorable missing mechanism. This result turns out to be useful also in a causal inference framework where usually some features of the distribution of two variables, namely $Y(T=0)$ and $Y(T=1)$, are of interest: the properties of the instrumental variable Z allow to estimate the two marginal distributions separately[7].

Furthermore, note that, even if not stated precisely in the same terms, the use of a *key variable*,

[6]With this respect, note that one may choose or construct an instrumental variable having such a property.

[7]Manski's result can be shown as follows. Suppose to have three variables Y, S, and Z, and suppose the distribution of Y, $F(Y)$ is of interest. A random sample from the population is available, however Y is observed only if $S=1$, and $F(Y) \neq F(Y|S=1)$. The following conditional distributions can be estimated from observed data: $F(Y,Z|S=1)$, $F(Z|S=1)$, and $F(Z|S=0)$. Assuming that $F(Y|Z) = F(Y)$ (exclusion restriction) and $P(S=1|Z=1) = 1$ (*perfect* instrument), $F(Y)$ can be identified.

with the same features of our Z, has been exploited by Frangakis et al. (2007) within a Principal Stratification Design, in order to draw valid inference for the marginal distribution of some input data nonignorably missing (missing due to death).

5.1 Alternative structural assumptions

Finding or constructing an instrumental variable that, in addition to the exclusion restriction, is perfect in the sense of Assumption PS5 can be a difficult task and sometimes cannot be defendable empirically. Alternative identification assumptions can be stated as forms of monotonicity of S:

$$Assumption\ \ PS6:\ S(t,0) \leq S(t,1)\ \ \forall t$$

$$Assumption\ \ PS7:\ S(0,z) \leq S(1,z)\ \ \forall z.$$

Assumption PS6 is a weaker behavioral assumption than Assumption PS5 and relates to the response behaviour w.r.t. the instrument: for a fixed treatment level, units responding when $Z = 0$ would respond also when $Z = 1$. Analogously, Assumption PS7 relates to the response behaviour w.r.t. the treatment: for a fixed value of the instrument, units responding when under control would respond also when treated. As far as the structural model is concerned, these monotonicity assumptions refer to a specific constraint on the function $h(.)$ for fixed values of η (Imbens, 2006):

$$Assumption\ \ SM6:\ h(t,0,\eta) \leq h(t,1,\eta)\ \ \forall t, \forall \eta$$
$$Assumption\ \ SM7:\ h(0,z,\eta) \leq h(1,z,\eta)\ \ \forall z, \forall \eta.$$

The two assumptions imply the non-existence of some of the 16 strata in Table 3. As should be clear from Table 5, it is not possible to identify the proportions of the strata, except for the first and the last ones[8]. Analogously, only for the last stratum we can estimate the distribution of $Y(T = 0)$, because the distribution of $Y(T = 0, Z = 0)$ is identified and is equal to $Y(T = 0, Z = 1)$, thank to

[8]Note, in fact, that the proportion of stratum 0000 can be estimated by the proportion of non-respondents within the observed group where $T = 1$ and $Z = 1$; whereas the proportion of stratum 1111 can be estimated by the proportion of respondents within the observed group where $T = 0$ and $Z = 0$.

G	S(0,0)	S(0,1)	S(1,0)	S(1,1)
1	0	0	0	0
2	0	0	0	1
4	0	0	1	1
6	0	1	0	1
8	0	1	1	1
16	1	1	1	1

Table 5: Principal strata with two binary treatments and a binary intermediate variable, under Assumptions PS6 and PS7

the exclusion restriction (Ass. PS4). Additional assumptions are required in order to either reduce the number of strata or state the equivalence of the distribution of Y across some strata.

Here we propose the following set of identifying assumptions:

*Assumption PS*8 : Stratum 0001 does not exists

which states that it is not possible to react to the instrument only if treated: if one reacts under treatment, the same must be true under control.

*Assumption PS*9 : $Y(1,0) \perp\!\!\!\perp S(1,0) | S(0,0) = 0, S(0,1) = 1, S(1,1) = 1$

which assumes the (latent) ignorability of nonresponse for a union of strata, having the same response behaviour for all the combinations of values of Z and T but one. Under previous assumptions, this amounts to state that the distribution of $Y(1,0)$ is the same within stratum 6 or 8.

We will now show that, under previously stated assumptions (PS1-PS4 and PS6-PS9), one can identify and estimate the causal effect of T on a specific subset of units, namely those *reacting to the instrument* under control and/or under treatment, which are precisely strata 6 and 8[9].

First of all, note that we are now left with 5 strata ($G = 1, 4, 6, 8, 16$); there are thus 4 free

[9]The same estimator estimates the effect only for latent stratum 6 if we assume $Y(0,1) \perp\!\!\!\perp S(1,0) | S(0,0) = 0, S(0,1) = 1, S(1,1) = 1$ instead of PS9.

of variation strata proportions to be estimated (π_4, π_6, π_8, π_{16}). These can indeed be identified because we have 4 sampling response proportions ($\hat{p}(S = 1|T = 1, Z = 1)$, $\hat{p}(S = 1|T = 0, Z = 0)$, $\hat{p}(S = 1|T = 0, Z = 1)$, $\hat{p}(S = 1|T = 1, Z = 0)$); estimation can be performed by a method of moment strategy, i.e. equating empirical and theoretical proportions (e.g.,$\hat{p}(S = 1|T = 1, Z = 1) = \pi_4 + \pi_6 + \pi_8 + \pi_{16}$).

For sake of simplicity, from now on we focus on the estimation of means and thus on Average Treatment Effects. Under control we can estimate the following mean $\bar{Y}(t, z|S = 1)$, where $\bar{Y}(0, 1|S = 1)$ is the mean of Y for respondents with $T = t$ and $Z = z$, which can be written as

$$\frac{\bar{Y}(0, 1|G = 6)\pi_6 + \bar{Y}(0, 1|G = 8)\pi_8 + \bar{Y}(0, 1|G = 16)\pi_{16}}{\pi_6 + \pi_8 + \pi_{16}}.$$

We can also estimate $\bar{Y}(0, 0|S = 1)$, which is equal to $\bar{Y}(0, 0|G = 16)$ and also to, by the exclusion restriction, $\bar{Y}(0, 1|G = 16)$. By estimating these quantities and the strata proportions, we can, by difference, obtain an estimate of $\bar{Y}(0, 1|G = 6)\pi_6 + \bar{Y}(0, 1|G = 8)\pi_8$.
Similarly, under treatment we can estimate the following means:

$$\bar{Y}(1, 0|S = 1) = \frac{\bar{Y}(1, 0|G = 4)\pi_4 + \bar{Y}(1, 0|G = 8)\pi_8 + \bar{Y}(1, 0|G = 16)\pi_{16}}{\pi_4 + \pi_8 + \pi_{16}}$$

and $\bar{Y}(1, 1|S = 1) =$

$$\frac{\bar{Y}(1, 1|G = 4)\pi_4 + \bar{Y}(1, 1|G = 8)\pi_8 + \bar{Y}(1, 1|G = 6)\pi_6 + \bar{Y}(1, 1|G = 16)\pi_{16}}{\pi_4 + \pi_6 + \pi_8 + \pi_{16}}$$

which, by Assumption 9, is equal to:

$$\frac{\bar{Y}(1, 1|G = 4)\pi_4 + \bar{Y}(1, 1|G = 8)(\pi_8 + \pi_6) + \bar{Y}(1, 1|G = 16)\pi_{16}}{\pi_4 + \pi_6 + \pi_8 + \pi_{16}}.$$

By difference, we can now obtain an estimate of $\bar{Y}(1, 1|G = 8)$ which can be contrasted to

$$\frac{\bar{Y}(0, 1|G = 6)\pi_6 + \bar{Y}(0, 1|G = 8)\pi_8}{\pi_6 + \pi_8}$$

to obtain an estimate of the average causal effect of T within strata 6 and 8 (i.e, for units reacting to the instrument), $E[Y(T = 1) - Y(T = 0)|G \in \{6, 8\}]$:

$$\frac{\hat{\bar{Y}}(1, 1|S = 1) \cdot (\hat{\pi}_4 + \hat{\pi}_6 + \hat{\pi}_8 + \hat{\pi}_{16}) - \hat{\bar{Y}}(1, 0|S = 1) \cdot (\hat{\pi}_4 + \hat{\pi}_8 + \hat{\pi}_{16})}{\hat{\pi}_6}$$

$$-\frac{\hat{\bar{Y}}(0,1|S=1)\cdot(\hat{\pi}_6+\hat{\pi}_8+\hat{\pi}_{16})-\hat{\bar{Y}}(0,0|S=1)\cdot\hat{\pi}_{16}}{\hat{\pi}_6+\hat{\pi}_8}$$

where $\hat{\bar{Y}}(t,z|S=1)$ is the sample average of $\bar{Y}$ for respondents with treatment $T=t$ and instrument $Z=z$.

Assumptions PS8 and PS9 may find a correspondence in terms of assumptions within the structural model approach. However, the assumptions would translate into specific constraints on the values of $h(.)$ and specific conditional independence hypotheses between η and ϵ that would not have a natural behavioral interpretation in this framework and would, thus, not easily suggested.

An assumption which could be an alternative to Assumption PS8, with a proper meaning within the structural approach, is a sort of monotonicity of $h(.)$ w.r.t. η:

$$Assumption\ \ SM8:\ h(t,z,\eta)\leq h(t,z,\eta')\ \ \forall t,z,\ \ \eta\leq\eta'.$$

Within the Principal Stratification framework this would be equivalent to introduce a *natural* ordering of the strata and eliminate the strata contradicting the order; for example, in Table 4 strata 4 and 6 could not simultaneously exist. Supposing that stratum 6 does not exist and formulating Assumption PS9 as $Y(1,0)\perp\!\!\!\perp S(1,0)|S(0,0)=0,S(0,1)=0,S(1,1)=1$, it could be easily shown that identification is now achieved for the causal effect of T for strata 8 and 16, i.e, for *the always respondents and the units reacting to the instrument under control*. In this case the estimator is equal to:

$$\hat{E}[Y(T=1)-Y(T=0)|G\in\{8,16\}]=$$

$$\frac{\hat{\bar{Y}}(1,0|S=1)\cdot(\hat{\pi}_4+\hat{\pi}_8+\hat{\pi}_{16})-[\hat{\bar{Y}}(1,1|S=1)\cdot(\hat{\pi}_2+\hat{\pi}_4+\hat{\pi}_8+\hat{\pi}_{16})-\hat{\bar{Y}}(1,0|S=1)\cdot(\hat{\pi}_4+\hat{\pi}_8+\hat{\pi}_{16})]\cdot\frac{\hat{\pi}_4}{\hat{\pi}_2}}{\hat{\pi}_8+\hat{\pi}_{16}}$$

$$-\hat{\bar{Y}}(0,1|S=1).$$

Alternatively, supposing that stratum 4 does not exist and formulating Assumption PS9 as $Y(0,0)\perp\!\!\!\perp S(0,0)|S(1,0)=1,S(0,1)=1,S(1,1)=1$, identification is again achieved for the causal effect of T for strata 8 and 16, but the estimator in this case simply:

$$\hat{\bar{Y}}(1,0|S=1)-\hat{\bar{Y}}(0,0|S=1).$$

Choosing between Assumption PS8 and SM8 is a subject matter; it can be also suggested by the data. Indeed, as well known in a causal inference problem, the identification strategy depends on the plausibility of the assumptions with respect to the specific empirical context.

5.2 Alternative assumptions for the instrument

So far, the instrument Z has been considered as an additional treatment, which is thus subject to assume any possible value, and its value could be under the control of an experimenter (see Rubin, 2004, and related discussion).

Consider now Z as an additional post-treatment variable, that precedes nonresponse and can be added to the other potential outcome variables. Because we now have only one binary treatment, T, two potential outcomes can be defined for each post-treatment variable, Z, S, and Y: $Z(t)$, $S(t)$, $Y(t)$ for $t = 0, 1$. The treatment is assumed randomized conditional on a set of pre-treatment covariates (on which we implicitly condition) so that the following hypothesis hold:

Assumption PS10 : $T \perp\!\!\!\perp Z(0), Z(1), S(0), S(1), Y(0), Y(1)$.

In order to characterize Z as an instrument, we impose the following exclusion-restriction type of assumption:

Assumption PS11 : $Y(0) \perp\!\!\!\perp Z(0)$ and $Y(1) \perp\!\!\!\perp Z(1)$.

Principal strata are now defined as the joint value of the 4 intermediate potential variables ($Z(0)$, $Z(1)$, $S(0)$, and $S(1)$) as shown in Table 6.

The specific selection problem with one instrument can be alternatively formalized in the simultaneous equations framework as follows:

$$Y_i = g(t_i, \epsilon_i) \tag{4}$$

$$S_i = h(Z_i, t_i, \eta_i)$$

$$Z_i = m(t_i, \nu_i)$$

G	Z(0)	Z(1)	S(0)	S(1)
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

Table 6: Principal strata with one binary treatments and two binary intermediate variables

where Y is observed only if $S = 1$ and $Y \not\perp\!\!\!\perp S$. The hypotheses concerning Z and T are stated as:

$Assumption\ \ SM10:\ T \perp\!\!\!\perp (\epsilon, \eta, \nu)$

$Assumption\ \ SM11:\ Z \perp\!\!\!\perp (\epsilon)$.

Note that Assumption SM11 corresponds to the exclusion restriction (Ass. PS11). The new classification of units in 16 strata can be retrieved by defining the type of unit function, G, such that:

$$G(\eta, \nu) = G(\eta', \nu') \qquad \text{if h(z, t, } \eta) = \text{h(z, t, } \eta') \text{ and m(t, } \nu) = \text{m(t, } \nu') \ \forall \text{z, t}$$

$$G(\eta, \nu) \neq G(\eta', \nu') \quad \text{if h(z, t, } \eta) \neq \text{h(z, t, } \eta') \text{ or m(t, } \nu) \neq \text{m(t, } \nu') \text{ for some z, t}$$

Some identifying assumptions are required in order to exploit the information gathered from the

instrument. The first one is again a monotonicity hypothesis of the instrument with respect to the treatment:

$$Assumption\ \ PS12:\ Z(0) \leq Z(1)$$

which eliminates strata 9 to 12 and can be alternatively stated in terms of the structural model as follows:

$$Assumption\ \ SM12:\ m(0,\nu) \leq m(1,\nu)\ \ \forall \nu.$$

An additional identifying assumption, concerning the response behavior, is a type of latent ignorability (Frangakis, Rubin 1999):

$$Assumption\ \ PS13:\ Y(0) \perp\!\!\!\perp S(0)|Z(0)=k, Z(1)=h\ \ \forall h,k;\ Y(1) \perp\!\!\!\perp S(1)|Z(0)=k, Z(1)=h\ \ \forall h,k.$$

which states that, within strata defined only by the joint values of the two potential outcomes $Z(0)$ and $Z(1)$, nonresponse is ignorable.

For model (4), the following condition holds by contruction, $\epsilon \perp\!\!\!\perp S|G(\eta,\nu)$; while latent ignorability can be formalized by defining a type of unit function such that:

$$\begin{aligned} G(\nu) &= G(\nu') && \text{if } \mathrm{m(t,}\nu\mathrm{)} = \mathrm{m(t,}\nu'\mathrm{)}\ \forall \mathrm{t} \\ G(\nu) &\neq G(\nu') && \text{if } \mathrm{m(t,}\nu\mathrm{)} \neq \mathrm{m(t,}\nu'\mathrm{)} \text{ for some t} \end{aligned}$$

and assuming:

$$Assumption\ \ SM13:\ \epsilon \perp\!\!\!\perp S|G(\nu).$$

Assumption PS13 states that, within strata defined only by the two potential values of the instrument Z, the nonresponse mechanism is ignorable, so that, if we could observe both values of Z, given those values the distribution of Y among respondents and nonrespondents would be the same. Note that the conditioning that induces independence between outcome and nonresponse is neither on ν (in the structural setting), nor on the observed values of Z, say $Z(t_i)$, but rather on specific subsets of values of ν, i.e., on the joint values of $Z(0)$ and $Z(1)$.

Indeed, conditional independence given ν, together with Assumption SM11, would also imply marginal independence of ϵ and S, so that Assumption PS13 would correspond to an assumption of (not latent) ignorable nonresponse, thus removing the motivating complication of nonignorability[10].

If assumption PS13 holds, we can then concentrate only on the 3 latent strata defined by Z. To show how these assumptions allow identifiability of a properly defined causal effect, consider the following equality:

$$f(Y(1)|Z(1) = 0, S(1) = 1) = f(Y(1)|Z(1) = 0, Z(0) = 0, S(1) = 1)$$

which holds because Assumption PS12 implies that if $Z(1) = 0$ then $Z(0) = 0$. By Assumptions PS13 and PS11 we also have:

$$f(Y(1)|Z(1) = 0, Z(0) = 0, S(1) = 1) = f(Y(1)|Z(1) = 0, Z(0) = 0)$$

$$= f(Y(1)|Z(1) = 0) = f(Y(1)).$$

The last two equalities show that respondents with $Z(1) = 0$ can be used to estimate $f(Y(1))$. Analogously, respondents with $Z(0) = 1$ can be used to estimate $f(Y(0))$. So, focussing on the average causal effect, an estimator of $E[Y(1) - Y(0)]$ is the following:

$$\hat{\bar{Y}}(1|Z = 0, S = 1) - \hat{\bar{Y}}(0|Z = 1, S = 1).$$

A numerical example is presented in the Appendix, which shows how the different assumptions may coexist and how identification is achieved.

6 Empirical illustration

We analyze a small intervention in the form of financial aids to Tuscan small and medium handicraft enterprises. In the framework of the Programs for the Development of Crafts in Tuscany (Regional Law n. 36, 4/4/95), we consider the effects of interest-relief grants on investments (PSA

[10]Note that $\epsilon \perp\!\!\!\perp S|\nu \equiv \epsilon \perp\!\!\!\perp \eta|\nu$ and $Z \perp\!\!\!\perp \epsilon \equiv \nu \perp\!\!\!\perp \epsilon$; so assuming $\epsilon \perp\!\!\!\perp \eta|\nu$, together with $\nu \perp\!\!\!\perp \epsilon$, implies $\epsilon \perp\!\!\!\perp \eta$ (Dawid, 1979).

2001-2002), delivered in years 2001 and 2002, on turnover in year 2005. Data are obtained by integrating different data sources: administrative archives (ASIA, 2001-2004; IRAP, 2001-2003), data on firms receiving aids collected by ArtigianCredito, and data from an "ad hoc" survey (Mauro, Mattei, 2007). The survey was conducted in order to gather additional information on firms not contained in administrative archives, in particular outcome variables on firms'performances (turnover, number of employees, production innovation). The survey was designed using propensity score matching (Rubin, 2001), in order to select a sample of firms not receiving aids (controls) as similar as possible to treated firms. Evidence on the effect of the program on some dimensions of firms' performances, not affected by missingness, such as the number of employees, suggests that the effect is negligible (Mauro, Mattei, 2007). This result is not surprising, given the limited amount of investments (50000 euros on average) and an interest relief of approximately 5% of the investment. Turnover might be a variable more sensitive to such an intervention, which is however affected by a high degree of nonresponse. We consider a subsample consisting of 101 treated firms and 101 controls: the sample of controls was constructed using nearest neighbor propensity score matching and resulted in a sample of firms with the same distribution of pre-treatment covariates as the sample of treated firms. There is some evidence that nonresponse may be nonignorable: nonresponse is in fact highly correlated with pre-treatment turnover, so we may suspect that it is the same with post-treatment turnover, even conditional on observed covariates. In order to estimate the program effect in the presence on nonresponse, we exploit the variable Z, the indicator variable which assumes value 1 if an employee responds to the phone interview and 0 if the owner responds, as an instrument for nonresponse. In Table 7, the relevant sample quantities are reported: the chosen instrument does play a role in determining nonresponse, in particular under control ($\hat{p}(S = 1|T = 0, Z = 1) > \hat{p}(S = 1|T = 0, Z = 0)$). Moreover, monotonicity of nonresponse with respect to both the treatment and the instrument is not contradicted by the data.

The way sample averages $\hat{\bar{Y}}(T, Z|S = 1)$ are used to obtain an estimate of some causal effect depends on the causal estimand and the additional identifying assumptions, in particular those related to the nonexistence of some strata. In Table 8 we report the estimated strata proportions under different identifying assumptions. In all the configurations considered, the sample proportions sug-

$\hat{\bar{Y}}(T=0\|S=1)$	414078
$\hat{\bar{Y}}(T=1\|S=1)$	451132
Diff	37054
$\hat{\bar{Y}}(T=0,Z=0\|S=1)$	362273
$\hat{\bar{Y}}(T=1,Z=0\|S=1)$	416327
$\hat{\bar{Y}}(T=0,Z=1\|S=1)$	676554
$\hat{\bar{Y}}(T=1,Z=1\|S=1)$	686734
$\hat{p}(S=1\|T=1,Z=1)$	0.722
$\hat{p}(S=1\|T=1,Z=0)$	0.710
$\hat{p}(S=1\|T=0,Z=1)$	0.714
$\hat{p}(S=1\|T=0,Z=0)$	0.628

Table 7: Relevant sample quantities: Y is 2005 turnover in euros, T is the program indicator, Z is the instrumental variable and S is the response indicator

gest that stratum 4 does not exist or is negligible, while data are consistent with the existence of strata 2 and 6.

We therefore estimate the average effect in strata $G \in \{8, 16\}$ under the assumptions PS1-PS4, PS6-PS7, nonexistence of stratum 0011 and $Y(0,0) \perp\!\!\!\perp S(0,0) | S(1,0) = 1, S(0,1) = 1, S(1,1) = 1$:

$$\hat{\bar{Y}}(1,0|S=1) - \hat{\bar{Y}}(0,0|S=1) = 54054 \text{ euros.}$$

The estimated effect must be interpreted as an average effect *for the always respondents and for firms reacting to the instrument under control*. It is slightly larger than the effect estimated by improperly comparing respondents under treatment and under control (Diff in Table 7).

7 Concluding remarks

In this paper we have dealt with the problem of a nonignorable nonresponse on an outcome variable, on which a causal effect of a treatment is of interest. Identification issues have been studied

	$\pi_2 = 0$	$\pi_4 = 0$	$\pi_6 = 0$
$\hat{\pi}_1$	0.278	0.278	0.273
$\hat{\pi}_2$	0	0.008	0.013
$\hat{\pi}_4$	0.008	0	0
$\hat{\pi}_6$	0.013	0.005	0
$\hat{\pi}_8$	0.074	0.082	0.086
$\hat{\pi}_{16}$	0.628	0.628	0.628

Table 8: Estimated strata proportions under different identifying assumptions

without and with an "instrument" for nonresponse. In the latter case, assumptions that characterize the instrument and allow identification of some causal estimands have been proposed, which are not the standard assumptions used in an IV setting with endogenuous regressors (Angrist et al., 1996, Abadie, 2003). An empirical analysis based on a real case study is currently being conducted, which has inspired the simplified evaluation setting considered here as a motivating example.

Results were derived within the PS framework, where the latent strata are generated by the primitive potential outcomes. Identification strategies exploit the comparison between observed groups and latent groups (strata). This comparison can sometimes imply only bounds for treatment effects (Zhang and Rubin, 2003; Imbens and Rubin, 1997); point identification can instead be reached by means of assumptions that usually relate to specific behavioral hypotheses about the strata. Some of such assumptions aim at reducing the number of strata; other hypotheses impose certain features of the distribution of outcomes within or among strata: these may include various forms of exclusion restrictions (Mealli *et al.*, 2004), various versions of stochastic dominance that assume, for example, that the distribution of the outcome in one or more strata stochastically dominate that of other strata (Zhang and Rubin, 2003; Zhang *et al.*, 2006), and various forms of ignorability and nonignorability for the selection mechanism (Frangakis and Rubin, 1999).

In an observational setting, most of the assumptions would be conditional on covariates, so that methods to accomodate covariates are required, because in finite samples it is infeasible to work

within cells defined by the covariates, in particular if they are continuous. These methods may include either flexible parametric specifications (Hirano et al., 2000; Grilli and Mealli, 2007) or be semiparametric as in Abadie et al. (2002), Abadie (2003), and Frolich (2007).

Indeed, identification and efficiency improvements could also be achieved by exploiting covariates: plausible behavioral hypotheses *within or among groups* defined by the values of the covariates can be translated into restrictions on coefficients *within or among strata* (Jo, 2002; Frangakis, 2006). However, the assumptions embedded in a parametric model, derived under PS, are more explicit, in terms of the behavior of units, than the ones characterizing structural models.

Using PS, whatever the assumptions made, the result of inference is always a causal effect within one or more strata. An issue that often arises regarding the PS approach is that we cannot univocally identify the group the causal effect refers to, so we cannot univocally estimate the individual causal effects. This issue also characterizes the Instrumental Variable literature where, under certain assumptions, only the effect on specific subpopulations can be identified (Angrist et al., 1996). Note, however, that the fact that proper causal effects can only be defined and estimated for latent subgroups of units is a limitation created by the selection mechanism, rather than a drawback of the framework of principal stratification.

Appendix

In this appendix we propose a numerical example, which illustrates the case presented in Section 5.2; we consider a nonignorable nonresponse mechanism for the outcome and an instrumental variable that is assumed as an additional post-treatment variable that preceeds nonresponse and can be added to the other potential outcome variables. Assume that $E(Y(1)) = 4$ and $E(Y(0)) = 3$, so that the estimand ATE is equal to 1; the outcome variable is subject to nonignorable nonreponse, $E(Y(1)|S(1) = 1) = 3.84 \neq E(Y(1)|S(1) = 0) = 4.16$ and $E(Y(0)|S(0) = 1) = 2.84 \neq E(Y(0)|S(0) = 0) = 3.16$. In Table 9 we describe in terms of principal strata a plausible setting that, under Assumptions 13, 14, 17, and 19, may have generated the figures stated above.

Assumption 17 is verified because $Z(0) \leq Z(1)$. Assumption 14 holds because $E(Y(0)|Z(0) =$

$\pi(Z(0),Z(1))$	Z(0)	Z(1)	$\pi(S(0),S(1)\|Z(0),Z(1))$	S(0)	S(1)	Y(0)	Y(1)
0.2	0	0	0.25	0	0	1	3
			0.25	0	1	1	4
			0.25	1	0	2	5
			0.25	1	1	0	4
0.4	0	1	0.1	0	0	3.4	0
			0.2	0	1	4.8	6
			0.3	1	0	2.5	4
			0.4	1	1	7	1.5
0.4	1	1	0.4	0	0	3.4	3.5
			0.3	0	1	2.8	6
			0.2	1	0	5	8
			0.1	1	1	1.5	2

Table 9: Principal strata, principal strata proportions and average outcome values

$0) = \frac{0.2\times1+0.4\times4}{0.6} = 3 = E(Y(0)|Z(0) = 1)$, where, e.g., $1 = 0.25\times0+0.25\times1+0.25\times2+0.25\times1 = E(Y(0)|Z(0) = 0, Z(1) = 0)$. Analogously, $E(Y(1)|Z(1) = 1) = 4 = E(Y(1)|Z(1) = 0)$. Further, note that nonresponse rate varies with Z. While nonresponse is nonignorable, latent ignorability (Ass. 19) can be easily verified to hold, e.g., $E(Y(1)|S(1) = 0, Z(0) = 0, Z(1) = 0) = \frac{0.25\times3+0.25\times5}{0.5} = 4 = E(Y(1)|S(1) = 1, Z(0) = 0, Z(1) = 0)$. Under treatment, using respondents with $Z(1) = 0$, $E(Y(1))$ is estimated to be 4, while, under control, using respondents with $Z(0) = 1$, $E(Y(0))$ is estimated to be 3, and so the ATE correctly estimated as 1.

References

[1] Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models, Journal of Econometrics, 113, 231–263.

[2] Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimates of the effects of subsidized training on the quantiles of trainee earnings, Econometrica, 70, 91–117.

[3] Ahn, H. and Powell, J. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism, Journal of Econometrics, 58, 329.

[4] Altonji, J., Matzkin, R. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors, Econometrica, 73, 10531103.

[5] Angrist, J.D., Imbens, G. W. and Rubin, D.B. (1996). Identification of causal effects using instrumental variables, (with discussion), Journal of the American Statistical Association, 91, 444–472.

[6] Barnard, J., Frangakis, C.E., Hill, J.L., and Rubin, D.B. (2003). A Principal Stratification approach to broken randomized experiments: a case study of School Choice vouchers in New York City, Journal of the American Statistical Association (with discussion), 98: 299-323.

[7] Chernozhukov, V., Hansen, C. (2005). An IV model of quantile treatment effects, Econometrica, 73, 245-261.

[8] Chernozhukov, V., Imbens, G., Newey, W. (2005). Instrumental variables identification and estimation via quantile conditions, Journal of Econometrics, forthcoming.

[9] Chesher, A. (2003). Identification in nonseparable models, Econometrica, 71, 1405-1443.

[10] Chesher, A. (2005). Nonparametric identification under discrete variation, Econometrica, 73, 1525-1550.

[11] Choi, K. (1990). The semiparametric estimation of the sample selection model using series expansion and the propensity score, mimeo, University of Chicago.

[12] Christofides, L.N., Li, Q., Liu, Z. and Min, I. (2003). Recent two-stage sample selection procedures with an application to the gender wage gap, Journal of Business & Economic Statistics, 21: 396–405

[13] Copas, J. B. and Li, H. G. (1997). Inference for non-random samples (with discussion), Journal of the Royal Statistical Society, Series B, 59: 55–95.

[14] Cosslet, S. (1991). Semiparametric estimation of a regression model with sample selectivity, in: Nonparametric and semiparametric methods in econometrics and statistics, W.A. Barnett, J. Powell and G.Tauchen (eds.), Cambridge University Press.

[15] Das, M. and Newey, W.K. and Vella, F. (2003). Nonparametric estimation of sample selection models, Review od Economic Studies, 52: 33–58.

[16] Dawid, A. P. (1979). Conditional independence in statistical theory, Journal of the Royal Statistical Society, Series B 41 1–31.

[17] Florens, J.P., Heckman, J., Meghir, C., Vytlacil, E. (2003). Instrumental variables, local instrumental variables and control functions, Working Paper, IDEI, Toulose.

[18] Forcina A. (2006). Causal effects in the presence of non compliance: a latent variable interpretation, Metron, LXIV : 1–7.

[19] Frangakis, C.E. (2006). Comment on Forcina (2006), Metron, LXIV : 8–14.

[20] Frangakis, C. E. and Rubin, D. B. (1999). Addressing Complications of Intention-to-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncomplianze and Subsequent Missing Outcomes, Biometrika, 86: 365–379.

[21] Frangakis, C.E. and Rubin, D.B.(2002). The defining role of "principal stratification and effects" for comparing treatments adjusted for posttreatment variables: from treatment noncompliance to surrogate endpoints, Biometrics, 58: 191–199.

[22] Frangakis, C.E., Rubin, D.B., An M., and MacKenzie E. (2007). Principal Stratification Designs to Estimate Input Data Missing Due to Death, Biometrics, 63 (3): 641–649.

[23] Frolich, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates, Journal of Econometrics, 139, 35–75.

[24] Gallant, A. and Nychka, D. (1987). Seminonparametric maximum likelihood estimation, Econometrica, 55: 363–390.

[25] Grilli, L. and Mealli, F. (2007). Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification, Journal of Educational and Behavioral Statistics, forthcoming.

[26] Gronau, R. (1974). Wage comparison - A selectivity bias, The Journal of Political Economy, 82: 1119-1143.

[27] Heckman, J. (1974). Shadow prices, market wages, and labor supply, Econometrica, 42: 679–694.

[28] Heckman, J. (1990). Varieties of selection bias, American Economic Review, 80: 313–318.

[29] Heckman, J., Vytlacil, E. (1999). Local instrumental variables, Nonlinear statistical inference: Essays in honour of Takeshi Amemiya, C. Hsiao, K. Morimune and J. powell (eds.), Cambridge University Press.

[30] Hirano, K., Imbens, G., Rubin, D. and Zhou, X. (2000). Assessing the effect of an influenza vaccine in an encouragement design, Biostatistics, 1: 69–88.

[31] Hoderlein S. (2005). Nonparametric demand systems, instrumental variables and an heterogeneous population, Working Paper; Mannheim.

[32] Hoderlein S. and Mammen E. (2007). Identification of marginal effects in nonseparable models without monotonicity, Econometrica, 75, 5: 1513–1518.

[33] Holland, P. (1986). Statistics and causal inference, Journal of American Statistical Association, 81: 945–970.

[34] Honore, B. E., Kyriazidou, E, Udry, C. (1997). Estimation of type-3 tobit models using symmetrics trimming and pairwise comparisons, Journal of Econometrics, 76: 107–128.

[35] Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications (with comment and rejoinder), Journal of Educational and Behavioral Statistics, 27, 385420.

[36] Klein, R. and Spady, R. (1993). An efficient semiparametric estimator for binary response models, Econometrica, 61: 387–421.

[37] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, Journal of Econometrics, 58: 71–120.

[38] Imbens, G.W. (2006). Nonadditive models with endogenouus regressors, mimeo.

[39] Imbens, G.W., Newey, W. (2003). Identification and estimation of triangular simultaneous equations models without additivity, Working Paper, MIT.

[40] Imbens, G.W. and Rubin, D.B. (1997). Bayesian inference for causal eects in randomized experiments with noncompliance, Annals of Statistics, 25: 305–327.

[41] Lee, L.F. (1982). Some approaches in the correction of selectivity bias, Review of Economic Studies, 49: 355–372.

[42] Lee, L.F. (1983). Generalized econometric models with selectivity, Econometrica, 51: 507–512.

[43] Lee, L.F. (1994). Semiparametric two-stage estimation of sample selection models subject to tobit-type selection rules, Journal of Econometrics, 61: 305–344.

[44] Li, Q., Wooldridge, J. (2002). Semiparametric estimationof partially linear models for dependent data with generated regressors, Econometic Theory, 18: 625–645.

[45] Little, R.J.A. (1985). A note about models for selectivity bias, Econometrica, 53, 1569–1474.

[46] McLachlan, G., Peel, D. (2000). Finite Mixture Models, Wiley series in probability and statistics.

[47] Manski, C.F. (2003). Partial Identication of Probability Distributions, Springer-Verlag.

[48] Mattei, A. and Mauro, V. (2007). Valutazione di politiche per le imprese artigiane, Rapporto di ricerca, IRPET.

[49] Mattei, A. and Mealli, F. (2007). Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-Examination, Biometrics, 63, 437-446.

[50] Mealli, F. and Rubin, D. B. (2003). Assumptions allowing the estimation of direct causal eects: Commentary on "health, wealth, and wise? test for direct causal paths between health and socioeconomic status" by Adams et al., Journal of Econometrics, 112: 79–87.

[51] Mealli F., G. Imbens, S. Ferro, A. Biggeri (2004). Analyzing a Randomized Trial on Breast Self Examination with Noncompliance and Missing Outcomes, Biostatistics, 5: 207-222.

[52] Newey, W. (1990). Semiparametric efficiency bounds, Journal of Applied Econometrics, 5: 99–135.

[53] Pagan, A. and Ullah, A. (1997). Nonparametric econometrics, Cambridge University Press.

[54] Pagan, A. and Vella, F. (1989). Diagnostic tests for models based on individual data: a survey, Journal of Applied Econometrics, 4: S29-S59.

[55] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients, Econometrica, 57(6): 1403-1430.

[56] Robinson, P.M. (1988). Root n-consistent semiparametric regression, Econometrica, 56: 931–954.

[57] Rosembaum, P. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment, Journal of the Royal Statistical Society, Series A, 147: 656–666.

[58] Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, Journal of Educational Psycology, 66: 688-701.

[59] Rubin, D.B. (1978). Bayesian inference for causal effects, Annals of Statistics, 6: 34–58.

[60] Rubin, D.B. (1980). Discussion of "Randomization analysis of experimental data: the Fisher randomization test" by D. Basu, Journal of the American Statistical Association, 75, 591-593.

[61] Rubin, D.B. (2001). Using Propensity Scores to Help Design Observa- tional Studies: Application to the Tobacco Litigation, Health Services and Outcomes Research Methodology, 2 169-188.

[62] Rubin, D.B. (2004). Direct and indirect causal effects via potential outcomes (with discussion), Scandinavian Journal of Statistics, 31: 161–170.

[63] Vella, F. (1998). Estimating models with sample selection bias: A survey, The Journal of Human Resources, 33: 127–169.

[64] Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result, Econometrica, 70: 331341.

[65] Wooldridge, J.M. (2002). Econometric analysis of cross section and panel data, MIT Press.

[66] Zhang, J.L. and Rubin, D.B. (2003). Estimation of causal effects via principal stratication when some outcomes are truncated by death, Journal of Educational and Behavioral Statistics, 28: 353–368.

[67] Zhang J., Rubin D.B., Mealli F. (2006). Evaluating The Effects of Job Training Programs on Wages through Principal Stratification, in: Modeling and Evaluating Treatment Effects in Econometrics, Advances in Econometrics vol. 21, D. Millimet, J. Smith, and E. Vytlacil (eds.), Elsevier Science Ltd, UK.