Evaluating the urban component of *Oportunidades.* Which methods for which parameters?*

Manuela Angelucci† Orazio Attanasio‡

February 2006

Abstract

In this note we discuss several methodological issues connected with the evaluation of the effects of the urban component of the program *Oportunidades.* The evaluation of *Oportunidades* is based on the comparison between treatment and control areas at two points in time. However, standard methods are complicated by the presence of a relatively low take up of the program. We start by considering two parameters, the Average Intent to Treat (AIT), which ignores the issue of participation into the program and simply compares treatment and control areas, and the Average Treatment on the Treated (ATT), which measures the effect of the program on beneficiaries. The first parameter can be estimated by standard diff-in-diff matching methods. We discuss two alternative approaches for the second one. The first uses matching to control for non random participation and to balance characteristics of households in control and treatment areas. The second exploits the fact that the ATT can be expressed as the AIT divided by the participation share, in the absence of spillover effects. We propose a semi-parametric estimator that exploits the flexibility of matching estimators within a standard Instrumental Variable approach. We discuss the different assumptions necessary for the identification of the ATT with each of the two approaches. In the final section of the note, we apply the different methods discussed to the estimation of the AIT and ATT of the program on consumption. We argue that estimates of the ATT obtained by IV seem more plausible than those obtained by matching.

*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.

†Department of Economics, University of Arizona, and IZA

‡University College London, NBER, BREAD, and CEPR.

1 Introduction.

In this note we discuss alternative approaches to the evaluation of the recently implemented *Oportunidades* program in urban areas. *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. *Oportunidades* is now the largest welfare program in Mexico by a large margin and covers more than one in ten Mexicans. As PROGRESA, *Oportunidades* is being rigorously evaluated using a systematic and large data collection effort. However, the methodological problems involved in the evaluation of *Oportunidades* are different from those relevant for the evaluation of PROGRESA for two important reasons. First, the allocation of the program within the localities in the evaluation sample was not random. Second, maybe due to the different procedure through which the registration into the program had to be carried out in urban areas, the enrollment rate of eligible household was, a year after the start of the program, much lower than in rural areas.

In this note we discuss different parameters of interest related to the impact of the programs and alternative assumptions under which they can be estimated with the available data. To provide concrete examples of the issues and of the methods we discuss, we will analyze the effect of the program on household consumption. Consumption has an important advantage in that one has strong priors about the effect of the program on such a variable, partly derived from the intuition that an increase in (permanent) income should be reflected in an increase in consumption and partly derived from other studies.

The rest of this note is organized as follows. In section 2 we briefly describe the program and its evaluation. In sections 3 and 4, we discuss the methodological issues. In particular, Section 3 discusses identification issues, while section 4 discusses estimation. Section 5, presents some results on consumption. Section 6 concludes.

2 *Oportunidades* and its evaluation.

From an evaluation point of view, the original PROGRESA program had two attractive features. First, the progressive geographic phasing-in of the program was exploited to create a randomized sample of 506 villages for evaluation purposes. In particular, the complex logistics of the program made it impossible to reach all the 50,000 targeted localities simultaneously: the expansion process took about two years. Given this situation, the 506 localities in the evaluation sample were randomly divided into two groups. In the first group, made of 320 localities, the program started in June 1998. In the second group, of 186 localities, the program did not start until the year 2000. Second almost all eligible households were incorporated into the program or, at least, households that decided not to participate into the program were very few. The success of the program in terms of coverage was due in part to the its apparent attractiveness in rural areas and in part, probably, to the registration process used in rural areas. Once a rural locality had been targeted, a census of all households living in the locality was taken (the pre-program Survey of Rural Household Socioeconomic Characteristics so-called ENCASEH survey). Based on a (small) subset of variables observed in that survey, each household was designated a beneficiary or not. This ensured that eligibility status was known by the household with certainty before enrolling to the program.

The expansion and incorporation processes of urban Oportunidades in 2002 were very different. First, allocation across geographic areas was not random, complicating the evaluation of the program (as discussed below). Second, a census to establish eligibility in all targeted communities was deemed not feasible. Instead, registration offices (módulos) were set up in some locations within eligible areas, and resources invested in spreading the news about the availability of the program in that area. To apply for the program, potential beneficiaries were supposed to visit the local office (módulo), where they would be administered a questionnaire (a version of the ENCASEH questionnaire, so-called ENCASURB since it referred to Urban Households), establishing preliminary eligibility. Households deemed eligible on this preliminary basis were visited by administration officials to verify the information provided in the questionnaire administered in the modulo, determining final eligibility. 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, according to self reported estimates in the tamizaje survey (see below for details) the fraction of eligible households registered onto Oportunidades was only 0.335. Based on administrative data, the coverage is higher, but does not reach 50%.

The fact that many eligible households did not sign up for the program has obvious implications both for the effectiveness of the program and for the techniques it is appropriate to use in its evaluation. With this in mind, in what follows we discuss the desirability of estimating both the so-called 'intent to treat' impact of the program (which ignores the fact that a substantial proportion of eligible households do not register for the program) and the effect of the 'treatment on the treated'.

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 the treatment areas are potentially very different from the control areas. The evaluation advisory group decided to sample manzanas[1] within treatment areas and, to each manzana in the treatment sample, to match a control manzana that would be similar in terms of a pre-estimated propensity score. Of course 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 aeras with the highest concentration of poor households, using such a variable would give no intersection between treatment and control samples.

In PROGRESA and rural Oportunidades, the way in which households were registered for the program enabled the evaluation database to be linked directly to the administrative database – in fact, the rural evaluation survey (so-called ENCEL) data set is an expansion of the ENCASEH administrative database. The same strategy could not be used in the case of urban Oportunidades because, as we mentioned above, many eligible (poor) households did not sign up for the program – so surveying only poor incorporated households could not be guaranteed to provide a representative sample of poor households. The evaluation advisory group decided to sample a number of blocks (manzanas) where they would conduct a census. Some of these manzanas were in areas where the program was about to start in 2002 (treatment), while others were not incorporated yet (control). Treatment and control manzanas were selected to be comparable across various dimensions; nevertheless, some differences between the two sets of manzanas were unavoidable. The census was called the tamizaje, and asked enough questions to allow the construction of the poverty index used to determine eligibility for Oportunidades itself. The total number of households in the tamizaje is 37,489, in the manzanas included in the evaluation study.

Once the tamizaje had been conducted, a sample was selected for the main socio-economic questionnaire (the questionnaire that provides all the consumption data we use), which is part of the Urban Evaluation Survey or "ENCELURB". The sample was made of 'poor', 'quasi-poor' and some non-poor households. While in the control areas, pre-established proportions of the three groups were included in the sample, in treatment areas, the sampling procedure was a little more complicated. As before, households from the three groups were sampled, but this time information about whether the household was incorporated into Oportunidades was also used. The original plan was to include about 6,000 incorporated households and 2,000 non-

[1] A *manzana* or block is much smaller than an area.

incorporated households. However, since far fewer households registered for the program than expected, the required number of incorporated households was not reached. As a result, on the one hand, all eligible incorporated households (according to the self reported status) were included in the sample. And the decision was taken to add a number of localities (typically in the vicinity of selected treatment manzanas) to the sample (named 'barrido' manzanas). Only poor incorporated households living into these localities were sampled. Moreover, for these barrido blocks, no tamizaje was conducted.[2]

The sampling scheme outlined above has an important implication. The fraction of poor participating into the program observed in our treatment sample is quite different from the true fraction of program participants. In other words, the final survey is, to a certain extent, what is called a 'choice-based' sample, as the probability of being included into the sample depends on the participation to the program (that is, which households were interviewed depended partly on whether they chose to apply for Oportunidades). Fortunately, we can estimate the true proportion of participating households in each manzana from the tamizaje.

The baseline socio-economic questionnaire was first conducted in 2002, after households had registered for Oportunidades, but before any payments had been made. A year later, the same households were interviewed using a similar questionnaire. Together, the results from these surveys make up the ENCELURB database (16,012 households were interviewed in both 2002 and 2003). This database contains a wide range of information. For the analysis in this paper the consumption questions in both years are obviously very important. Here it is worth stressing that information on the consumption of many commodities is available, and includes (at least for food) consumption in kind. The information is collected with retrospective questions that refer to the past week, month or six months (depending on the commodity considered), but in processing the data we transform all the figures to monthly equivalents.

In what follows, we discuss the problems involved in estimating the effect of the program on the 'poor', that is on eligible households. We neglect estimators that use regression discontinuity design idea by using the 'quasi-poor' as a control for a part of the eligible sample. For expositional simplicity, we also neglect the issues involved with the choice-based nature of the evaluation sample. All the arguments we present below can be easily adjusted and the estimates re-weighted to take into account that treatment areas over-weight participants into the program.

[2]Some of the groups in the evaluation team used the *barrido* blocks, while others did not. For the consumption analysis, this choice does not make much difference.

3 Identification of program impacts.

In this section we define the parameters of interest 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.

Before discussing identification, we need to stress that we observe the variables of interest and a large set of individual characteristics in 2002, the year prior the beginning of Oportunidades, and 2003, its first year of implementation. Hence, we can implement difference-in-difference (DD) 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. Hence, our estimate take into account the possibility that pre-program levels may differ between the groups that we compare, and are based on the assumption that, had the program not been implemented, they would have changed by the same magnitude. We do not discuss the assumptions needed for the validity of the (conditional) DD estimator as they are completely standard and will be maintained across the various estimators we consider below.

3.1 Identification of AIT and ATT

We are interested in estimating two main parameters: average intention to treat effects (AIT) and average treatment on the treated effects (ATT). Both of these parameters are of interest from a policy point of view. The first parameter, AIT, estimates the average effect of Oportunidades on the consumption of all eligible households, irrespective of their participation to 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. In other words, under plausible assumption and with an estimate of the AIT we will only be able to state that the ATT is at least as large as the AIT. 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 households.

We define the notation as follows. The superscripts T and C denote area of residence - treatment and control, respectively. As usual, y_1 and y_0 denote potential outcomes with and without the treatment, while y denotes observed outcomes. There are $i = 1, N$ households, observed in $t = 1, 2$ time periods. The program starts in the second period, $t = 2$. Thus,

$y_1^T(it)$ and $y_0^T(it)$ are the potential outcomes with and without the treatment for household i living in a treatment area at time t, while $y(it)$ is the household's observed outcome. The parameter β is the treatment effect, i.e. $\beta^T(it) = y_1^T(it) - y_0^T(it)$ is the treatment effect at time t for household i living in a treatment area. Analogous quantities can be defined for control areas. Of course only some of these quantities are observables. In control areas, $y_1^C(it)$ is never observed as the program is not available. In treatment areas, if all households participated, we would not observe $y_0^T(it)$. In reality, however, as only some households participate into the program, we observe $y_1^T(it)$ for some households and $y_0^T(it)$ for others. The following equation express observed outcomes as a function of potential outcomes:

$$
\begin{aligned}
& y^k(it) = y_0^k(it) + \beta^k(it)D(it) \\
& k = \{T, C\}
\end{aligned}
$$

$D(it)$ is an indicator that takes the value 1 for households who participate and 0 for households who do not. $D(it)$ is 0 for all households in control areas, and for treatment households in $t = 1$. However, for households in treatment areas at time 2, $D(it)$ is a decision variable and, as such, is likely to be correlated with potential outcomes. Lastly, we define $\Delta^k y(i) = y^k(i2) - y^k(i1)$, for $k = \{T, C\}$, and $\Delta y(i) = y(i2) - y(i1)$ as the difference in potential and observed outcomes for household i living in area k. Given this notation we can define the following parameters.

Average Intention to Treat. Neglecting both household and time indices for simplicity (but bearing in mind that all treatment effects are potentially identified only in $t = 2$, after the program is implemented), we define the AIT effect as:

$$
E(\beta^T D) \tag{1}
$$

The AIT is an interesting policy parameter, because it measures the effect of the program on eligible households, regardless of whether they participate or not into the program. As long as the policy maker has little influence on participation (which might or might not be the case) the AIT represents the average effect on eligible households and is therefore one of the most relevant parameters for policy analysis. The AIT is also interesting because it provides a lower bound to the ATT if the program has no positive effect on the non participants in treatment areas.[3]

Since we do not observe what the variable of interest in treatment areas would have been in

[3] We have made these statements relative to a positive effect. Similar considerations hold for a negative one. Effectively, when the program does not affect non-participants, the AIT is a scaled version of the ATT.

the absence of the program, we require some assumptions that allow us to use the 'control' areas to estimate this counterfactual. It should be remembered that the assumptions refer here to the change in the outcome of interest. Constant unobservables are allowed.The basic assumption used here is that of 'selection on observables', that is, that conditional on a set of observable variables, the assignment of the program across treatment and control manzana is random. In particular, we (as it is standard in this literature) make the following two assumptions. :

- Assumption (1.1) - Stable Unit Treatment Value Assumption (SUTVA).[4] 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 consumption change of poor households living in C areas is not affected by Oportunidades.
- Assumption (1.2) - Conditional Indepedence Assumption (CIA). Conditional on observables (X), the expected value of changes in unobservable individual characteristics does not systematically differ between treatment and control areas.

These two assumptions imply that $E(\Delta y_0^T(i) \mid X) = E(\Delta y^C(i) \mid X)$. This solves the missing counterfactual problem and allows to compute the AIT by comparing observed outcomes of households in T and C areas, conditional on a set of variables X: $E(\Delta y^T(i) \mid X) = E(\Delta y^C(i) \mid X) = E(\Delta y_0^T(i) \mid X) - E(\Delta y^C(i) \mid X) + E(\beta^T D|X)$. The parameter in 1 is then estimated integrating the difference $E(\Delta y^T(i) \mid X) = E(\Delta y^C(i) \mid X)$ over the relevant values of X.[5]

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

$$E(\beta^T|D = 1) \tag{2}$$

This parameter measures the effect of the program on households who actually receive it. If participation into the program (compliance) were complete, the ATT and the AIT would coincide. Analogously to what discussed above for the AIT, we cannot observe what the change in consumption for participants would have been in the absence of the treatment. There are at least two ways to estimate this effect. One is to use some DD matching estimator. The

[4] Formalized by Rubin (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 households. 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.

[5] If one is willing to accept the validity of SUTVA and CIA, the only potential problem is that of non-overlapping support. It is possible that some treated households have no control households to be matched to. In this case, the AIT would be only estimated for the subset of households for whom a match can be found. A related issue is that of the lack of precision caused by having few control households over some parts of the support. These issues, that are standard in the matching literature, are not the main point of this paper and will only be noted.

idea is to condition on observables in the hope that, conditional on such variables, not only the assignment of the program across regions, but also participation into the program is random. The alternative is to exploit the idea that the AIT constitute a bound on the ATT and unravel the relationship between the two using information on participation rates. 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.

3.2 *Matching approaches to participation.*

If one decides to follow a matching approach to estimate the ATT, there are several alternatives that correspond to the use of different control groups and different versions of the same identification assumptions. In general, consider the set of compliers and some control group. The identification assumptions are:

- Assumption (2.1.a) - SUTVA. One consequence of the SUTVA is that the change in the variable of interest for the control group does not depend on Oportunidades.
- Assumption (2.2.a) - CIA. Conditional on observables, the expected value of the change in the variable of interest for the control group is the same as the change for compliers in the absence of the program.

Given this set of assumptions, the following equality holds: $E(\Delta y_0^T|D = 1, X) = E(\Delta y_0^C|X)$, where the latter object is the change in observed outcomes for the control group. For example, one could use non-compliers in treatment areas as a possible control group for participants in the same areas. In this case, the parameter in (2) is identified by the difference $E(\Delta y^T|D = 1, X) - E(\Delta y^T|D = 0, X)$.[6] An advantage of comparing eligible households within treatment areas is that one does not need to worry about heterogeneity across areas. However, a disadvantage is that the SUTVA might not be plausible when is referred to individuals in treatment areas, because it ignores the possibility of program spillover to non-compliers.

One of the possible alternative approaches within the matching framework is the one proposed by Petra Todd and used in most of the evaluation. This approach is based on the comparison of compliers and poor households from control areas, i.e. individuals with similar observable characteristics who live in areas where the program is not available. It consists of estimating a propensity score using compliers and non-compliers in treatment areas, in order

[6]This is because the above difference equals $E(\Delta y_0^T|D = 1, X) - E(\Delta y^T|D = 0, X) + E(\beta|D = 1, X) = E(\beta|D = 1, X)$ by assumption.

to compare compliers with households in control areas with a similar propensity score. For this latter group, the propensity score is imputed using the coefficients estimated from households in treatment areas. The idea behind this choice of comparison group is that it avoids making the assumption that the program has no effect on non-participants in treatment areas. We simply need to assume that the program does not affect the change in outcomes of households living in different areas, which is more plausible given the bigger (economic and geographic) distance between these two groups. At the same time, though, the Conditional Independence Assumption becomes much stronger: as before, we require that, conditional on the variables used in the propensity score, participation into the program is random. That is, we have to condition on all variables that affect both participation and outcome (this is a standard requirement with matching estimators). However, this particular type of matching further requires that the observable variables that balance the sample of participant and non participants in treatment areas also balance treatment and control areas. It is easy to think of examples that violate this second assumption. Let's consider two conditioning variables, X_a and X_b, both of which are observed. For instance, assume X_a is the number of school age children, and X_b household head's unemployment status. Suppose that both variables affect consumption, but only X_a determines participation in treatment areas. In other words, the X_b variable is balanced between participants and non-participants in treatment areas, hence $P(X_a, X_b) = P(X_a)$. This means that households in treatment and control areas are matched only on X_a (when we regress the probability of participation on X_b, its coefficient is zero). However, if fewer individuals are unemployed in control areas, X_b is not balanced between treatment and control households. In this particular example, in order for the ATT to be identified we need to assume that household heads' employment status does not influence the change in consumption, or, more generally, that the change in potential outcomes without the treatment is the same in expectation for households with different X_b characteristics. It is plausible to think, instead, that the effect of the differences in X_b will confound the effect of the program on participating households. Note that we picked a time varying variable X_b in this example, as fixed differences among the outcome determinants would not create a problem in our context because of the Diff-in-diff approach.

A similar argument can be made for yet another possibility, to estimate a propensity score that predicts area of residence, and then compare treated households with households in control areas who have the same propensity score. In this case, we are balancing observables between areas, but nothing insures us that we are also balancing the variables that determine program participation (and outcome).

3.3 *A LATE approach to participation.*

Obviously, the sets of assumptions needed to identify the ATT are much stronger than the ones required for the identification of the AIT effect. The approach we discussed in the previous section burdens the matching procedure with the task of balancing both the participation into the program and the non random assignment of the program across areas. It is therefore worthwhile exploring an alternative strategy to identify the missing counterfactual, which consists in using the type of area of residence (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)[7], with the difference that in our context, unlike in these two papers, the issue of non-compliance arise on top of a non-random assignment of the program, In this setting, we need the following 5 assumptions to identify the parameter of interest:

- Assumption (3.1) - SUTVA. Program participation and the outcome of interest y for eligible households living in control and treatment areas, *including non-compliers*, are not affected by the participation status of others.

- Assumption (3.2) - CIA. Conditional on observables, unobservable characteristics of individuals do not systematically differ between treatment and control areas.[8]

- Assumption (3.3) - Exclusion restriction. Conditional on observables, the outcome y in the absence of the program is not affected by the type of area of residence (treatment or control).

- Assumption (3.4) - Nonzero average causal effect of area of residence on program participation. Being in treatment areas increases the likelihood of participating to the program.

- Assumption (3.5) - Monotonicity. There is no individual who would have participated to Oportunidades if he or she had lived in a control area, but not if he or she had lived in a treatment area.

With this approach one normally identifies the LATE, i.e. the average treatment effect for the set of agents who are induced to participate to the program because of the instrument, called compliers in the literature. In this particular case, though, the instrument is the type of area of residence and the participation rate in control areas is zero. Thus, the households

[7] See also the discussion in Heckman (1996) and Moffitt (1996).
[8] This is an implication of the ignorability assumption made by Angrist, Imbens and Rubin (1996).

who are induced to participate to Oportunidades because they live in treatment areas are all the treated households. Hence, given the above assumptions this instrument identifies the ATT effect:

$$\frac{E(\Delta y^T|X) - E(\Delta y^C|X)}{E(D = 1|X)} = E(\beta^T|D = 1, X) \tag{3}$$

Following this approach the ATT is given by the AIT divided by the proportion of participants. Assumption (3.4) is obvious in this context, as the program is available only to households living in treatment areas, and assumption (3.5) seems quite realistic. Hence, the main differences between identifying ATT effects by matching or by IV are related to differences in the SUTVA, CIA, and exclusion restriction assumptions.

3.4 Comparing the two approaches

In the matching cases we have to assume that, conditional on observables, the participation into the program of households in treatment areas is unrelated to potential outcomes. This implies observing all variables determining both participation and outcome, including transitory shocks. Note that this holds also for the second type of matching proposed, where eligible participants are compared to households in control areas. On the positive side, it should be stressed that the assumptions require that participation is unrelated to the change in the no-program outcomes. Even with diff-diff matching, participation could be related to the program effect, as long as its correlation comes through the outcome after the program. That is, the assumptions do not rule out the decision D being correlated with post-program outcomes. People who anticipate a large benefit can be the ones that select in.

In the IV case, instead, we relax this assumption, because we only require conditional independence for area of residence, rather than for participation, and the exclusion restriction to hold. In other words, we only need to observe all the variables by which households in treatment and control areas differ. While unobservable transitory shocks may differ between participants 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.a), and eligible households of control areas (as in 2.1.b). This may not be the case if the program has any effect on the local goods, labor, and financial markets that may change non-participants consumption. 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 be likely to change. The absence of indirect program effects on non-participants is a non-testable assumption. However, we suspect that these indirect effects are unlikely to occur in the case of some variables such as consumption, given that the program has only been implemented for a few months. It is clear, however, that such an assumption might be more problematic for other variables.

4 Estimation of program impacts

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

4.1 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 (at least in non linear transforms of the variables used as controls), one can estimate the AIT by a simple OLS regressor where y is regressed on various controls and a dummy for area of residence. Alternatively, if one wants to use a less parametric approach, one can use Propensity Score Matching techniques where one first computes the probability of being in a treatment area as a function of some variables X, and the compares changes in y for households in treatment areas to changes in y for households in control areas with similar propensity scores, i.e.

$$E(\beta^T D) = E(\Delta y^T - \Delta y^C) \tag{4}$$

$$= \int_p [E(\Delta y^T|P(X) = p) - E(\Delta y^C|P(X) = p)]dF(p|Z = 1) \tag{5}$$

where $P(X) = E(Z|X)$ and Z is an indicator variable that takes the value 1 for observations in treatment areas. This comparison is by necessity limited to the subset of treatment and control households that lie in the common support of the propensity score. In the specific case we are considering we only need to keep in mind some important caveats specific to the exercise we are discussing.

- As stressed above, the conditional independence assumption is made on the changes in y, not its level. Working in first difference has the desirable feature that fixed (area level) unobservable differences between treatment and control samples are eliminated. The same of course is not true for time varying unobservables.

- The deterministic assignment of the program across areas prevents the use of specific area level variables as determinants of the propensity score, as they would make the common support empty. For instance, if one were to condition on the proportion of poor in a given area, one would perfectly discriminate between treatment and control areas, ending up with no common support. Similarly, using level variables correlated with the proportion of poor would make the common support very small and result in very imprecise estimates. There is nothing that can be done about this issue: it is an undesirable, yet unavoidable feature of this evaluation exercise.

In the application below, we employ both propensity score matching techniques and linear controls. Controlling for a wide variety of individual variables does not control common support problems. The same is true for a number (but obviously not all) area level variables.

4.2 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 households in treatment areas 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 try to 'inflate' the AIT estimates using an Instrumental Variable approach. We discuss the estimators in turn.

4.2.1 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 control groups. The first is given by non-participant household in treatment areas. The second is given by eligible households in control areas.

1. *Comparing participants and non participants in treatment areas.* The first step in this procedure is to estimate a participation equation using treatment area eligible individuals. These estimates are used to compute propensity scores that are then used in a second step to compare participants and non participants eligible in treatment areas.

2. *Comparing participants in treatment areas to eligibles in control areas.* The first step of this procedure is the same as the one above. The coefficients of the participation model are then used to impute a propensity score to eligible households in control areas. Participants in treatment areas are then matched to eligible in control areas with similar probabilities to participate into the program if it would have been available in control areas.[9]

Since these methods are commonly implemented in the literature, we do not discuss their implementation in details. However, 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 households in control areas. However, as we discussed above, this comes at the cost of using only information on participation in treatment areas to balance both treatment and control areas, and participants and non participants within the treatment areas.

4.2.2 An IV approach to estimate the ATT

A different approach is to start from the idea that the AIT constitutes a lower bound to the ATT. Suppose first that one is willing to assume that the average effect on non participants is zero. Then a simple estimate of the effect is simply obtained by dividing the AIT by the participation rate. In practice, one has to construct the sample equivalent of equation 3. Alternatively, one can assume that such an assumption is true conditional on a set of variables X. As the participation level (as well as the effect), one obtains a different estimate of the ATT. In particular, in this case the average treatment effect will be given by:

$$E(\beta^T|D=1) = \frac{E(y^T) - E(y^C)}{E(D|Z=1)} = \tag{6}$$

$$\int_x \frac{E(y^T|X=x) - E(y^C|X=x)}{E(D|Z=1, X=x)} dF(x|D=1) \tag{7}$$

An estimator can then derived by considering the sample equivalents of the moment in equation

[9] One could also combine the two groups. That is one could compare participants in treatment areas to eligibles in control areas and non participants in treatment areas. That is one would be pooling eligible households in control areas with all participating households in treatment areas, and compute the following propensity score: $P(X) = Prob(V = 1|X)$, where $V = 0$ for all households in control areas and for non-compliers in treatment areas, and $V = 1$ for all participating households in treatment areas.

7.[10] In particular, one can first estimate the probability of participating into the program using data from the treatment areas. This probability of participation can be computed as a function of observables both in treatment and control areas and used to 'inflate' the changes in y in both treatment and control areas. One can then compare the 'inflated' changes between treatment and control areas using matching techniques aimed at balancing the treatment and control areas.

If the change in potential outcomes is independent of area type conditional on X, it is also independent conditional on the propensity score, $P(X) = E(Z|X)$ (Rosenbaum and Rubin, 1983). This insures that the following equality holds, $E(\Delta y_0^T|P(X)) = E(\Delta y_0^C|P(X))$. This means that we can rewrite the expected difference in changes in observed outcomes, conditional on the propensity score, as follows, using Assumptions (3.1) to (3.5):

$$\begin{aligned} & E(\Delta y^T|P(X)) - E(\Delta y^C|P(X)) \\ = \; & E(\Delta y_0^T|P(X)) - E(\Delta y_0^C|P(X)) + E(\beta^T D|P(X)) \\ = \; & E(\beta^T|D=1, P(X))E(D|Z=1, P(X)) \\ = \; & E(\beta^T|D=1, P(X))E(D|Z=1, P(X)). \end{aligned}$$

Therefore,

$$E(\beta^T|D=1, P(X)) = \frac{E(\Delta y^T|P(X)) - E(\Delta y^C|P(X))}{E(D|Z=1, P(X))}$$

Thus, to reduce the dimensionality, we can estimate the following object

$$\int_p \frac{[E(\Delta y^T|P(X)=p) - E(\Delta y^C|P(X)=p)]}{E(D|Z=1, P(X)=p)} dF(p|D=1)$$

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 $X(P(D=1|X))$, while using area type as instrumental variable we only need random area assignment, again conditional on X $(P(Z=1|X))$.

[10] Once again, we stress that, if participation in control areas is zero, the estimator for the ATT we propose in equation 7 is equivalent to a LATE estimator where a dummy indicating whether the household lives in a treatment or control area is used as instrument for participation. In reality, for unknown reasons, some (very few) households in control area seemed not be participating into the program. We do not consider this issue.

5 Results

To illustrate the different methodologies, in this section we report the results we obtain trying to estimate the effects of the program on consumption. As we discussed above, matching estimators play an important role in what follows. Therefore we start the section by presenting estimates of different propensity scores. We start by showing how the common support varies as we introduce different sets of variables. We start by discussing estimates of the probability of being in a treatment vs a control area that are used to compute AIT. In addition to the propensity score that measures the probability of being in a treatment area, we also present estimates of the propensity score that measures the probability of participating into the program in treatment areas. 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 way 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.

5.1 Propensity score and common support

Table 1 shows the frequencies of the propensity scores for area of residence obtained changing the set of conditioning variables. The first specification (s1) includes a large set of individual and area characteristics, including the proportion of poor (we will describe these variables in details below). As expected, there is no common support, as the concentration of poor is one of the criteria for the choice of treatment areas. In the second specification (s2) we drop all area-specific variables, with the exclusion of number of primary, middle, secondary schools, and health centers per households. Now there is complete common support. However, the right tail for the control group is very thin, so estimation in this part of the support will not be very precise.[11] Nevertheless, we believe it is important to condition on these aggregate variables because treatment areas are on average less poor than control areas, thus poorer areas may also have worse infrastructure. Lastly, the third specification (s3) uses only household demographics

[11] We also experimented with more parsimonious versions of s2, dropping the area-specific variables. In this case, the right tail for the control group is fatter (28% of $Z = 1$ households and 7% of $Z = 0$ households have a propensity score of 0.8 or higher. After dropping these aggregate variable, further omitting 1999 to 2001 earnings and employment variables, and 2002 negative shocks (loss of job, business, life, and natural disasters) does not change the distribution of the propensity score in a sizeable way.

and poverty index as covariates. Again, we have full support, and a higher density of the propensity score right tail for households in control areas. As expected, this last propensity score has a lower variance than the previous two. The right panel of the Table presents the results from a different propensity score. This time, we compute the probability of program participation for households in treatment areas, $P(D = 1|Z = 1, X)$, and predict the value of the propensity score for control households, i.e. $P(D = 1|Z = 0, X)$. In this case we have full common support with each set of covariates.[12]

Table 1: Frequencies of the propensity score (ps) for area of residence, Z=0 (control) and Z=1 (treatment), and for participation, compliers (D=1) and non-compliers (D=0) in treatment areas

	ps for area type: $P(Z = 1\|X)$						ps for area type: $P(D = 1\|Z = 1, X)$						participation: $P(D = 1\|Z = 1, X)$	
ps	s1		s2		s3		s1		s2		s3		s1	
	Z=1	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1	Z=0	D=1	D=0
$[0, 0.1)$	0.01	0.72	0.01	0.16			0.02	0.03	0.00	0.01			0.0	0.02
$[0.1, 0.2)$	0.01	0.09	0.01	0.08	0.00	0.00	0.06	0.12	0.03	0.05	0.01	0.00	0.02	0.10
$[0.2, 0.3)$	0.01	0.06	0.02	0.10	0.00	0.01	0.12	0.17	0.12	0.14	0.09	0.08	0.06	0.16
$[0.3, 0.4)$	0.01	0.04	0.03	0.12	0.01	0.03	0.19	0.21	0.23	0.24	0.27	0.24	0.13	0.22
$[0.4, 0.5)$	0.01	0.02	0.05	0.12	0.06	0.12	0.21	0.19	0.23	0.24	0.28	0.29	0.20	0.21
$[0.5, 0.6)$	0.02	0.02	0.08	0.12	0.27	0.35	0.18	0.13	0.20	0.18	0.20	0.22	0.20	0.14
$[0.6, 0.7)$	0.02	0.02	0.13	0.12	0.36	0.34	0.12	0.08	0.12	0.09	0.10	0.09	0.18	0.07
$[0.7, 0.8)$	0.03	0.01	0.19	0.10	0.24	0.13	0.07	0.04	0.05	0.04	0.04	0.05	0.12	0.03
$[0.8, 0.9)$	0.04	0.00	0.27	0.06	0.06	0.02	0.03	0.01	0.02	0.01	0.02	0.02	0.05	0.01
$[0.9, 1]$	0.84	0.01	0.23	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Propensity scores (ps) estimated by weighted probit. When we compute the second ps as $P(D = 1|Z = 1, X)$, we predict values of the ps for Z=0 households using the estimated coefficients. s1=all household- and area-level variables; s2=all household-level variables, number of primary, middle, secondary schools, and health centers per households, excluding all other area variables; s3=household demographics and poverty level. The frequencies may not add up to 100 because of rounding error.

The third panel of Table 1 shows the distribution of the propensity score computed from households in treatment areas using the first specification (s1). The value of the score are imputed using the coefficient values for households in control areas. The first two columns predict participation for control non-treated (D=0) and treated (D=1) households in treatment areas (Z=1). Recall that both groups of households are eligible for the program. The inspection of this Table reveals some interesting facts: first, the common support assumption holds for our sample, irrespective of whether we compare households within treatment areas, or whether we use information from households in control areas.[13] Second, the comparison of compliers

[12] Note that, when we use this propensity score to estimate the ATT using an instrumental variable approach, the possible omission of unobservables that affect both program participation and consumption change - such as transitory idiosyncratic shocks - does not pose a problem, as long as these variables are balanced between treatment and control areas (****Manu non capisco questa footnote sarei per toglierla.). Instead, their omission may violate the CIA when we estimate the ATT through propensity score matching.

[13] Obviously, in this case the common support assumption would also hold for more parsimonious specifications.

and non-compliers in treatment areas shows that there is a certain amount of self-selection into the program based on observable characteristics. However, it is not very strong: more than a quarter of non-compliers have a probability of participation of at least 50%. This fact may suggest that lack of information on the program existence, or on its features (eligibility rules and expected returns) may be partly responsible for the low enrolment rate. For example, suppose that only a random subset of eligible households is sufficiently informed about the program. Self-selection based on expected returns occurs only among the informed households, while others that would potentially benefit from the program are not aware of its existence. If this were true, we would expect participation rates to increase in the future, as knowledge of the program existence and characteristics diffuse among the eligible population. One the other hand, these results are also consistent with the possibility that households self-select into the program based on unobservables. Further research is needed to distinguish between these alternative hypotheses.

As mentioned above, a surprisingly small proportion of eligible households actually registered for the program. According to the tamizaje survey, among the 279 original treatment blocks included in our sample (this excludes Barrido blocks that do not appear in the tamizaje) the average proportion of eligible households that are registered for the program is 0.335. The block at the 75th percentile has half of eligible households participating, while the block at the 25th percentile has only eight per cent. 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 to predict participation into the program. In Table 2, we present weighted probit estimates of the propensity score coefficients. The left column models participation among eligible households in treatment areas, $P(D = 1|Z = 1, X)$, while the right column predicts the likelihood of living in a treatment area, $P(Z = 1|X)$. In both cases we condition on the set of variables from specification s2. These results are useful for three reasons: first, they are informative about the process that determines participation (left column). In particular, we can check the role played by several variables, ranging from individual variables (such as socio-economic indicators - assets, education of the household head, and so on) to environmental variables (features of the block and area such as distance from the Oportunidades enrolment module). Second, they show the characteristics by which households in treatment and control areas differ (right column). Lastly, a comparison of the two sets of estimates reveals whether matching on the predicted probability of participation balances all the variables by which T and C households differ, as we discussed in Section 3.2.

Two interesting facts emerge from looking at the determinants of program participation.

Table 2: Probit regression of determinants of program participation (left column) and of household characteristics by area of residence (right column) - marginal effects

Children tot. and in school:			**Employment and earnings:**		
nokids0-5	-0.012	0.032	hhinc02	-0.002	-0.003
	[0.012]	[0.009]***		[0.002]	[0.002]
nokids6-12	-0.029	-0.039	hhinc02^2	0	0
	[0.025]	[0.016]**		[0.000]	[0.000]
nokids13-15	-0.042	-0.018	Dhhworked01	0.071	0.035
	[0.025]*	[0.020]		[0.038]*	[0.033]
nokids16-20	-0.006	-0.013	Dhhworked00	-0.025	-0.018
	[0.016]	[0.013]		[0.038]	[0.032]
nosch0-5	0.067	-0.006	Dhhworked99	0.001	0.028
	[0.028]**	[0.022]		[0.034]	[0.029]
nosch6-12	0.082	0.06	hhinc01	-0.001	0.00
	[0.025]***	[0.016]***		[0.002]	[0.001
nosch13-15	0.1	0.04	hhinc00	-0.006	-0.00
	[0.029]***	[0.023]*		[0.004]	[0.003
nosch16-20	0.022	0.056	hhinc99	-0.001	-0.00
	[0.027]	[0.022]**		[0.002]	[0.002
Asset ownership:			saving02	0	0
Dcar	0.058	-0.133		[0.000]	[0.000]
	[0.090]	[0.069]*	debt02	0	0
Dtruck	-0.151	-0.057		[0.000]	[0.000]*
	[0.083]*	[0.069]			
Dfridge	-0.043	0.07	**Hoh and spouse:**		
	[0.021]**	[0.016]***	Female hoh	0.029	0.084
Dstove	0.049	-0.058		[0.031]	[0.026]***
	[0.023]**	[0.017]***	Dnopartner02	-0.096	0.029
Dwash	0.014	-0.09		[0.038]**	[0.031]
	[0.030]	[0.023]***	Hoh education (base=0):		
Schools/health centers per hh:			Dhohlit1	-0.054	0.049
Primary	18.704	110.218		[0.029]*	[0.026]*
	[7.912]**	[7.249]***	Dptrlit1	0.017	-0.109
Secondary	-180.943	-138.909		[0.031]	[0.027]***
	[30.569]***	[16.304]***	Dhoheduc1	0.053	-0.054
High school	177.831	97.866		[0.030]*	[0.027]**
	[32.982]***	[20.302]***	Dhoheduc2	0.021	-0.151
Health center	21.917	-435.04		[0.035]	[0.031]***
	[30.317]	[19.882]***	Dhoheduc3	0.041	-0.149
Housing:				[0.045]	[0.040]***
Dirt floor	0.024	0.171	Dhoheduc4	0.028	-0.163
	[0.019]	[0.015]***		[0.039]	[0.033]***
cardboard, tyres roof	0.023	-0.139	Dhoheduc5	0	-0.071
	[0.021]	[0.016]***		[0.047]	[0.042]*
cardboard, tyres wall	-0.027	0.118	Dhoheduc9	-0.06	-0.062
	[0.027]	[0.021]***		[0.050]	[0.043]
rooms number	-0.047	-0.06	Dptreduc1	-0.059	0.131
	[0.011]***	[0.009]***		[0.034]*	[0.027]***
piped water	0.003	0.026	Dptreduc2	-0.085	0.056
	[0.019]	[0.015]*		[0.037]**	[0.031]*
Toilet	-0.004	-0.114	Dptreduc3	-0.099	0.101
	[0.017]	[0.013]***		[0.049]**	[0.038]***
Own home	0.062	-0.022	Dptreduc4	-0.1	0.071
	[0.018]***	[0.014]		[0.041]**	[0.033]**
			Dptreduc5	-0.083	0.128
				[0.055]	[0.040]***
Continues					

Continued					
Welfare receipt:			Dptreduc9	0.063	-0.029
Dtortilla02	0.065	0.175		[0.066]	[0.049]
	[0.027]**	[0.020]***	**Transitory shocks:**		
Dmilk02	0.068	-0.134	Death	0	0.028
	[0.034]**	[0.024]***		[0.024]	[0.019]
Ddif02	0.133	0.161	Unemployed Hoh	0.019	0.109
	[0.033]***	[0.025]***		[0.020]	[0.016]***
Dbrek02	-0.002	0.007	Lost business	0.165	-0.14
	[0.031]	[0.022]		[0.120]	[0.103]
Deducs02	0.097	-0.106	Natural disaster	0.04	0.086
	[0.056]*	[0.043]**		[0.037]	[0.032]***
Dtrans02	-0.052	0.055	Hoh doctor visit	not signif.	not signif.
	[0.222]	[0.191]	Spouse to doctor	not signif.	not signif.
Dprobecat02	0.311	0.172	Children to doctor	not signif.	not signif.
	[0.180]*	[0.158]	**Household poverty index:**		
Dhouse02	0.018	-0.03	Index	0.129	-0.093
	[0.143]	[0.100]		[0.021]***	[0.016]***
Dprocampo02	-0.203	0.181	**Observations**		
	[0.082]**	[0.066]***		5318	8413

Robust standard errors in brackets; clustering at the locality level. Note: hh=household; D=dummy (e.g. DhhworkedXX=dummy for whether household head was employed in year XX). Unless otherwise specified, all variables are from 2002. * means significant at 10%, ** significant at 5%, *** significant at 1%.

The first is that individuals appear to 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, fewer assets and rooms per person, receipt of other government programs, and large number of children in school are more likely to be Oportunidades recipients. In particular, each extra child with some schooling increases the enrolment likelihood by 6.7, 8.2 and 10 percentage points for children aged 0 to 5, 6 to 12 and 13 to 15, respectively. Participation is also higher among households with illiterate (5 percentage points) or low educated heads and spouses. A lower cost of participation (proxied by the availability of schools) is also associated with higher participation: a 10 percent increase in primary and high school per hundred households around the means (which are 0.0021 and 0.0004) increases the participation probability by 0.4 and 0.8 percentage points, respectively.

The second interesting conclusion from Table 2 is that participation seems to be associated with measures of permanent, rather than transitory poverty. Indeed, most of the variables that are significantly associated with program enrolment, such as number of rooms and parental education, are proxies of permanent poverty levels. More temporary income measures, such as contemporary and previous income and household head employment status are not significantly related to program participation. Moreover, transitory idiosyncratic shocks such as job or business loss, illness or death of relative and natural disasters do not seem to occur more frequently among participants. This point is especially important given that one key identification assumption needed when estimating ATT effects through matching is that after conditioning on observables, participation is not caused by unobservable transitory shocks. The lack of association between the aforementioned transitory variables and enrolment seems to suggest that participation may not be driven by time-varying unobservables.

The signs of two coefficients are not consistent with the general patter of correlations described: middle school availability and lack of home ownership appear to be associated with lower program participation. Given the above discussion, one would expect the opposite. An explanation for the latter case is readily available: poverty is related to participation in an inverse u-shaped fashion, although it is positive for most households. Higher poverty levels are associated with a higher likelihood of enrolment up to a level of about 3 (the aggregate poverty measure used varies between .7 and 6 for poor households), and with a negative participation likelihood after that threshold. Only about 5 percent of the households have a poverty level larger than 3. Thus households who do not own a house may be the poorest.[14]

Lastly, we proceed to compare the estimates from the two propensity scores. We already explained how 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. Its partial effect as a determinant of program participation is small, 0.019, and not significant. However, household heads in treatment areas are 11 percentage points more likely not to have a job than household heads in control areas. When we match households based on the first 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. Unemployment is not the only variable that is balanced within treatment areas, but unbalanced between treatment and control areas: other such variables are housing characteristics, asset ownership, and education of household heads. By and large, the analysis of the respective partial effects reveals that using the first propensity score, treated households are matched to wealthier control households. For instance, households in treatment areas are 13 percentage points less likely to own a car, and 17 and 12 percentage points more likely to live in a house with dirt floor and with walls made of cardboard or tyres. The household head is much more likely to be uneducated, and the spouse is 12 percentage points more likely to be illiterate. All these variables do not seem to affect participation. If the change in consumption in the absence of the program is positively correlated with household wealth, the consumption

[14] The results from this additional regression are not reported, because the general pattern of sign and significance of the remaining variables does not change. Further investigation is required to understand the sign of the secondary school coefficient.

change for households in control areas is higher than the change for treated households in the absence of the program, i.e. $E(\Delta Y|Z=0) > E(\Delta y_0|D=1, Z=1)$, thus matching using the imputed propensity score underestimates the ATT.

5.2 Estimates of AIT and ATT effects

Table 3: Average Intention to Treat and IV Average Treatment on the Treated estimates for consumption, logs and levels

	IV with ps	OLS-IV	Matching	Matching
ps:	$P(Z=1\|X)$		$P(D=1\|Z=1,X)$	$P(D=1\|X)$
	(1)	(2)	(3)	(4)
		logs		
AIT	0.118	0.037	—	—
	[0.034]***	[0.041]		
ATT	0.252	0.139	0.057	0.080
	[0.074]***	[0.072]*	[0.018]***	[0.044]*
		levels		
AIT	239.72	70.44	—	—
	[70.20]***	[71.30]		
ATT	507.36	213.30	85.68	141.25
	[153.85]***	[145.49]	[32.39]***	[79.75]*

Note that the proportion incorporated is 0.475. Standard errors [in brackets] and 95% confidence interval (in parentheses) clustered at the area level. Block-bootstrap standard errors for ATT estimates, the block is the area. The matching estimates in the first two columns use all household and are variables to build the propensity score. These same variables are added as controls to the OLS-IV specifications in the last column. Instead, we use specification s2 (household variables and school and health care availability) to compute the propensity score in the IV estimates from column 3. Local linear regression matching estimates. The estimates from llr are similar to the ones obtained using the 5 nearest neighbors with replacement.

Table 3 provides estimates of the average Intent to Treat and Treatment on the Treated effects on consumption level and log, using different estimators. The first column shows the estimates from our favorite estimator, which uses equation (5) to estimate the AIT, and (**??**) to estimate the ATT. The second column estimates the AIT by OLS, and then divides this estimate by average program participation. The third and fourth columns provide propensity score matching estimates of the ATT. In the third column we are comparing compliers and non-compliers in treatment areas, i.e. the propensity score we estimate is the probability of participation for eligible households in $Z=1$ areas, $P(D=1|Z=1,X)$. In the fourth column, instead, we use the estimated coefficients from this propensity score to impute the probability of participation for poor households in control areas. We then compare outcomes for compliers and households in control blocks with the same level of the propensity score. In both cases, we present local linear regression estimates.[15] Note that, while the last three sets of estimates

[15] We also experimented using the 5 nearest neighbor with replacement, and the results are almost identical.

use all household and all area variables as controls, column one uses specification s2, i.e. all household variables and school and health center availability.

The estimated AIT from column one reveals that consumption is 12%, or 240 *pesos* larger for eligible households in treatment areas, compared to households with similar observable characteristics who live in control areas. This parameter is a lower bound of the ATT under fairly general assumptions: as long as there are no program spillover effects on both non-compliers and households in control areas, the parameter is an estimate of (1). However, it remains a lower bound of the ATT even in case of spillover effects for non-compliers (i.e. eligible non-participants in treatment areas), as long as they are smaller than the program effect on the treated.

The estimates based on the two matching estimators discussed above are considerably smaller. Given that the average grant received by these housholds is around 450 pesos, we feel that the effects one obtains by matching are too small to be pplausible. An increase in consumption by 141 pesos would imply an astonishing saving rate among these poor households. The comparison between the two propensity scores in Table 2 is suggestive about the reason for a possible negative bias of the matching estimator.

6 Conclusions

In this note we have considered different strategies for the estimation of the effects of a large welfare program in Mexico. The evaluation has to tackle several methodological problems. The first is the non random assignment of the program across areas. This is dealt, in a fairly standard fashion, combining difference in difference and matching methods. The second problem arises from the remarkably low participation into the program in treatment areas. WE discuss two alternative ways to deal with this program. The two approaches are valid under different set of assumptions. Our empirical example, based on the effect of the transfer on household consumption, seems to indicate that the second approach, based on inflating Intent to Treat estimates using an Instrumental Variables approach is more likely to deliver plausible answers than the approach based on matching.

7 References

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.

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.

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.

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.