ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Angelucci, Manuela; Attanasio, Orazio P.

Working Paper

Estimating ATT effects with non-experimental data and low compliance

IZA Discussion Papers, No. 2368

Provided in Cooperation with:
IZA Network @ LISER, Luxembourg Institute of Socio-Economic Research (LISER)

Suggested Citation: Angelucci, Manuela; Attanasio, Orazio P. (2006) : Estimating ATT effects with non-experimental data and low compliance, IZA Discussion Papers, No. 2368, Institute for the Study of Labor (IZA), Bonn,
https://nbn-resolving.de/urn:nbn:de:101:1-20090406301

This Version is available at:
https://hdl.handle.net/10419/33804

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 2368

Estimating ATT Effects with Non-Experimental Data and Low Compliance

Manuela Angelucci
Orazio Attanasio

October 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Estimating ATT Effects with Non-Experimental Data and Low Compliance

Manuela Angelucci
University of Arizona
and IZA Bonn

Orazio Attanasio
University College London,
NBER, BREAD and CEPR

Discussion Paper No. 2368
October 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2368
October 2006

ABSTRACT

Estimating ATT Effects with Non-Experimental Data and Low Compliance*

In this paper we discuss several methodological issues related to the identification and estimation of Average Treatment on the Treated (ATT) effects in the presence of low compliance. We consider non-experimental data consisting of a treatment group, where a program is implemented, and of a control group that is non-randomly drawn, where the program is not offered. Estimating the ATT involves tackling both the non-random assignment of the program and the non-random participation among treated individuals. We argue against standard matching approaches to deal with the latter issue because they are based on the assumption that we observe all variables that determine both participation and outcome. Instead, we propose an IV-type estimator which exploits the fact that the ATT can be expressed as the Average Intent to Treat divided by the participation share, in the absence of spillover effects. We propose a semi-parametric estimator that couples the flexibility of matching estimators with a standard Instrumental Variable approach. We discuss the different assumptions necessary for the identification of the ATT with each of the two approaches, and we provide an empirical application by estimating the effect of the Mexican conditional cash transfer program, *Oportunidades*, on food consumption.

JEL Classification: C31

Keywords: program evaluation, treatment effects

Corresponding author:

Manuela Angelucci
Department of Economics
University of Arizona
McClelland Hall 401
Tucson, AZ 85721-0108
USA
E-mail: angelucm@eller.arizona.edu

* We would like to thank Erich Battistin, Jere Behrman, Kei Hirano, Costas Meghir, Barbara Sianesi, Petra Todd, Marcos Vera-Hernandez, and Iliana Yaschine for very useful discussions of the issues covered by this note. Several staff at *Oportunidades* were very helpful with questions about the details of the data.

1 Introduction.

This paper discusses alternative approaches for the identification and estimation of Average Treatment on the Treated (ATT) effects of non-experimental programs. Random allocation across geographic areas, or randomization of eligible subjects within one area, is often considered infeasible or unethical or politically unappealing. In these cases, the data for the evaluation of the program frequently consist of subjects with similar observable characteristics, non-randomly allocated to a group that is offered the treatment, and a group that is not. Often these individuals live in different geographic areas and the program is available only in some locations. A common practice, in such a situation, is to use matching estimators to "balance" control and treatment samples in terms of a set of observable characteristics. The presumption on which these methods are based is that, conditional on observables, the exposure to the program is random. Following the literature, we call this assumption the Conditional Independence Assumption (CIA).

An additional problem arises when there are non-compliers - eligible individuals in treatment areas who do not enroll into the program.[1] While the reasons for non-participation may vary, the general presumption is that compliance is not random. This situation poses a problem and faces the econometrician with a temptation. The problem is that, even if one is willing to believe that treatment and control group assignment is random conditional on observables, comparing these groups' observed outcomes will not yield Treatment on the Treated effects because of the presence of non-compliers. The temptation is to solve the issue of participation using the same matching methods used to balance the characteristics of compliers and non-compliers, or of compliers and subjects from the control group, assuming that, conditional on observables, participation into the program is random. In this paper, we argue that this temptation should be resisted: since compliance is a choice variable, unlike assignment to the treatment and control groups, it is likely based on unobservable information. Consequently, the CIA may be valid, for a given set of observables, when comparing subjects from the treatment and control group, but not when comparing participants and non-participants, or participants and control subjects. For example, consider the empirical application, in which we estimate the effect of a program offering scholarships to poor children on household food consumption: with perfect (or very high) take-up rates, when estimating the ATT we simply need to worry about finding households coming from similarly disadvantaged backgrounds who have not been

[1] We refer to this group as non-compliers or non-participants in the rest of the paper. Analogously, we call eligible households who receive the treatment compliers or participants.

offered the scholarship.[2] When the take-up rate is low, instead, we need to be able to observe all variables inducing the subject to participate to the program which are related to outcomes. To illustrate this point, consider a family with one eligible child, who normally attends school, but who did not enroll the year before the beginning of the program because the household head was temporarily unemployed due to transitory health problems. This family joins the program because the child would have gone back to school in any case. If we do not observe all the relevant information, we may end up matching this household with families whose head is regularly unemployed, and whose children never attend school. These families' consumption is likely lower than the consumption of the treated family in the absence of the treatment, resulting in an upward-biased estimate of the treatment effect. Clearly, observing pre-program characteristics helps, but it may not take care of selection based on transitory shocks that also affect outcomes.[3] We propose instead an alternative method to deal with endogenous participation.

The method we propose to estimate ATT effects consists of inflating the Average Intent to Treat (AIT) estimates using information on participation. In the presence of non-compliance, the AIT is an interesting policy parameter *per se*, because it measures the average program effect on the eligible individuals. To estimate the AIT, one can use the standard methods proposed in the literature, which are valid under well-understood assumptions. In the absence of spillover effects, dividing the AIT by the share of treatment subjects who do receive the treatment identifies the ATT. This is an Instrumental Variable estimator, in the sense that it exploits the assumption that group type (treatment or control) is conditionally independent from potential outcomes. Thus, this estimator is in the spirit of Imbens and Angrist (1994), Angrist, Imbens and Rubin (1996) and Heckman, Smith and Taber (1998), with the difference that their instrumental variable is excludable unconditionally, because they consider situations where the program is randomly assigned, or of Abadie (2003), in the sense of considering Instrumental Variable estimators for models with covariates.[4] Further, our paper belongs to the literature that warns against considering matching estimators as the "magic bullet" that

[2]Obviously, if these subjects live in different geographical areas, one needs to consider the additional issues of different geographic characteristics, while if they come from the same area one may worry about spillover or peer effects.

[3]The same reasoning (probably even more poignantly) applies for the plausibility of the common trend assumption invoked when using pre- and post-program data to difference out constant unobservables. It is possible and even likely that transitory shocks play different roles in program assignment across areas and in individual program enrollment.

[4]Heckman, Smith and Taber (1998) consider also the case of partial compliance, in which the treatment has an effect on non-compliers. In that case they use an instrument, different from program assignment, that affects program participation among the treated but does not affect outcomes. If we had such an instrument, we could follow a similar approach.

solves the evaluation problem, but rather that the choice of an estimator depends on the features of the data and the institutions involved (see, for example, Smith and Todd (2005)).

After discussing identification and estimation of ATT effects contrasting our estimator with the standard matching approach (Sections 2 and 3), we provide an empirical application by estimating the AIT and ATT effects of the conditional cash transfer program *Oportunidades* on food consumption in urban areas (Section 4). *Oportunidades* is the successor of PROGRESA, the flagship welfare program of the Mexican government aimed at fostering the accumulation of human capital in rural areas. Extended to urban areas, it is now the largest welfare program in Mexico by a huge margin, as it covers more than one in ten Mexicans. The program design fits the type of evaluation we described above: the allocation of the program within the localities in the evaluation sample was not random, and the enrollment rate of eligible household was, a year after the start of the program, lower than 50%. Applying our favorite method we obtain a substantial effect of the grant on food consumption. An implication of our estimates is that the beneficiary families spend most of the grant on food, saving about 20%. On the other hand, applying matching methods to deal with participation, we obtain substantially lower effects, implying that the beneficiary households consume less than one third of the grant on food and have an astonishing saving rate, in stark contrast with the empirical evidence from the rural sample and from similar programs in other countries. By comparing the determinants of the probability of participation and of the probability of being in a treatment area, we notice that different variables are important in the two equations. A consequence of this imbalance is that, when using matching, we compare participants to relatively better off households.

2 Identification of program impacts.

We are mainly interested in a situation where the assignment of the program of interest is not random and, *at the same time,* a sizeable proportion of eligible of individuals decide not to participate into the program. In this section we define the parameters of interest in such a situation and spell out the identification assumptions necessary to estimate them. We discuss why these parameters are interesting and what we learn from them in terms of the program effect on the variables of interest. Finally, we present different ways to estimate them that tackle various identification issues.

Consider data providing information on a large set of characteristics for treatment and control subjects. Assignment to the treatment is not randomized. However, control group members have similar observable characteristics to the treatment group ones. Program take-up

is not 100%, hence there is a set of subjects in the treatment group who does not receive the treatment. This type of sample design and the presence of non-compliers are very common features of non-experimental evaluations.

2.1 Identification of AIT and ATT

We are interested in estimating two main parameters: average intention to treat (AIT) effects and average treatment on the treated (ATT) effects. Both of these parameters are of interest from a policy point of view. The first parameter, AIT, estimates the average effect of the treatment on the outcome of interest of all eligible subjects, irrespective of their participation into the program. As it effectively ignores the issue of what determines participation into the program, the identification of AIT requires less restrictive identification assumptions than that of ATT. However, as we argue below, the AIT only provides a lower bound of the size of the ATT (for a positive ATT). The direct estimation of ATTs has the disadvantage of requiring alternative sets of more restrictive identification assumptions but it has the advantage of measuring the effect of the program on recipients, rather than on all the eligible subjects.

We define potential outcomes on the basis of program availability. Individual i's potential outcome is $Y_i(1)$ in the presence of treatment availability, irrespective of compliance, and $Y_i(0)$ without treatment availability. Obviously, it is impossible to observe the two potential outcomes for the same individual.

Average Intention to Treat. We define the AIT effect as the average program effect for individuals who are offered the treatment:[5]

$$E[Y(1) - Y(0)] \tag{1}$$

The AIT measures the effect of the program on eligible subjects, regardless of whether they participate into the program or not. Since often the policy maker has little influence on participation, the AIT is one relevant parameters for policy analysis. The AIT is also interesting because it provides a lower bound to the ATT under the assumption that the program effect on non participants in the treatment group is lower than its effect on compliers.[6]

Suppose the treatment is made available in some areas but not in others. Treatment ares are identified by a variable $Z = 1$. The AIT in these areas would then be given by $E[Y(1)-Y(0)|Z = 1]$: We can observe actual outcomes of individuals who live in areas where the treatment is

[5] Although we drop the individual-specific subscript i for notational ease, we do allow for heterogeneous effects.

[6] We have made this statement relative to a positive ATT effect. Similar considerations hold for a negative one.

offered, hence $E[Y(1)|Z=1] = E[Y|Z=1]$. Instead, we cannot observe what the variable of interest in the treatment group would have been in the absence of the program. Thus, we require some assumptions that allow us to use the control groups observed outcome to estimate this counterfactual. As it is standard in this literature, we make the following two assumptions:

- Assumption (1.1) - Stable Unit Treatment Value Assumption (SUTVA).[7] Potential outcomes of each individual are unrelated to the treatment status of other individuals. This rules out general equilibrium effects. In this particular context, we require that the outcome of interest for control group subjects is not affected by the treatment.

- Assumption (1.2) - Conditional Independence Assumption (CIA). Conditional on observables (X), potential outcomes are independent of treatment availability, $Z \perp Y(0), Y(1)|X$.

- Assumption (1.3) - Common support. $Pr(Z=1|X) < 1$.

These three assumptions imply that the expected potential outcome in the absence of the treatment offer for subjects who have been offered to participate to the program, $E[Y(0)|X = x]$, equals the expected observed outcome for subjects who have not been offered the program, $= E[Y|Z=0, X=x]$. This solves the missing counterfactual problem and allows one to compute the AIT by comparing observed outcomes of subjects in the treatment and control groups, Y, conditional on a set of variables X, and by integrating this difference over the relevant values of X:

$$
\begin{aligned}
AIT \quad &= E[Y(1) - Y(0)] && (2)\\
&= \int_x E[Y|Z=1, X=x] - E[Y|Z=0, X=x] dF(x|Z=1) && (3)\\
&= \int_p E[Y|Z=1, P(X)=p] - E[Y|Z=0, P(X)=p] dF(p|Z=1) && (4)
\end{aligned}
$$

where the last equality reduces the dimensionality of the problem by conditioning on the propensity score, $Pr(X) = Pr(Z=1|X)$ (Rosenbaum and Rubin, 1983).

To define the ATT we further need to consider compliance. $D_i(1)$ indicates potential compliance, $D_i(1) = 1$, and non-compliance, $D_i(1) = 0$, when the treatment is available. $D_i(0)$ is the compliance status when the individual is not offered the treatment, hence it is always 0. $Y_i(1, D(1))$ and $Y_i(0, D(0))$ are the potential outcomes with and without being offered the

[7] Formalized by Rubin (1980, 1986), SUTVA means that potential outcomes depend on the treatment received, and not on what treatments other units receive. Hence, the SUTVA rules out any effect of the program on non-treated subjects. In each of the cases discussed in this context, however, we require this assumption to hold only for particular sub-sets of our sample in order for our parameter of interest to be identified. We will indicate which sub-group we need the SUTVA to hold for each parameter or identification strategy.

treatment: $Y_i(1,1)$, $Y_i(1,0)$, and $Y_i(0,0)$ are the potential outcomes with compliance, without compliance, and without treatment availability, respectively. The potential outcome $Y_i(0,1)$ is the missing counterfactual, as there are no compliers in the group that is not offered the treatment. Compliance is a decision variable and, as such, is likely correlated with potential outcomes.

Average Treatment on the Treated. Our second parameter of interest is the ATT effect, defined as:

$$E[Y(1,1) - Y(0,1)] \tag{5}$$

This parameter measures the effect of the program on subjects who actually receive it ($D(1) = 1$). If participation into the program (compliance) were complete, the ATT and the AIT would coincide. Analogously to what discussed above for the AIT, we observe the first object, as $E[Y(1,1)] = E[Y|Z = 1, D = 1]$, but we cannot observe participants' outcome in the absence of the treatment. There are at least two ways to identify this effect. One is by assuming that, conditional on observables, participation into the program is random. The alternative is to use treatment availability as an exclusion restriction to identify a conditional version of a Local Average Treatment Effect. While the details by which one implement these two approaches might vary, it is clear that the nature of the assumptions under which one approach or the other provides reliable estimates are different. We discuss them in turn.

2.2 *Matching approaches to participation.*

A first alternative to deal with the issue of non-participation, is to use a matching approach in the hope that, conditional on observables, participation into the program is random. If one decides to follow such an approach to estimate the ATT, there are several alternatives that correspond to the use of different control groups. The identification assumptions are:

- Assumption (2.1) - SUTVA. One consequence of the SUTVA is that the change in the variable of interest for the control group does not depend on the treatment.
- Assumption (2.2) - CIA. Conditional on observables, compliance is independent of potential outcomes, $D(1) \perp Y(0,1), Y(1,1), Y(0,0)|X$.
- Assumption (2.3) - Common support. This assumption varies according to which group one uses as counterfactual, as we will show below.

This set of assumptions insures that the unobservable $E[Y(0,1)|D(1) = 1, X = x]$ equals both $E[Y|D(1) = 0, X = x]$ and $E[Y|Z = 0, X = x]$. Thus, there are at least two options to identify

the ATT. One is to use non-compliers in the treatment group as a possible counterfactual for participants in the same groups, in which case the common support assumption is $Pr(D = 1|Z = 1, X) < 1$ and the ATT is

$$ATT = E[Y(1,1) - Y(0,1)] \tag{6}$$

$$= \int_x E[Y|Z = 1, D = 1, X = x] - E[Y|Z = 1, D = 0, X = x]dF(x|D = 1) \tag{7}$$

One advantage of comparing eligible subjects within the treatment group is that the counterfactual outcome comes from subjects living in the same geographic areas as the compliers. However, a disadvantage is that the SUTVA might not be plausible when referred to individuals in the treatment group, because it ignores the possibility of program spillover effects to non-compliers.

There are at least two other approaches within the matching framework that are robust to the presence of spillover effects in treatment areas. The first consists in comparing compliers in treatment areas to eligible households in control areas, i.e. individuals with similar observable characteristics who are not offered the program. In this case, matching balances the differences in the X variables for these two groups and the common support assumption is $Pr(V = 1|X) < 1$, where $V = 1$ for compliers and $V = 0$ for individuals who are not offered the program. The ATT is then defined as:

$$ATT = E[Y(1,1) - Y(0,1)] \tag{8}$$

$$= \int_x E[Y|Z = 1, D = 1, X = x] - E[Y|Z = 0, X = x]dF(x|D = 1) \tag{9}$$

A second approach is the one employed by Todd *et al.* (2006) to evaluate some of the effects of the program we study in our empirical application. This approach compares compliers in treatment areas with eligible households in control areas who have the same probability of participating in the program, were the program available. Todd *et al.* (2006) estimate a propensity score using compliers and non-compliers in the treatment group, and impute the probability of participation among eligible households in control areas. Implicit in this procedure is that balancing participation also balances the differences across areas.

The equations (7) and (9) are valid also conditional on the propensity scores, which are $Pr(D = 1|Z = 1, X)$ and $Pr(V = 1|X)$, respectively. While the assumptions above guarantee that both methods yield a consistent estimate of the ATT, these assumptions might be more or less plausible when applied to individuals within the treatment areas or to individuals living in

treatment and control areas. If one thinks that spillover effects might be important, the second method might be more appropriate. If one thinks that differences across areas might be harder to control for, one might prefer the first method.

In both cases, the Conditional Independence Assumption plays a very important role as we need to observe all variables that affect both participation and outcome. In the latter case, the matching procedure has the task of balancing both the participation into the program and the non-random assignment of the program across groups. We come back to these issues below.

2.3 A LATE approach to participation.

Since the sets of assumptions needed to identify the ATT through matching are quite strong, we propose to identify the missing counterfactual by using group type (treatment or control) as an instrumental variable that is correlated with participation, but conditionally uncorrelated with outcome. The discussion below is related to the one in Angrist, Imbens, and Rubin (1996), and Heckman, Smith, and Taber (1998), with the difference that in our context, unlike in these other papers, the issue of non-compliance arises on top of a non-random assignment into the treatment and control groups.[8] In this setting, we need the following assumptions to identify the parameter of interest:

- Assumption (3.1) - SUTVA. Program participation and the outcomes of interest for subjects in the control and treatment groups, *including non-compliers*, are not affected by the participation status of others.

- Assumption (3.2.a) - CIA. Conditional on observables (X), potential outcome are independent of treatment availability, $Z \perp Y(0,0), Y(1, D(1))|X$.[9]

- Assumption (3.2.b) - Common support. $Pr(Z = 1|X) < 1$.

- Assumption (3.3) - Exclusion restriction. Conditional on observables, the expected value of potential outcomes in the absence of the program is the same in the treatment and control groups, $E[Y(1,0) - Y(0,0)|D(1) = 0]$.

- Assumption (3.4) - Nonzero average causal effect of group type on program participation. Conditional on X, the average compliance rate is greater than zero, $E[D(1)|X] > 0$.

[8] See also the discussion in Heckman (1996) and Moffitt (1996).

[9] This is an implication of the ignorability assumption made by Angrist, Imbens and Rubin (1996).

- Assumption (3.5) - Monotonicity. There is no individual who would have participated to the program if he or she had been in the control group, but not if he or she had been in the treatment group.

With this approach one normally identifies the LATE, i.e. the average treatment effect for the set of agents who are induced to participate in the program because of the instrument. In this particular case, though, the instrument is the group type and the participation rate in the control group is zero. Thus, the subjects who are induced into treatment because they belong in the treatment group are all the treated subjects. Hence, given the above assumptions, this instrument identifies the ATT effect:

$$
\begin{aligned}
ATT &= E[Y(1,1) - Y(0,1)|D(1) = 1] && (10)\\
&= \int_x \frac{E[Y|Z=1, X=x] - E[Y|Z=0, X=x]}{E(D=1|X=x)} dF(x|D=1) && (11)\\
&= \int_p \frac{E[Y|Z=1, P(X)=p] - E[Y|Z=0, P(X)=p]}{E(D=1|P(X)=p)} dF(p|D=1) && (12)
\end{aligned}
$$

where the last equality follows from Rosenbaum and Rubin (1983) and the propensity score is $Pr(X) = Pr(Z = 1|X)$.

Assumptions (3.2.a) and (3.2.b) are the same needed to identify the AIT. Assumption (3.4) is obvious in this context, as the program is available only to subjects living in the treatment group, and assumption (3.5) seems quite realistic. The main differences between identifying ATT effects by matching or by IV are related to differences in the versions of SUTVA, CIA, and exclusion restrictions needed.

2.4 Comparing the two approaches

In the matching cases we have to assume that, conditional on observables, program participation is unrelated to potential outcomes. This implies observing all variables determining both participation and outcome, including transitory shocks. In the IV case, instead, we relax this assumption because we only require the exclusion restriction to hold and conditional independence to apply for group type, rather than for compliance. In other words, we only need to observe all the variables by which the subjects in the treatment and the control groups differ. While unobservable transitory shocks may differ between compliers and non-compliers (if the shock is one of the causes for participation), it is less likely that the distribution of such shocks will differ systematically between control and treatment. The same argument applies for assumption (3.3). On the other hand, we need to assume away any program indirect effect

on both non-compliers, as in (2.1), and control group subjects (as in 1.1). This assumption may be questionable in some applications.[10]

3 Estimation of program impacts

In this section we discuss the estimation of AIT and ATT effects which correspond to the alternative identification strategies we discussed above. We assume that program effects are heterogeneous. We start by considering estimates of the AIT and then move on to the more controversial ATT.

3.1 Estimating the AIT

The case of the AIT is reasonably standard. The assumption made above are those typically used in the literature. If one is willing to make linearity assumptions, one can estimate the AIT by a simple OLS regression, regressing Y on various controls and a dummy for group type. Alternatively, if one wants to use a less parametric approach, one may use Propensity Score Matching techniques, first computing the probability of being in a treatment group as a function of some variables X, and then comparing outcomes of subjects in the treatment group to outcomes of subjects in the control group with similar propensity scores. One can estimate the AIT by replacing the objects in equation 4 with their sample equivalents.

3.2 Estimating the ATT

As we mentioned above, we consider two approaches to estimate the ATT. The main issue we need to deal with is the fact that a substantial fraction of eligible subjects in the treatment group did not participate to the program. The first approach is to employ matching to try to balance participants and non-participants. The alternative is to use the treatment availability as an Instrumental Variable. We discuss the estimators in turn.

[10] Hirano *et al.* (2000), for example, estimate a positive treatment effect for non-compliers. In the application we use as an example, the program may have effects on the local goods, labor, and financial markets, which may change non-participants' consumption (non-participants may be non-compliers, control subjects, or both). For example, participants' higher consumption may increase goods prices; alternatively, the liquidity injection caused by the program may result in larger loans or transfers to non-participants. In both these cases non-participants' consumption would likely change. However, we suspect that these indirect effects are unlikely to occur in the case of consumption, given that the program has only been implemented for a few months in the second and last data wave. It is clear, however, that such an assumption might be more problematic for other variables.

3.2.1 Controlling for limited participation through matching estimators

Even if one decides to use matching methods to deal with the issue of participation, there are different possible estimators available, corresponding to different comparison groups and slightly different identification assumptions. What these different approach have in common is the assumption that, conditional on observable, participation into the program is random. In practice we can use two different groups as the control group. The first is given by non-participant subject in the treatment group. The second is given by eligible subjects in the control group.

1. *Comparing participants and non participants in treatment areas.* The first step in this procedure is to estimate the probability of compliance using treatment group eligible individuals. One then uses these estimates to match compliers with similar non-compliers, weighting differences in observed outcomes by the density of the propensity score among compliers to obtain the ATT.

2. *Comparing compliers to eligibles in the control group.* There are several ways to compare these two groups. One way consists of pooling eligible subjects in the control group with compliers in the treatment group, and of computing the following propensity score: $Pr(X) = Prob(V = 1|X)$, where $V = 0$ for subjects in the control group and $V = 1$ for compliers. Alternatively, one may estimate the probability of compliance from the treatment group, and impute a propensity score to eligible subjects in the control group. Irrespective of the method chosen to estimate the propensity score, one then matches compliers to control group subjects with similar probabilities to participate into the program as described in the previous paragraph.

Since these are standard methods in the literature, we do not discuss their implementation in details. Yet, we want to emphasize once more that, besides the CIA, the first type of matching relies on the assumption that there are no spillover effects of the program on non-compliers. The second type of matching replaces this assumption with the milder requirement of no spillover effects for subjects in the control group. However, as we discussed above, this comes at the cost of using only information on participation in the treatment group to balance both treatment and the control group, and participants and non participants within the the treatment group.

3.2.2 An IV approach to estimate the ATT

The assumptions listed in Section 2.3 suggest a different approach. If the effect of the program on non participants is smaller than that on participants, the AIT constitutes a lower bound to the ATT. If one is willing to assume that, conditional on X, the average treatment effect on non-compliers is zero, and treatment availability is random, then we can estimate the ATT using a conditional version of the LATE estimator, with treatment availability as instrument. One can derive an estimator by considering the sample equivalents of the moments in equation (12). In particular, for each value of the propensity score one can first estimate the Intention to Treat and then divide it by the proportion of compliers. Weighting these conditional ATT effects by the density of the propensity score among compliers, one can estimate the unconditional ATT.

This last set of considerations emphasizes the difference between the conditional independence assumptions required for the two approaches: with standard matching we need to assume that program participation is random, conditional on $X(Pr(D = 1|X))$, while using group type as instrumental variable we only need random group assignment, again conditional on X $(Pr(Z = 1|X))$.

4 An application: estimating the effect of *Oportunidades* on food consumption

To compare the estimators we discussed in Sections 2 and 3, we estimate the ATT effect of a conditional cash transfer program in urban areas in Mexico, *Oportunidades,* on food consumption. Unlike its rural component at the start of the program, *Oportunidades* was not assigned randomly across areas. Therefore even to evaluate the AIT we need to balance the treatment and control samples. As we will see, the low participation into the program introduces an important difference between AIT and ATT.

Besides being interesting in itself, the analysis of consumption is useful in the present context because one typically has strong priors about the effect of the program on such a variable. These priors are partly derived from the intuition that an increase in (permanent) income should be reflected in an increase in consumption, and partly based on existing empirical evidence.[11]

After discussing the characteristics of both the program and the evaluation sample, we present estimates of different propensity scores, corresponding to different comparison groups.

[11] Hodinott and Skoufias (2003) and Gertler et al (2005) show that in the rural *Oportunidades*, originally called PROGRESA, beneficiary households increased consumption by an amount that corresponded to 75 to 85% of the grant.

In addition to the propensity score that measures the probability of being in a treatment group, we also present estimates of the propensity score that measures the probability of participating into the program in the treatment group. These are used to compute the estimates of the ATT based on matching we mentioned above. We then move on to present the estimated AITs, and the ATT estimates obtained using the two alternative approaches discussed above. For the first approach, which deals with participation using matching methods, we present two different sets of results, based on two different control groups. For the alternative approach, based on "inflated" AIT estimates obtained using LATE methods, we also present two approaches that correspond to two different ways to control for observables in computing the AIT: parametric and non-parametric. We conclude the section with a comparison of the results obtained with the two approaches.

4.1 *Oportunidades* and its evaluation.

Started in 2003 and still ongoing, *Oportunidades* is the urban component of the Mexican conditional cash transfer program previously named PROGRESA. The program targets poor urban households, providing scholarships and other smaller transfers to improve recipients' education, nutrition, and health. Unlike PROGRESA, in *Oportunidades* the allocation across geographic areas was not random. The program operated setting up registration offices (módulos) within eligible areas, and investing resources in spreading the news about the availability of the program in that area. Potential beneficiaries had to visit a local office first, and then found out whether they qualified for the program, based on an estimated poverty status. The consequence of this scheme was that, at least in the first year of operation, many potentially eligible households did not apply for the program - possibly because they were not aware of its existence, or because of uncertainty over their eligibility status, or because *Oportunidades* was simply less attractive in urban areas than rural areas. Indeed, administrative data indicate that program take-up does not reach 50% of eligible households.

The evaluation of the impact of *Oportunidades* is based on the comparison of households living in treatment areas to households living in control areas. As mentioned above, the assignment of the urban *Oportunidades*, however, was not random. Given the budget available, it was decided to start the program first in areas with a high concentration of poor households. This implies that treatment areas are different from control areas. The evaluation advisory group decided to sample blocks[12] within treatment areas and to match each block in the treatment sample to a control block with similar observable characteristics (based on a pre-estimated

[12] A block is much smaller than an area.

propensity score). This implies that the very expansion of the program prevented the use of certain variables to form the propensity score. For instance, as the program was assigned to the areas with the highest concentration of poor households, using such a variable would give no intersection between treatment and control samples.

A further issue arises from the fact that the data are "choice-based", as they oversample compliers. Thus, the fraction of eligible (i.e. poor) households participating into the program observed in our treatment sample is quite different from the true fraction of program participants. Fortunately, we can estimate the true proportion of participating households in each block from a different data set.[13]

Our sample consists of data on about 8000 poor households from treatment and control areas, interviewed both in 2002, after households had registered for *Oportunidades*, but before any payments had been made, and in 2003, after payments started. We have information on the consumption of many types of food in the week prior the interview, including consumption in kind. We transform all the figures to monthly equivalents. The availability of pre-program data enables us to implement difference-in-difference estimators. The advantage of this class of estimators is that the required assumptions are on the *change* in the variable of interest, rather than on its level. It is easy to modify all the above assumptions considering changes, rather than levels of potential outcomes, allowing for the possibility that pre-program outcomes may differ between the groups one compares.

4.2 Propensity score and common support

The first panel of Table 1 shows the frequencies of the propensity scores for area of residence, $Pr(Z = 1|X)$, using our preferred set of covariates in the estimation of the propensity score. This set of regressors (using 2002 values, unless otherwise specified) is: child number by age and school enrollment; heads of household's gender, education, literacy, employment status since 1999; spouse presence in the households and education level; household poverty status, income since 1999, savings and debts; transitory shocks such as natural disasters, loss of business or death of household members; number of doctor visits by children and parents; state dummies. We use this set of covariates in all the propensity scores discussed in this Section.

Although the common support is complete, the right tail for the control group is very thin, so the estimation in this part of the support may not be very precise. Unreported robustness checks show that dropping state dummies improves the support but fails to control for differential trends between control and treatment areas, resulting in biased estimates of the treatment

[13] See Angelucci, Attanasio, and Shaw (2004) for further details.

effects.

The second panel of Table 1 shows the distribution of the propensity score computed considering compliers and non-compliers in treatment areas, $Pr(D = 1|Z = 1, X)$. This is the propensity score to compute the ATT using only individuals in treatment areas: participants are compared to non-participants under the CIA and the assumption of no spill-over effects. In this case, common support is definitely not a problem. While there is some evidence of self-selection into the program based on observable characteristics, it is not very strong: 42% of non-participants and 74% of participants have a probability of participation above 50%. This fact may be caused by a lack of information on the program existence or features - e.g. only a random subset of eligible households is sufficiently informed about the program. If this were true, we would expect participation rates to increase in the future, as knowledge of the program existence and characteristics spreads among the eligible population. Apparently, this is not the case, because enrolment did not increase in 2004. The density of the propensity score is also consistent with the possibility that households self-select into the program based on time-varying unobservables. Further research is needed to distinguish between these alternative hypotheses.

The third panel of the Table reports the distribution of propensity scores for the case in which compliers, $V = 1$, are compared to eligible households from control areas, $V = 0$, i.e. $Pr(V = 1|X)$. The distribution of the propensity score is not dissimilar to that in the first panel: while we do have households in control areas with relatively high propensity scores, these are not very many. Moreover, we have a large number of compliers that have propensity scores above 0.9.

Finally, in the fourth panel we compute the probability of program participation for households in treatment areas, $Pr(D = 1|Z = 1, X)$, and predict the value of the propensity score for control households, i.e. $Pr(D = 1|Z = 0, X)$. This is the propensity score used by Todd *et al.* (2006). The distribution of this propensity scores among the $Z = 1$ and the $Z = 0$ groups is remarkably similar, as well as similar to that in Panel 2, even though the control group is taken from areas where the program was not available, as in Panels 1 and 3. While this might seem comforting for the application of the the second matching method discussed above, the differences between Panels 3 and 4 indicates that the variable that balance participants and non participants in treatment areas are different from those that balance treatment and control areas. As we discuss below, this may create important biases.

After showing the differences in the propensity scores in the various panels of Table 1, it is worth comparing the probit estimates of the coefficients used to compute them. In the case of the propensity score for participation, this is also informative about the process that

Table 1: Frequencies of the propensity score (ps)

	(1) $Pr(Z=1\|X)$		(2) $Pr(D=1\|Z=1,X)$		(3) $Pr(V=1\|X)$		(4) imputed	
	Z=0	Z=1	D=0	D=1	V=0	V=1	D=0	D=1
$[0, 0.1)$	0.13	0.01	0.02	0.00	0.25	0.02	0.01	0.00
$[0.1, 0.2)$	0.14	0.02	0.04	0.01	0.32	0.06	0.05	0.01
$[0.2, 0.3)$	0.19	0.04	0.11	0.03	0.14	0.05	0.12	0.03
$[0.3, 0.4)$	0.17	0.05	0.18	0.08	0.07	0.04	0.16	0.09
$[0.4, 0.5)$	0.12	0.06	0.22	0.14	0.05	0.04	0.17	0.15
$[0.5, 0.6)$	0.08	0.05	0.18	0.19	0.05	0.04	0.15	0.19
$[0.6, 0.7)$	0.06	0.05	0.13	0.20	0.04	0.08	0.14	0.19
$[0.7, 0.8)$	0.05	0.08	0.08	0.20	0.03	0.10	0.12	0.19
$[0.8, 0.9)$	0.04	0.15	0.03	0.13	0.02	0.15	0.05	0.12
$[0.9, 1]$	0.03	0.48	0.003	0.02	0.01	0.42	0.01	0.02

Propensity scores (ps) estimated by weighted probit. The frequencies may not add up to 100 because of rounding error.

induces eligible households to register into the program. As mentioned above, a surprisingly small proportion of eligible households actually registered for the program, 47.5% in our sample. Although it is not completely obvious why such a small proportion of eligible households registered for the program, it is not difficult to find variables that help predict program participation. In Table 2, we present weighted probit estimates of the propensity score coefficients. Column 1 predicts the likelihood of living in a treatment area, $Pr(Z = 1|X)$; column 2 models participation among eligible households in treatment areas, $P(D = 1|Z = 1, X)$, while column 3 predict participation using compliers and poor households in control areas, $P(V = 1|X)$. Besides being informative about the process that determines participation, these results are useful because they show the characteristics by which households in treatment and control areas differ, and because a comparison of the first two sets of estimates reveals whether matching on the predicted probability of participation balances all the variables by which $Z = 1$ and $Z = 0$ households differ.

Two interesting facts emerge from looking at the determinants of program participation (column 2). First, individuals respond to economic incentives, in the sense that participating households seem to be the ones with the highest net benefits from the program: households with a higher poverty index and a large number of children in school are more likely to be *Oportunidades* recipients; participation is also higher among households with illiterate or low educated heads and spouses. Second, participation seems to be associated with permanent poverty indicators, and to a smaller extent by transitory shocks. Low parental education and high poverty index, which increase the likelihood of participation, are proxies of permanent poverty levels. Current and past income and employment, job or business loss, illness or death of relative and natural disasters do not seem to be related to participation, with the exception

Table 2: The coefficients of the propensity scores - marginal effects

	(1) $P(Z=1\|X)$	(2) $P(D=1\|Z=1,X)$	(3) $P(V=1\|X)$
nokids0-5	0.037	-0.008	0.041
	[0.010]***	[0.014]	[0.013]***
nokids6-12	0.011	-0.018	0.02
	[0.017]	[0.028]	[0.022]
nokids13-15	-0.003	-0.045	-0.018
	[0.020]	[0.029]	[0.027]
nokids16-20	-0.013	-0.02	-0.024
	[0.014]	[0.020]	[0.018]
nosch0-5	-0.006	0.058	0.013
	[0.023]	[0.032]*	[0.030]
nosch6-12	-0.005	0.073	0.002
	[0.017]	[0.027]***	[0.021]
nosch13-15	-0.013	0.127	0.016
	[0.023]	[0.033]***	[0.030]
nosch16-20	-0.016	0.014	-0.013
	[0.024]	[0.031]	[0.031]
sfem	0.072	0.042	0.129
	[0.030]**	[0.040]	[0.040]***
nopartner	0.03	-0.04	
	[0.034]	[0.050]	
hohlit-1	0.023	-0.062	-0.005
	[0.027]	[0.036]*	[0.035]
hohlit-2	-0.018	-0.043	-0.021
	[0.085]	[0.110]	[0.115]
ptrlit-1	-0.174	0.036	-0.163
	[0.026]***	[0.035]	[0.035]***
ptrlit-2	0.071	0.193	0.182
	[0.061]	[0.066]***	[0.080]**
hoheduc-1	-0.093	0.058	-0.086
	[0.029]***	[0.036]	[0.036]**
hoheduc-2	-0.142	0.028	-0.134
	[0.034]***	[0.040]	[0.040]***
hoheduc-3	-0.111	-0.013	-0.12
	[0.044]**	[0.054]	[0.051]**
hoheduc-4	-0.176	0.028	-0.18
	[0.038]***	[0.044]	[0.042]***
hoheduc-5	-0.229	0.004	-0.208
	[0.049]***	[0.055]	[0.049]***
ptreduc-1	0.151	-0.047	0.118
	[0.025]***	[0.040]	[0.038]***
ptreduc-2	0.145	-0.077	0.102
	[0.027]***	[0.045]*	[0.043]**
ptreduc-3	0.155	-0.147	0.094
	[0.030]***	[0.058]**	[0.058]
ptreduc-4	0.139	-0.128	0.057
	[0.028]***	[0.049]***	[0.048]
ptreduc-5	0.17	-0.153	0.095
	[0.032]***	[0.063]**	[0.065]

Continues

Continued			
	-1	-2	-3
	$P(Z=1\|X)$	$P(D=1\|Z=1,X)$	$P(V=1\|X)$
hohemploy-1	0.1	0.077	0.141
	[0.030]***	[0.038]**	[0.037]***
hohse-1	0.078	0.097	0.116
	[0.028]***	[0.039]**	[0.039]***
ptremploy-1	0.103	0.044	0.131
	[0.017]***	[0.026]*	[0.024]***
ptrse-1	0.104	0.067	0.157
	[0.019]***	[0.029]**	[0.027]***
hhinc	0.004	-0.001	00
	[0.002]*	[0.002]	[0.002]
hhinc^2	-3.73e-09	1.08e-10	-6.48e-10
	[1.76e-10]***	[1.84e-10]	[1.73e-09]
hhinc01	-0.003	-0.004	-0.004
	[0.003]	[0.004]	[0.003]
hhinc00	0.002	0.006	0.001
	[0.002]	[0.005]	[0.003]
hhinc99	-0.0002	-0.013	-0.004
	[0.0001]	[0.005]**	[0.004]
hhworked01	0.004	0.03	0.026
	[0.035]	[0.045]	[0.046]
hhworked00	-0.022	-0.089	-0.063
	[0.033]	[0.043]**	[0.044]
hhworked99	0.05	0.042	0.06
	[0.030]*	[0.041]	[0.037]
saving	0.00001	-3.07e-06	0.00001
	[9.41e-06]	[9.01e-06]	[7.27e-06]*
debt	6.10e-06	-5.16e-06	2.37e-06
	[3.49e-06]*	[3.76e-06]	[3.65e-06]
death	0.037	0.021	0.063
	[0.020]*	[0.026]	[0.028]**
unemp	0.101	0.022	0.117
	[0.017]***	[0.023]	[0.025]***
bust	-0.237	0.163	-0.236
	[0.118]**	[0.117]	[0.109]**
disaster	0.125	-0.048	0.162
	[0.029]***	[0.044]	[0.047]***
poverty	-0.129	0.16	-0.095
	[0.011]***	[0.017]***	[0.015]***
Household size dummies	Yes	Yes	Yes
State dummies	Yes	Yes	Yes
Doctor visit dummies	Yes	Yes	Yes

Robust standard errors in brackets; clustering at the locality level. Note: hh=household; hoh=head of household; ptr=partner (e.g. hhworkedXX=dummy for whether household head was employed in year XX). Unless otherwise specified, all variables are from 2002. We provide a more detailed variable description in the Appendix. *, **, *** mean significant at 10%, 5%, 1%.

of 1999 income and 2000 employment status, both of which point out to higher enrolment to the program for households hit by a shock.

Lastly, we proceed to compare the estimates from the first two propensity scores. As we mentioned above, the presence of variables that are balanced among compliers and non-compliers within treatment areas, but not between poor households in treatment and control blocks, may cause one of the identification assumptions to fail. For example, consider the case of household heads' unemployment status (the variable *unemp*). Its partial effect as a determinant of program participation is small, 0.02, and not significant. However, household heads in treatment areas are 10 percentage points more likely not to have a job than household heads in control areas. When we match households based on the second propensity score, we are comparing treated families with control families with a lower proportion of unemployed heads. In this case, it seems unrealistic to assume that treated households' change in consumption in the absence of the program would have been the same as the observed change in consumption of households with different characteristics. More generally, when we fail to control for variables that do not seem to affect participation, but which ar not balanced between treatment and control areas, we end up with biased estimates of the treatment on the treated effect. For example, if the change in consumption in the absence of the program is positively correlated with household wealth, the consumption change for households in control areas is higher than the change for treated households in the absence of the program, thus matching using the imputed propensity score underestimates the ATT.

4.3 Estimates of AIT and ATT effects

Table 3 provides estimates of the Average Intent to Treat and Treatment on the Treated effects on the level of food consumption using different estimators. Columns 1.1 and 2.1 show the estimates of the AIT and ATT from our favorite estimators, based on equations (4) and (12). The remaining columns provide propensity score matching estimates of the ATT. In column 2.2 we are comparing compliers and non-compliers in treatment areas, while in the remaining two columns we compare compliers with similar households from control areas (in column 2.4 we imput the probability of participation for poor households in control areas using the coefficients from a probit for compliance).

The estimated AIT from column 1.1 reveals that food consumption is about 100 *pesos* larger for eligible households in treatment areas, compared to households with similar observable characteristics who live in control areas. This parameter is identified using a more general set of assumptions than the ones required for each of the ATT estimates, as we discussed above,

Table 3: Average Intention to Treat and IV Average Treatment on the Treated estimates for food consumption, monthly *pesos*.

AIT	ATT			
(1.1) Matching $Pr(Z=1\|X)$	(2.1) IV with ps $Pr(Z=1\|X)$	(2.2) Matching $Pr(D=1\|Z=1,X)$	(2.3) Matching $Pr(V=1\|X)$	(2.4) Matching (imputed)
101.90	270.81	106.14	126.62	110.87
[20.14; 240.84]	[111.63; 552.49]	[47.15; 166.72]	[16.32; 340.87]	[22.96; 183.66]

The participation proportion is 0.475. Block-bootstrap bias-corrected 95% confidence interval in brackets, the block is the area. 200 repetitions. Local linear regression matching estimates. The estimates from llr are similar to the ones obtained using the 4 nearest neighbors with replacement.

and is a lower bound of the ATT even in case of spillover effects for non-compliers, as long as they are smaller than the program effect on the treated. The estimates of the ATT effects differ considerably: while using our favorite estimator, in column 2.1, we find an effect of 270 pesos, the other three point estimates are more than twice as small, ranging from 106 to 126 pesos. Given that the average grant received by these households is around 350 pesos, the IV estimates of the ATT suggest that the program has a large effect on food consumption, and that perhaps a small part of the transfer is saved, similarly to what happened in rural areas (in unreported regressions we estimated a small and insignificant effect on non-durable, non-food consumption, and a very small and significant effect on durable consumption). There are several reasons why we believe that the matching estimates of the ATT effects are too small to be plausible. First, from an economic viewpoint, a 100 peso increase in food consumption would imply an astonishing saving rate among these poor households. Second, the fact that the estimates of the ATT are as low as the AIT suggests that the program has huge spillover effects, and that its effect on non-compliers is as big as the effect on compliers. Lastly, we expect the matching estimates of the ATT to be downward biased, as discussed in the previous section.

5 Conclusions

This papers deals with the identification and estimation of Average Treatment on the Treated effects from non-experimental data. A treatment is offered to a group of eligible subjects, and a

set of individuals with similar observable characteristics, but who are not offered the program, is used as control group. Researchers often use matching type estimators when using data with this structure, assuming that one observes all variables that affect both participation and outcome. We study the case in which many individuals in the treatment group decide not to take up the program, although it is offered to them. In this situation, we argue that the matching assumptions may be unrealistic when one tries to balance not only the non-random assignment of the program but also non participation, if one suspects a strong self-selection into the program based also on transitory shocks. In such a case, assuming that assignment to the treatment or control group is random, conditional on observables, may be less stringent than assuming that selection into the program is conditionally independent from potential outcomes. We propose an estimator which uses the conditional independence of group assignment as an exclusion restriction, and which estimates the ATT dividing the Average Intent to Treat (AIT) by the probability of participation in the treatment group. We estimate treatment on the treated effects using the probability of being in the treatment group to match individuals from the treatment group with observationally similar subjects from the control group, and divide the difference in observed outcomes by the average probability of participation for individuals with the same (or similar) values of propensity score. We obtain the ATT by integrating these effects over the density of the propensity score among program participants.

To illustrate the different methods discussed in this paper we have presented an empirical example that uses non-experimental data from the Mexican conditional cash transfer program *Oportunidades* to estimate the ATT effect on food consumption. We obtain substantially different results using our favorite procedure and matching: while the ATT estimates using our IV procedure are consistent with our expectations, the matching estimates suggest that household save most of their grant, rather than spending it, contradicting the existing empirical evidence.

References

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

[2] Angelucci, M., Attanasio, O. and J. Shaw :"The effect of Oportunidades on the level and composition of consumption in urban areas", in *External Evaluation of the Impact of* Oportunidades *Program 2004: Education*, Henandez-Prado, B. and M. Henandez-Avila, Eds., Chapter 3, Vol. 4, 105 - 152, 2005.

[3] Angrist, J.D., Imbens G.W. and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 444-455.

[4] Gertler, P., Martinez, S. and M. Rubio (2005) "Investing Cash Transfers to Raise Long Term Living Standards", Mimeo.

[5] Heckman, J.J. (1996): "Identification of Causal Effects Using Instrumental Variables: Comment", *Journal of the American Statistical Association,* Vol. 91, No. 434, pp. 459-462.

[6] Heckman, J.J.; Smith, J. and C. Taber (1998): "Accounting for Dropouts in Evaluations of Social Programs", *The Review of Economics and Statistics*, Vol. 80, No. 1., pp. 1-14.

[7] Hirano, K., Imbens, G, Rubin, D., and Zhou, X. (2000), "Assessing the Effect of an Influenza Vaccine in an Encouragement Design with Covariates," *Biostatistics* 1, pp. 69-88.

[8] Hoddinot, J. and Skoufias, E., (2003) " The impact of Progresa on food consumption", IFPRI WP No 150, http://www.ifpri.org/divs/fcnd/dp/papers/fcndp150.pdf.

[9] Imbens, G.W., Angrist, J.D., (1994) "Identification and estimation of local average treatment effects", *Econometrica*, Vol. 62, pp. 467–476.

[10] LaLonde, R. J. (1995), "The promise of public sector-sponsored training programs", *The Journal of Economic Perspectives*, Vol. 9, No. 2, pp. 49-68.

[11] Moffitt, R.A. (1996):"Identification of Causal Effects Using Instrumental Variables: Comment", *Journal of the American Statistical Association,* Vol. 91, No. 434. , pp. 462-465.

[12] Rosenbaum, P., Rubin, D. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, pp. 41–55.

[13] Rubin, D., (1980), "Discussion of Randomization Analysis of Experimental Data: The Fisher Randomization Test by D.Basu", Journal of the American Statistical Association, 75: 591-93.

[14] Rubin, D, (1986), "Which ifs have causal answers? Discussion of Hollands Statistics and causal inference", *Journal of the American Statistical Association*, 81: 961-62.

[15] Smith, J.A., and Todd, P.E. (2005), "Does matching overcome LaLonde's critique of non-experimental estimators?", *Journal of Econometrics*, Vol. 125, pp. 305-53.

[16] Todd, E.P., Gallardo-Garcia, J., Behrman R.J., and S.W. Parker (2006), "*Oportunidades* Impact on Children and Youth Education in Urban Areas after One-year of Program Participation", in *External Evaluation of the Impact of* Oportunidadesu *Program 2004: Education*, Henandez-Prado, B. and M. Henandez-Avila, Eds., Chapter 3, Vol. 1,

6 Appendix: variable description

Consumption. Food expenditure was calculated for each category in 2002 and 2003, excluding households with more than 18 (out of a total of 37) missing food consumption responses.

Food includes:

- Starch - Maize tortillas, bread, pasta, kidney beans, rice, crisps, flour, maize, other cereals, etc.
- Protein - Red meat, chicken, pork, canned fish, fish, seafood, eggs, milk, cheese, other animal products.
- Fruit and vegetables - Tomatoes, onions, potatoes, chile peppers, carrots, pumpkin, bananas, apples, oranges, other fruits and vegetables.
- Miscellaneous food - Sugar, soft drinks, water purification tablets, coffee, vegetable oil, fried potatoes, fried pig skin, other manufactured food.
- Food eaten outside the home.

Household Income. We add incomes for the main and second job of all household members and other sources such as other jobs, pensions, and compensations. Incomes from domestic helpers and their relatives, and individuals whose relationship to other family members is missing, are not included.

Variables used to calculate our propensity score. With the exception of past employment and earnings information, all variables relate to 2002.

hhinc; hhincsq - *Household Income; squared(see above)*
saving - *Household saving excluding domestic helpers and their relatives, and individuals whose relationship to other family members is missing.*

Dmissav - *Dummy for missing household saving*

Dtothh1-5 - *Size of household dummies for households with between 1 and 5 members.*

Dnohhsize - *Dummy for missing size of household*

Dnot1hoh - *Dummy for households recorded as having more than one head.*

nokids0-5, 6-12, 13-15, 16-20 - *Number of children in the household in the specified age range.*

sfem -*Dummy for female headed household*

Dnopartner-*Dummy for households where there is no partner.*

hohlit - *Head of household is literate*

ptrlit - *Partner is literate.*

hoheduc - *Head of household education level (included as a series of dummies). The categories are: no qualification, incomplete primary, complete primary, incomplete secondary, complete secondary, higher education.*

ptreduc - *Partner education level (included as a series of dummies, same categories as for head).*

hohdoc - *Head of household went to a conventional doctor in last four weeks.*

ptrdoc - *Partner went to a conventional doctor in last four weeks.*

nosch0-5, 6-12, 13-15, 16-20 - *Number of children in the household in the given age range attending school.*

kiddoc - *At least one child in household went to conventional doctor in last four weeks.*

hohemployee - *Head of household is an employee.*

ptremployee - *Partner is an employee.*

hohse - *Head of household is self-employed.*

ptrse1 - *Partner is self-employed*

misshohemp - *Head of household employment status is missing.*

missptremp - *Partner employment status is missing.*

hhworked01, 00 ,99 - *Either head of household or partner worked in 2001, 2000, 1999.*

hhinc01, 00, 99 - *Income of head and partner in 2001, 2000, 1999.*

misinc01, 00, 99 - *Dummy for missing income in 2001, 2000, 1999.*

death - *Household suffered death or illness of non-resident family member in the last year.*

unemp - *A resident family member was unemployed in the last year.*

bust - *A resident family member lost a business in the last year.*

disaster - *Household suffered a natural disaster in the last year.*

poverty (and its square) - *Socio-economic household-specific poverty index (program eligibility is based on this index);*

entidad6-30 - *State dummies.*