

Income Gains to the Poor from Workfare: Estimates for Argentina's Trabajar Program

Jyotsna Jalan and Martin Ravallion¹

Indian Statistical Institute and World Bank

¹ The work reported in this paper is one element of the ex-post evaluation of the World Bank's Social Protection II Project in Argentina. The support of the Bank's Research Committee (under RPO 681-39) is gratefully acknowledged. The paper draws on data provided by the SIEMPRO unit of the Ministry of Social Development, Government of Argentina. The authors are especially grateful to Joon Hee Bang and Liliana Danilovich of SIEMPRO for their help with the data. The authors' thanks also go to staff of the Trabajar project office in the Ministry of Labor, Government of Argentina who provided the necessary data on their program and gave this evaluation their full support. Petra Todd kindly advised us on matching methods. Useful comments were received from Polly Jones, Dominique van de Walle, and seminar participants at the World Bank, the Indian Statistical Institute, Delhi, and the Institute of Fiscal Studies, London.

1. Introduction

Workfare programs require that participants must work to obtain benefits.² They are often turned to in crises such as due to macroeconomic or agro-climatic shocks, in which a large number of poor able-bodied people have become unemployed. Typically, the main aim of workfare is to raise the current incomes of poor families hurt by the crisis.

To assess the impact of such a program, we need to measure the income gain conditional on income in the absence of the program. The income gain is the difference between household income with the program and that without it. The “with” data can be collected without great difficulty. But the “without” data are fundamentally unobserved, since an individual cannot be both a participant and a non-participant of the same program. Common practice has been to estimate the gains by the gross wages paid.³ In other words, the unobserved income without the program is taken to be equal to income with the program, minus wages received.

This assumption would be a reasonable one if labor supply to a workfare program came only from the unemployed. But that is difficult to accept. Even if a participating worker was unemployed at the time she joined the program, that does not mean that she would have remained unemployed had the program not existed. Even a worker who has been unemployed for some time will typically face a positive probability of finding extra work during a period of search, including self-employment in an informal sector activity. Joining the program will leave less time for search. There are also ways in which behavioral responses help reduce foregone income. There are likely to be effects on time allocation within the household. For example,

² On the arguments and evidence on this class of interventions see Ravallion (1991, 1999a), Besley and Coate (1992), Lipton and Ravallion (1995), Mukherjee (1997), and Subbarao (1997).

³ See, for example, the various assessments of the cost-effectiveness of workfare programs reviewed in Subbarao et al., (1997).

Datt and Ravallion (1994) find that other family members took up the displaced productive activities when someone joined a workfare program in rural India. Such behavioral responses will reduce foregone income, though we can still expect it to be positive.

This paper estimates the income gains from a workfare program and how those gains vary with pre-intervention incomes. We study the Trabajar Program instituted by the Government of Argentina, and supported by a World Bank loan and technical assistance. We use propensity-score matching methods (Rosenbaum and Rubin, 1983, 1985; Heckman et al., 1997, 1998) to draw a comparison group to workfare participants from a larger survey of non-participants. A number of features of this setting lend themselves to matching methods. It was possible to assure that the same questionnaire was administered to both the participants and the non-participants, and that both groups were from the same economic environment. The Trabajar participant could be identified in the larger survey.⁴ Furthermore, using kernel density estimation techniques, we are able to ensure that participants are matched with the non-participants over a common region of the matching variables. Any remaining bias in the matching estimator can thus be attributed to unobserved characteristics. The design of the program can be expected to entail considerable rationing of participation according to observables; the sample of non-participants is very likely to include people who wanted to participate but were unable to do so due to say non-availability of the program. While our application is well suited to matching methods, bias due to unobservables cannot be ruled out by matching alone, since the method is based solely on observables. So we also propose and implement a test for any remaining selectivity bias after matching.

⁴ The researcher may not be able to identify whether an individual participated in the program or not in the larger population sample. In such cases, one can still go ahead with the matching procedure though this adds a “contamination bias” to the impact estimator. In our application this is not an issue.

The following section discusses the evaluation problem and our methods. Section 3 describes the Trabajar program. Our data are described in Section 4. Section 5 presents the results, and offers an economic interpretation. Section 6 concludes.

2. Estimating the Income Gains from Workfare

In assessing the gains from a workfare program, the workers' earnings are naturally the main focus, and that will be the case here. However, it should be noted that earnings net of foregone income are only one of the potential benefits. There could also be risk benefits from knowing that the program exists. There may well also be benefits from the outputs, depending on (amongst other things) how well targeted the workfare projects are to poor areas.⁵

We first outline what we see as the model of self-targeting underlying arguments for workfare, pointing to the key role played by foregone incomes. We then describe the matching method we use to estimate foregone incomes.

2.1 The Problem

The following rudimentary model has the essential features necessary to characterize the “self-targeting” argument often made in favor of workfare (Ravallion, 1991). The model assumes that foregone income from accepting a workfare job is $F(Y)$, a smoothly increasing function of pre-intervention income Y (scaled to lie between zero and one). Foregone income increases with pre-intervention income due to differences in education, experience and so on that are naturally correlated with both earnings and family income. The workfare program offers a wage W , with $F(0) < W < F(1)$. Workers only care about the net wage gain (i.e., the work alternatives are judged to be the same in other respects).

⁵ This issue is examined further in Ravallion (1999b), which presents results on poor-area targeting for the same program studied here.

It is evident that under these assumptions, only those workers with pre-intervention income less than $F^{-1}(W)$ will participate; the program will perfectly screen “poor” ($Y < F^{-1}(W)$) from “non-poor” ($Y > F^{-1}(W)$). The schedule of gains is $G = W - F(Y)$ for $Y < F^{-1}(W)$ and $G = 0$ for $Y > F^{-1}(W)$, yielding post-intervention incomes $Y + G$.

In this simple model, underestimating the foregone income will lead the evaluator to overestimate the impact on poverty. To see why, suppose that, in assessing the gains from the program, we use a biased estimate of foregone income, namely $M(Y) < F(Y)$ for all Y . Then we will overestimate the gains for all Y up to $M^{-1}(W)$. The distribution of incomes under the biased estimate of foregone incomes must first-order dominate the actual distribution. So the error in assessing foregone incomes will overestimate the impact on income poverty.⁶

This model also suggests that in the extreme—though commonly assumed—case in which the foregone income is zero, a workfare program would make little sense as a means of reaching the poor. There will be no self-targeting mechanism, and the government would have to rely on some form of indicator targeting or means test. So using the program wage to measure the income gain is antithetical to the logic of a workfare program as a means of self-targeting.

How can one estimate the foregone income? This is a counterfactual concept in that participants’ incomes in the absence of the program cannot be data. There are several methods one might adopt to assess the counter-factual, drawing on the literature on impact evaluation. One can do reflexive comparisons by collecting baseline data on probable (eligible) participants before the program was instituted. These data are then compared with data on the same individuals once they have actually participated in the program. In this case, the counterfactual group is the set of participating individuals themselves, but observed before the program is

⁶ This holds for a broad-class of poverty measures (Atkinson, 1987).

actually implemented. This method can be extended to include observations on non-participants, before and after the intervention, allowing a “double-difference” estimate of the program’s impact. Alternatively, potential participants are identified and data are collected from them. However, only a random sub-sample of these individuals is actually allowed to participate in the program. The identified participants who do not actually participate in the program (the “randomized out” group) form the counterfactual in this case.

Another possible approach is to use propensity-score matching methods, following Rosenbaum and Rubin (1983, 1985) and Heckman et al. (1997, 1998). Here, the counterfactual group is constructed by matching program participants to non-participants from a larger survey such as the population census or an annual national budget survey. The matches are chosen on the basis of similarities in observed characteristics.

Each of these methods has both strengths and weaknesses. For example, reflexive and double-difference comparisons raise concerns about attrition, whereby a non-random subset of the baseline sample drops out for various reasons. Randomization is ideal in theory, since the comparison group has the same expected distribution of characteristics as the treatment group in the absence of the intervention. However, randomization is not often feasible, and there can also be problems of selective non-participation amongst those randomly chosen for the program. Both baseline survey methods and randomization also require that the evaluation is set up prior to the program. This is unlikely to be feasible in crisis situations. A government concerned about the social impact of a macroeconomic or agro-climatic crisis is not likely to agree to wait for the evaluation to be put in place.

Matching methods can avoid these problems, though they create their own. An advantage is that, since most countries now have a nationally representative socio-economic survey instrument, the marginal cost of using matching methods only includes the survey of program

participants. The same survey instrument can then be taken to a sample of participants after the program has started, possibly with an extra module to cover specific questions related to the program. Matching estimates will be reliable if: (i) participants and controls have the same distribution of unobserved characteristics; failure of this condition to hold is often referred to as the “selection” problem in econometrics; (ii) they have the same distribution of observed characteristics; (iii) the same questionnaire is administered to both groups; and (iv) participants and controls are from the same economic environment. In the absence of these features, the difference between the mean earnings of the participants in a social program and the matched non-participants will be a biased estimate of the mean impact of the program.

2.2 *Method of Estimating the Gains from Workfare*

Suppose we have data on N participants in a workfare program, and another random sample of size rN ($r > 1$) from the population. The second set of data might be the national population census or an annual national household budget survey that has information relevant in the participation decisions of the individuals. Using the two sets of data, we try to match the N program participants with a comparison group of non-participants from the population.

The two surveys must include information that helps predict participation in the program. Let X be the vector of such variables. Ideally, one would match a participant with a non-participant using the entire dimension of X , i.e., a match is only declared if there are two individuals, one in each of the two samples, for whom the value of X is identical. This is impractical, however, because the dimension of X could be very high. Rosenbaum and Rubin (1983) show that matching can be performed conditioning on $P(X)$ alone rather than on X , where $P(X) = \text{Prob}(D=1 | X)$ is the probability of participating conditional on X , the “propensity score” of X . If outcomes without the intervention are independent of participation given X then they are

also independent of participation given $P(X)$. This is a powerful result, since it reduces a potentially high-dimensional matching problem to a single dimensional problem.

The propensity score is calculated for each observation in the participant and the comparison-group samples using standard logit models.⁷ Choice-based sampling methods suggested by Manski and Lerman (1978) can be used to weight the observations given that there is over-sampling of participants. In our case however, we do not know the sampling weights to do the choice-based sample re-weighting. But we can still carry out the matching using the odds ratio $p_i = P_i/(1-P_i)$ where P_i is the estimated probability of participation for individual i . Using the propensity score, one constructs matched-pairs on the basis of how close the scores are across the two samples. The nearest neighbor to the i 'th participant is defined as the non-participant that minimizes $[p(X_i) - p(X_j)]^2$ over all j in the set of non-participants, where $p(X_k)$ is the predicted odds ratio for observation k .

In their comparisons of non-experimental methods of evaluating a training program with a benchmark experimental design, Heckman et. al (1997, 1998) find that failure to compare participants and controls at common values of matching variables is the single most important source of bias — considerably more important than the classic econometric problem of selection bias due to differences in unobservables. To ensure that we are matching only over common values of the propensity scores, we estimated the density of the scores for the non-participants at 100 points over the range of scores. We use a biweight kernel density estimator and the optimal bandwidth value suggested by Silverman (1986). Once we estimate the density for the non-

⁷ One could use semi- and non-parametric methods to estimate the propensity scores though Todd (1995) argues that such methods do not make any difference to the impact estimator. Thus for computational simplicity, we use standard parametric likelihood methods to compute the estimated propensity scores.

participants, we exclude those non-participants for whom the estimated density is equal to zero. We also exclude 2% of the sample from the top and bottom of the non-participant distribution.

The mean impact estimator of the program is given by:

$$\bar{G} = \sum_{j=1}^P (Y_{j1} - \sum_{i=1}^{NP} W_{ij} Y_{ij0}) / P \quad (1)$$

where Y_{j1} is the post-intervention household income of participant j , Y_{ij0} is the household income of the i^{th} non-participant matched to the j^{th} participant, P is the total number of participants, NP the total number of non-participants and the W_{ij} 's are the weights applied in calculating the average income of the matched non-participants. There are several different types of parametric and non-parametric weights that one can use. In this paper we use three different weights and thereby report three different matching estimators. Our first matching estimator is the “nearest neighbor” estimator where we find the closest non-participant match for each participant and the impact estimator is a simple mean over the income difference between the participant and its matched non-participant.^{8,9} Our second estimator takes the average income of the closest five matched non-participants and compares this to the participant's income. We also report a kernel-weighted estimator where the weight are given by:

$$W_{ij} = K_{ij} / \sum_{j=1}^P K_{ij} \quad (2)$$

where

$$K_{ij} = \frac{K[(P(X_i) - P(X_j)) / a_{N0}]}{\sum_{j=1}^P K[(P(X_i) - P(X_j)) / a_{N0}]} \quad (3)$$

⁸ The closest match is chosen by the distance metric discussed above. Also we allow for replacement of the non-participants, so a non-participant could be the closest match for more than one participant.

and where a_{NO} is the bandwidth parameter, $K(\cdot)$ is the kernel as a function of the difference in the propensity scores of the participants and the non-participants. In our analysis, we have used Silverman's (1986) optimal bandwidth parameter and a biweight kernel function. (The results were very similar using either a rectangular or parzen kernel function.)

Lastly, in each of these cases, the associated standard errors of the mean impact estimator are also calculated. We calculated both the parametric and bootstrapped standard errors for the impact estimators. The two were virtually identical. We report the parametric standard errors in the paper. (The bootstrapped standard errors are available from the authors on request.)

2.3 *Testing for Bias due to Unobservables*

The matching estimate described above will give a biased estimate of the income gains from workfare if there are unobserved variables that jointly influence incomes and workfare participation, conditional on the observed variables in the data used for matching.

A natural test for such a bias is look for a partial correlation between incomes and the residuals from the participation model (used to construct the propensity scores) controlling for actual participation. We call this the test for selection bias in the matching estimator. It is a straightforward application of the standard Sargan-Wu-Hausman test. There will, of course, also be heterogeneity in other characteristics relevant to incomes. By performing the test for selection bias on a sample combining the participants and their matched non-participants we will have already eliminated some of this heterogeneity. One can also explicitly introduce a vector of control variables (Z) to give a test equation for income Y of the form:

$$Y_i = \mathbf{a} + \mathbf{b}P_i + \mathbf{g}R_i + \mathbf{d}Z_i + \epsilon_i \quad (4)$$

⁹ In calculating our mean impact, if the income of the participant is less than the income of the matched non-participant, we treat the impact to be zero rather than the observed negative number.

for household i in the combined participant and matched control sample of nearest neighbors, where R_i denotes the residuals from the participation model. Selection bias is indicated if we can reject the null that $\gamma=0$. In a linear model, identification requires the usual condition that there is at least one variable in X that is not in Z . The non-linearity of the propensity scores in X means that this condition is not essential. However, specifics of the program's design (discussed below) will provide a seemingly plausible exclusion restriction allowing identification without relying solely on the non-linearity.

We can use this method to test for selection bias only in the nearest neighbor case where there is one matched non-participant for each participant. In the other two cases, it does not appear to be possible to use a regression to replicate the complex weighting of the data on non-participants used in forming the matching estimator.

3. The Trabajar Program

Argentina experienced a sharp increase in unemployment in mid 1990s, reaching 18% in 1996/97. This was clearly hurting the poor; for example, the unemployment rate (on a comparable basis) was 39% amongst the poorest decile in terms of household income per capita in Greater Buenos Aires. Unemployment rates fell steadily as income per person increases.¹⁰

With financial and technical support from the World Bank, the Government of Argentina introduced the Trabajar II program in May 1997. This is a greatly expanded and reformed version of a previous program, Trabajar I. The program aimed to help in two ways. Firstly, by providing short-term work at relatively low wages, the program aimed to self-select unemployed workers from poor families. Secondly, the scheme tried to locate socially useful projects in poor

¹⁰ These data are from the Permanent Household Survey (EPH) for Greater Buenos Aires in May 1996.

areas to help repair and develop local infrastructure. This paper only assesses progress against the first objective (on the second see Ravallion, 1999b).

The national program budget is allocated across provinces by the center, leaving the provincial governments with considerable power to determine how the moneys are allocated within the province. There is evidence of horizontal inequality in the outcomes of this process, in that equally poor local areas (“departments”) obtained very different allocations in expectation from the program depending on which province they belong (Ravallion, 1999b). This decentralized nature of the program is the basis of our identification strategy in testing for selectivity bias. Following Ravallion and Wodon (1998) we use province dummy variables as the instruments. Clearly the province of residence matters to participation given the way program funds are allocated. We then assume that province of residence does not matter to incomes independently of participation. Given that we will include a wide range of local geographic control variables in the income regression this assumption is defensible.

The projects are proposed by local governmental and non-governmental organizations who must cover the non-wage costs. The projects have to be viable by a range of criteria, and are given priority according to ex ante assessments of how well targeted they are to poor areas, what benefits they are likely to bring to the local community, and how much the area has already received from the program. Workers cannot join the program unless they are recruited to a project proposal that is accepted on the basis of these criteria. The process of proposing suitable sub-projects is thus key to worker participation in the program. There are other factors. The workers cannot be receiving unemployment benefits or be participating in any other employment or training program. It is unlikely that a temporary employment program such as this would affect residential location, though workers can commute.

The wage rate is set at a maximum of \$200 per month. This was chosen to be low enough to assure good targeting performance, and to help assure workers would take up regular work when it became available. To help locate the Trabajar wage in the overall distribution of wages we examined earnings of the poorest 10% of households (ranked by total income per person) in Greater Buenos Aires (GBA) in the May 1996 Permanent Household Survey. For this group, the average monthly earnings for the principal job (when this entails at least 35 hours of work per week) in May 1996 was \$263.¹¹ (As expected, the poorest decile also received the lowest average wage, and average wages rose monotonically with household income per person.) So the Trabajar wage is clearly at the low end of the earnings distribution.

There are two further concerns about the project that the evaluation can throw light on. One concern has been the low level of female participation; only 15% of participating workers in the first six months were female. The key question is why. If it is because women choose not to participate then one would be less concerned than if it arose from impediments to their participation due to discrimination in allocating Trabajar jobs. If there is such a gender bias then there will be unexploited welfare gains from higher female participation. We cannot measure the welfare gain, but we can determine whether the net income gain is higher for women than men, implying an income loss from low female participation.

Secondly, while Trabajar I had been targeted to middle aged heads of households, it was decided not to impose this restriction on the new program since it risked adding to the forgone income of participants by constraining their ability to adjust work allocation within the household in response to the program. However, the past practice under Trabajar I may still

¹¹ This includes domestic servants. This is an unusual labor-market group, given that they often have extra income-in-kind. If one excludes them, the figure is \$334.

have influenced local implementation of the new program. Then one might expect to find that there are unexploited income gains by increasing participation by the young. We will test this.

4. Data

Two household surveys are used. One is of program participants and the other is a national sample survey, used to obtain the comparison group. Both surveys were done by the government's statistics office, the Instituto Nacional De Estadística Y Censos (INDEC), using the same questionnaire, the same interviewing teams and at approximately the same time.

The national survey is the Encuesta de Desarrollo Social (EDS), a large socio-economic survey done in mid-1997. The EDS sample covers the population residing in localities with 5,000 or more residents. The comparison group is constructed from the EDS. According to the 1991 census, such localities totaled to 420 in Argentina and represented 96% of the urban population and 84% of the total population. 114 localities were sampled.

The second data set is a special purpose sample of Trabajar participants done for the purpose of this evaluation. The sample design involved a number of steps. First among all the projects approved between April and June of 1997, 300 projects in localities which were in the EDS sample frame were randomly selected, with an additional 50 projects chosen for replacement purposes. The administrative records on project participants did not include addresses, so Ministry of Labor (MOL) had to obtain these by field work. From these 350 projects, the Labor Ministry could find the addresses of nearly 4,500 participants. However, for various reasons about 1,000 of these were not interviewed. The reasons given by INDEC were that the addresses were found to be outside the sample frame, or they were incomplete, or even non-existent, or that all household members were absent when the interviewer went to interview

the household, or that they did not want to respond. In all 3,500 participant households were surveyed. (The number of Trabajar participants during May 1997-January 1998 was 65,321.)

We restrict the analysis to households with complete income information, and those who completed all the questions asked of them. Also, we only consider participants who were actually working in a Trabajar project at the time they were surveyed. Since the EDS questionnaire does not ask income-related questions to those below 15 or above 64 years of age, we also had to restrict our attention to the age group 15-64 years for our analysis. We focus on current Trabajar participants in the reference week, fixed at the first week of September 1997, who received wages from the Trabajar program during August 1997. 80% of the Trabajar sample had current participants by this definition.¹² With these restrictions, the total number of active participants that we have used is 2,802.

5. Results and Interpretation

5.1 Descriptive Statistics

In Table 1, we present selected descriptive statistics for the Trabajar and EDS samples. The Trabajar sample has lower average income, higher average family size, are more likely to have borrowed to meet their basic needs, receive less from informal sources, are more likely to participate in some form of political organization, and less likely to own various consumer durables (with the exception of a color TV, which appears to be a necessity of life in Argentina.)

Table 2 gives the percentage distribution of Trabajar participants' families across deciles formed from the EDS with households ranked by income per capita, excluding income from Trabajar. (The poorest decile is split in half.) This is the type of tabulation that is typically made

¹² The remaining 20% of the participants are assumed to be beneficiaries of the program that had left work by August 1st, 1997 (i.e. at the start of the survey), or had not yet started the Trabajar job.

in assessing such a program. It assumes zero foregone income, so each participating family's pre-intervention income is simply actual income minus wage earnings from the program.

Table 2 suggests that a high proportion of the families of participants come from poor families.¹³ 40% of the program participants have a household income per capita which puts them in the poorest 5% of the national population; 60% of participants are drawn from the poorest 10% nationally. By most methods of measuring poverty in Argentina, the poverty rate is about 20%. So 75-85% of the participants are poor by this standard. Such targeting performance is very good by international standards.

Does relaxing the assumption of zero foregone income change the results in Table 1? Using the matching methods described above, we will now see whether that assumption is justified, and how much it matters to an assessment of average gains and their incidence.

5.2 *Propensity-Score Matching Estimates*

Table 3 presents the logit regression used to estimate the propensity scores on the basis of which the matching is subsequently done. The results accord well with expectations from the simple averages in Table 1. Trabajar participants are clearly poorer, as indicated by their housing, neighborhood, schooling, and their subjective perceptions of welfare and expected future prospects (relative to their parents). The participation regression suggests that program participants are more likely to be males who are head of households and married. Participants are likely to be longer-term residents of the locality rather than migrants from other areas. The model also predicts that (controlling for other characteristics) Trabajar participants are more

¹³ We have only considered participants who have earned at least \$150 from participating in the Trabajar job while calculating the impact of the project. The minimum wage offered under Trabajar is \$200; those reporting less than \$150 as their Trabajar earnings imply that these participants are in their last phase of their Trabajar job and or have misreported their income. Since we are interested in the impact on currently active participants in the program, we excluded the observations to get a better albeit a more conservative measure of the impact.

likely to be members of political parties and neighborhood associations. This is not surprising given the design of the program, since social and political connections will no doubt influence the likelihood of being recruited into a successful sub-project proposal. However, participation rates in political parties and local groups are still low, even for Trabajar participants (Table 1).

Based on Table 3, the mean propensity score for the national sample is 0.075 (with a standard deviation of 0.125). This is of course much lower than the mean score for the Trabajar sample, which is of 0.405 (0.266). However, after following the matching method outlined in section 2.2, the comparison group of nearest neighbors drawn from the national sample has a mean score of 0.394 (0.253), very close to that of the Trabajar sample.

Tables 4 and 5 give our estimates of average income gains, and their incidence according to pre-intervention incomes. For this purpose we have first estimated the gain for each participating household, by each of the three methods described in section 2.2. We then assign each household to a decile using the same decile bounds calibrated from the EDS, but this time the participants are assigned to the decile implied by their estimated pre-intervention income as given by actual income minus the estimated net gain.

The nearest neighbor estimate of the average gain is \$157, about three-quarters of the Trabajar wage. The nearest five and non-parametric estimator give appreciably lower gains, of around \$100. Since the latter estimates use more information and are presumably more robust we will use them in preference to the nearest neighbor estimate. For computational convenience and to circumvent the small sample problem in the sub-group cases, the rest of the paper is based on the “nearest five” rather than the non-parametric estimate.

The average gain using the “nearest five” estimator of \$103 is about half of the average Trabajar wage. Given that there is sizable foregone income, the crude incidence numbers in Table 1 overestimate how pro-poor the program is, since pre-intervention income is lower than is

implied by the net gains. Where this bias is most notable is amongst the poorest 5%; while the non-behavioral incidence analysis suggests that 40% of participant households are in the poorest 5%, the estimate factoring in foregone incomes is much lower at 10%. Nonetheless, over half of the participant households are in the poorest decile nationally even allowing for foregone incomes. Given that the poverty rate in Argentina is widely reckoned to be 20%, our results suggest that four out of five Trabajar participants are poor by Argentinean standards.

Figure 1 gives mean income gain at each level of pre-intervention income, estimated by a locally-weighted smoothed scatter plot of the data. Gains fall sharply (though not continuously) up to an income of about \$200 per person per month (which is about the median of the national distribution), and are roughly constant after that. We will return later to interpret this finding.

Figure 2 gives the empirical cumulative distribution functions (CDFs) implied by our results. We give the CDFs for both the Trabajar participants and the national distribution. We also give the counter-factual (pre-intervention) CDF for the Trabajar participants. There is a spike of zero incomes in the national sample, much of which is probably measurement error. If one takes this spike out there is first-order dominance comparing the Trabajar samples and the national samples, with higher poverty in the Trabajar sample at all possible poverty lines. There is automatically first-order dominance of the post-intervention incomes for the Trabajar sample given that we have ruled out negative gains on a priori grounds.

The absolute gains are highest for the third decile, but do not vary greatly across the deciles containing participants. The percentage net gains are highest for the poorest, reaching 74% for the poorest 5%. In section 5.3 we will offer an interpretation of these findings.

Tables 7-9 report the net wage gains for three different demographic groups: female participants, participants between the ages 15-24 years (typically identified as those who are new entrants into the job market), and workers in the age group 25-64 years.

The estimates in Table 7 are not consistent with the existence of income losses due to a gender bias in the program. The net wage gains from the program accruing to female participants are virtually identical to the gains for male participants. However, the distribution of female participation is less pro-poor, as indicated by household income per capita; while over half of the members of participating families are in the poorest decile nationally, this is true of less than 40% of the members of female participants' families. This probably reflects lower wages for women in other work, making the Trabajar wage more attractive to the non-poor.

For the younger cohort however, the net gains are significantly higher (comparing Tables 8 and 9). Foregone incomes are lower for the young, probably reflecting their lack of experience in the labor market. Because of this, there would be income gains from higher participation by the young. (To the extent that any young participants leave school to join the program, future incomes may suffer.) This suggests that the older workers may well be favored in rationing Trabajar jobs. However, the distribution of gains is more pro-poor for the older workers, with almost 60% coming from the poorest decile. Pushing for higher participation by the young entails a short-term trade-off between average gains and a better distribution. It may also entail a longer-term trade off with future incomes of the young, by reducing schooling.

Finally we test whether our impact estimator is biased due to selection on unobservables. For identification, we exclude the province dummy variables from the set of controls in the income regression, as discussed in section 3. The regression coefficient on participation (β in equation 4) was 154.358 ($t=5.049$) which is very close to the matching estimate for the nearest neighbor case.¹⁴ The coefficient on the residuals from the participation regression (γ in equation

¹⁴ Of course, if one drops the control variables and the participation residuals then the estimate is identical to that based on the mean differences between the participants and their nearest neighbors.

4) was 4.064 but this was not significantly different from zero ($t=0.402$). Evidently, selection bias on unobservables is not an important concern in our matching estimates.

5.3 *Economic Interpretation*

Although we find that program participation falls off sharply as household income rises, the net gains conditional on participation do not fall amongst the upper half of the income distribution (Figure 1). Since the program wage rate is about the same for all participants, foregone income amongst participants appears to be independent of family income above about \$200 per person per month. This may be surprising at first sight. The standard model of self-targeting through work requirements postulates that foregone income tends to be higher for higher income groups (section 2.1).

We can offer the following explanation. The Trabajar wage is almost certainly too low to attract a worker out of a regular job. For a worker with such a job, let the foregone income from joining the program be $f_e(Y) > W$ where (as in section 2.1) Y is the pre-intervention income of the worker's household, W is the wage rate offered on the Trabajar program and the function f_e is strictly increasing.

For an unemployed worker, however, only miscellaneous odd-jobs are available. Anyone can get this work, and it does not earn any more for someone from a well-off family than a poor one. Let this "odd-job" foregone income be $f_u < W$ and assume that f_u is independent of Y . Let the rate of unemployment be U and assume that this is a decreasing function of Y ; that is also consistent with the evidence for this setting (section 3). Average foregone income if one joins the Trabajar program is then:

$$F(Y)/U(Y)f_u + [1-U(Y)]f_e(Y) \tag{5}$$

This is strictly increasing in Y , as in the standard model of self-targeting (section 2.1).

In this model, unemployed workers will want to participate in the Trabajar program, while the employed will not be interested in participating (assuming that the alternative work is judged equal in other respects, although this can be relaxed without altering the main point of this model.) The program will successfully screen the two groups. We will see a fall in Trabajar participation as income rises, as in Table 4. However, when we calculate the foregone income of actual participants we will get f_u not $F(Y)$. Measured net gains amongst actual participants will not vary systematically with pre-intervention income, even though self-targeting of the poor is excellent. Our finding that foregone income conditional on participation does not fall as income rises amongst the upper half of the distribution is still consistent with good overall targeting through self-selection.

6. Conclusions

The assumptions made in program evaluations are often dictated by data availability. In assessing the gains from workfare programs it is common to only have access to a single cross-sectional survey. However, the assumption that the forgone income of participants is zero can still be tested with such data. Propensity-score matching methods of evaluation combine a single cross-sectional survey of program participants with a comparable larger cross-sectional survey from which a comparison group is chosen. With sufficiently detailed cross-sectional data on both participants and non-participants, these methods can allow an assessment of behavioral responses without pre-intervention baseline data or randomization. The accuracy of this method will depend on how well one can assure that treatment and comparison groups come from the same economic environment and were given the same survey instrument. The method cannot rule out the possibility of selection bias due to unobserved differences between participants and even a well-matched comparison group, though there is some evidence this may well be an over-rated

problem (Heckman, et al., 1998). We have applied recent advances in matching methods to Argentina's Trabajar Program. While neither a baseline survey nor randomization were feasible options in this case, the program is well suited to matching methods. We have also offered an over-identification test for selectivity bias after matching.

We find that program participants are more likely to be poor than non-participants by a variety of both objective and subjective indicators. The participants tend to be less well educated, they tend to live in poorer neighborhoods, and they tend to be members of neighborhood associations and political parties. The relatively low wage rate clearly makes the program unattractive to the non-poor.

Using our model of program participation to find the best matches from the national sample for each Trabajar worker, we have estimated the net income gain from the program. We find that ignoring foregone incomes greatly overstates the average gains from the program, though sizable gains of about half the gross wage are still found. Even allowing for foregone incomes, the distribution of the gains is decidedly pro-poor, reflecting the self-targeting feature of the programs' design. Average gains are very similar between men and women, but are higher for younger workers. Higher female participation would not enhance average income gains, and the distribution of the gains would worsen. Higher participation by the young would raise average gains, but also worsen the distribution. After matching, our tests suggest that selectivity bias (due to unobservables) is a negligible problem.

References

- Atkinson, A., 1987, "On the Measurement of Poverty", *Econometrica* 55: 749-64.
- Besley, Timothy and Stephen Coate., 1992, "Workfare vs. Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Programs", *American Economic Review* 82: 249-261.
- Datt, Gaurav and Martin Ravallion, 1994, "Transfer Benefits from Public Works Employment", *Economic Journal*, 104: 1346-1369.
- Drèze, Jean and Amartya Sen, 1989, *Hunger and Public Action*, Oxford: Oxford University Press.
- Heckman, J., H. Ichimura, and P. Todd, 1997, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economic Studies*, 64: 605-654.
- Heckman, J., H. Ichimura, and P. Todd, 1998, "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, 65: 261-294.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1998, "Characterizing Selection Bias using Experimental Data", *Econometrica*, 66: 1017-1099.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1996, "Nonparametric characterization of selection bias using experimental data: A study of adult males in JTPA. Part II, Theory and Methods and Monte-Carlo Evidence," Mimeo, University of Chicago.
- Lipton, Michael and Martin Ravallion, 1995, "Poverty and Policy", in Jere Behrman and T.N. Srinivasan (eds) *Handbook of Development Economics Volume 3* Amsterdam: North-Holland.
- Manski, Charles and Steven Lerman, 1977, "The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45: 1977-88.
- Mukherjee, Anindita, 1997, "Public Works Programmes: Some Issues", *Indian Journal of Labor Economics*, 40: 289-306.
- Ravallion, Martin, 1991, "Reaching the Rural Poor Through Public Employment: Arguments,

- Evidence and Lessons from South Asia”, *World Bank Research Observer* 6: 153-75.
- Ravallion, Martin, 1999a, “Appraising Workfare,” *World Bank Research Observer* 14: 31-48.
- Ravallion, Martin, 1999b, “Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved,” Policy Research Working Paper 2080, World Bank.
- Ravallion, Martin and Quentin Wodon, 1998, “Evaluating a Targeted Social Program When Placement is Decentralized”, Policy Research Working Paper 1945, Washington DC, World Bank.
- Rosenbaum, P. and D. Rubin, 1983, “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70: 41-55.
- Rosenbaum, P. and D. Rubin, 1985, “Constructing a Control Group using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *American Statistician* 39: 35-39.
- Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- Subbarao, K., 1997, “Public Works as an Anti-Poverty Program: An Overview of Cross-Country Experience”, *American Journal of Agricultural Economics*, 79: 678-683.
- Subbarao, K., Aniruddha Bonnerjee, Jeannine Braithwaite, Soniya Carvalho, Kene Ezemenari, Carol Graham, and Alan Thompson, 1997, *Safety Net Programs and Poverty Reduction: Lessons from Cross-Country Experience*, World bank, Washington DC.
- Todd, Petra, 1995, “Matching and Local Linear Regression Approaches to Solving the Evaluation Problem with a Semiparametric Propensity Score”, *mimeo*, University of Chicago.

Table 1: Descriptive Statistics

	Trabajar sample		National sample	
Per capita income (\$/person/month)	73.205		366.596	
	(101.843)		(792.033)	
Average household size	4.894		3.448	
	(2.509)		(1.981)	
Private pensions (\$/person/month)	10.821		18.927	
	(36.106)		(67.813)	
Social pensions (\$/person/month)	1.250		0.749	
	(6.719)		(6.896)	
Help from friends and relatives (\$/person/month)	1.515		11.893	
	(16.013)		(71.977)	
% of households who need to borrow to meet basic needs	32.777		18.820	
	(0.887)		(0.263)	
% of population participating in some form of political organization	2.910		1.450	
	(0.318)		(0.009)	
% of households who own a telephone	22.660		66.150	
	(0.791)		(0.318)	
% of households who own a color TV	75.600		77.040	
	(0.811)		(0.283)	
% of households owning a refrigerator with inbuilt freezer	26.450		48.280	
	(0.833)		(0.336)	
% of households owning an automatic washing machine	11.660		37.680	
	(0.606)		(0.326)	
	Male	Female	Male	Female
Average age at which currently active household members started working (years)	15.945	17.809	15.658	17.689
	(9.716)	(9.683)	(6.193)	(6.772)
Average age at which those household members who are no longer at school dropped out of school (years)	15.333	15.455	16.857	16.789
	(8.137)	(8.813)	(8.649)	(7.1306)
% of people in household who were unwell (accident or sick) in the last month	19.030	23.260	22.130	26.700
	(0.742)	(0.798)	(0.279)	(0.298)

Notes: Above averages are population-weighted averages. Monetary units are in \$/month, 1997 prices. Standard deviations are reported in the parentheses.

Table 2: Location of Trabajar participants in the national distribution of household income per capita

	Trabajar sample		National sample	
	Households	Persons	Households	Persons
Poorest 5%	40.2	38.8	5.0	5.6
Next 5%	18.0	21.3	5.0	7.8
Decile 2	17.5	18.5	10.0	13.1
Decile 3	9.9	9.5	10.0	11.7
Decile 4	6.8	5.8	10.0	10.9
Decile 5	2.2	1.9	10.0	9.7
Decile 6	2.5	1.6	10.0	9.1
Decile 7	1.7	1.6	10.0	9.2
Decile 8	0.6	0.5	10.0	8.2
Decile 9	0.4	0.3	10.0	7.9
Decile 10	0.2	0.1	10.1	6.7
Total	100.0	100.0	100.0	100.0

Table 3: Logit regression of participation in the Trabajar Program

	Coefficient	t-ratio
Cordoba	3.5084	8.395
Chaco	1.0953	2.750
La Pampa	1.2023	3.053
La Rioja	3.1152	7.505
Misiones	1.4492	3.630
Neuquen	1.0367	2.597
Salta	1.3164	3.332
San Juan	1.4462	3.513
Santa Fe	1.5063	3.897
Santiago del Estero	1.4058	3.572
Whether household is located in an emergency town	-0.5455	-3.284
.....a settlement of 5 ⁺ years	-0.9622	-3.998
.....a social housing area	0.3536	4.479
.....an area in very damaged condition	-0.3197	-2.747
Dwelling has 1 room (beside bathroom/kitchen)	0.7733	7.654
.....2 rooms	0.5247	6.805
.....3 rooms	0.2734	3.902
Main material of interior floors is cement/bricks	0.3028	2.579
Water is obtained from manual pumps	-0.9468	-2.902
Water shortages in last 12 months	-0.2707	-4.535
Portable gas is used for cooking	-0.5661	-2.807
Household gets hot water through a central heating service	0.6968	2.444
Located <3 blocks from a place where trash is placed habitually	-0.3360	-5.015
..... <3 blocks from a place which gets flooded	0.2218	3.284
.....in an area where there is daily collection of trash	0.1795	2.016
.....in an area with a water network	0.7348	4.396
.....in an area with sewer network	0.2779	4.073
.....<5 blocks from closest public transportation	-0.2674	-2.202
.....<5 blocks from closest public phone	-0.3044	-3.109
.....<5 blocks from closest public primary school	-0.4211	-4.419
.....5-9 blocks from closest public primary school	-0.3027	-3.180
.....<5 blocks from closest neighborhood health center	0.1675	2.309
.....5-9 blocks from closest neighborhood health center	0.1678	2.315
.....<5 blocks from closest pharmacy	-0.4265	-5.129
.....<5 blocks from closest mail	-0.2709	-2.655
.....< 10 blocks from a secondary school	-1.0198	-4.231
.....10-30 blocks from a secondary school	-1.0127	-4.253
.....30-50 blocks from a secondary school	-0.4955	-1.954
..... <10 blocks from a public hospital	-0.3943	-3.325
Safety is the major concern in the neighborhood	0.2708	2.917

It is a dangerous street for pedestrians to cross	0.1472	2.040
Shortages of electricity	0.2925	3.084
Drug addiction problem in neighborhood	0.3855	-3.786
Male	2.2307	13.961
Head of the household	0.3169	2.735
Spouse of the household head	-0.6185	-3.858
Legally married	0.2211	2.343
Separated after being married	0.4397	2.911
Divorced	0.3769	2.202
During last 12 months has been absent from h'hold for > 1 month	-0.4450	-3.182
Born in this locality	0.8215	5.019
..... in another locality of same province	0.5672	3.373
.....in another province	0.6523	3.867
Lived habitually in this locality for last 5 years	0.5326	4.876
Affiliated to a health system only through social work	-0.6388	-7.750
.....to a health system through unions and private hospital	-0.4694	-3.839
.....to health system through social work & mutual benefit society	-1.0715	-3.291
.....to health system because he is a worker	-1.1213	-6.530
Currently attends an educational establishment for primary/sec school	-0.7551	-2.117
Currently a student at tertiary school	0.8775	2.650
Dropped out of school because found syllabus uninteresting	-0.5386	-3.656
.....he/she was finding school difficult	0.6700	3.048
.....location of school was inconvenient	-0.3996	-1.951
Dropped out of school for personal reasons	0.3671	2.100
Taken a course in labor training in the last 3 years	0.4252	5.244
Never a member of a sports association	0.3444	2.826
Regular member of a neighborhood association with some admin. Responsibilities	0.9705	2.482
Regular member of a neighborhood association with no responsibilities	0.8259	2.526
Never a member of union/student association	0.5973	2.413
Member of a political party with some administrative responsibilities	0.7523	1.900
Member of a political party	1.6387	6.020
Occasional member of a political party	1.3609	5.041
Thinks that 20 years hence, economic situation will be the same as parents now	0.3981	5.401
Reason for above is lack of schooling	-0.3705	-5.632
Reason for above is economic situation of country	-0.7596	-7.291
Thinks that he and his family is very poor	0.5976	6.078
Children born in the last 12 months	0.2281	2.693
Pregnant currently	-0.9295	-2.435
Constant	-5.6210	-4.390
Log Likelihood	-5580	

Notes: Only significant coefficients in the logit regression are reported in the above table. For omitted categories and for other variables included in the regression see Addendum (available from the authors).

Table 4: Net income gains from the program using different estimators

Groups	Nearest neighbor	Nearest five estimator	Non-parametric estimator
Full sample	156.770 (296.083)	102.627 (247.433)	91.678 (230.327)
Ventile 1	372.010 (409.053)	108.543 (210.543)	107.862 (222.831)
Ventile 2	132.662 (260.851)	83.351 (200.379)	63.331 (161.769)
Decile 2	112.166 (230.161)	119.044 (285.357)	93.506 (197.679)
Decile 3	102.058 (176.515)	136.349 (263.939)	120.430 (240.703)
Decile 4	78.740 (248.272)	82.386 (281.863)	89.295 (277.294)
Decile 5	148.711 (434.210)	107.125 (208.313)	205.050 (597.605)
Decile 6 – 9	80.965 (191.337)	111.229 (278.584)	114.913 (196.906)
Decile 10	No participants in this decile		

Note: Standard errors in parentheses.

Table 5: Persons of participant households using different estimators

Groups	Nearest neighbor estimator	Nearest five estimator	Non-parametric estimator
Full sample	100.000	100.000	100.000
Ventile 1	21.525	10.207	8.671
Ventile 2	41.278	42.284	39.460
Decile 2	20.732	26.908	27.734
Decile 3	8.084	10.892	13.460
Decile 4	5.403	6.307	7.302
Decile 5	1.842	2.069	1.652
Decile 6 – 9	1.135	1.334	1.722
Decile 10	No participants in this decile		

Table 6: Net income gains from the program

Groups	% of participants in ventile/decile	Persons of participant households	H'hold income of Trabajar participants	Net income gain due to the program	Net gain as % of pre-intervention income
Full sample	100.000	100.000	501.181 (364.632)	102.627 (247.433)	25.926
Ventile 1	6.070	10.207	299.102 (221.119)	108.543 (210.543)	74.830
Ventile 2	36.535	42.284	369.194 (265.054)	83.351 (200.379)	24.746
Decile 2	26.700	26.908	548.789 (353.237)	119.044 (285.357)	26.566
Decile 3	12.601	10.892	685.413 (358.139)	136.349 (263.939)	23.056
Decile 4	11.833	6.307	543.680 (441.794)	82.386 (281.863)	13.483
Decile 5	3.496	2.069	749.443 (384.025)	107.125 (208.313)	14.975
Decile 6 – 9	2.766	1.334	879.382 (496.091)	111.229 (278.584)	11.469
Decile 10	No participants in this decile				

Notes: These numbers correspond to the nearest five estimator reported in Table 4. Standard errors in parentheses.

Table 7: Net income gains for female participants

Groups	% of participants in ventile/decile	Persons of participant households	H'hold income of Trabajar participants	Net income gain due to the program	Net gain as % of pre-intervention income
Full sample	100.000	100.000	571.890 (382.580)	103.904 (277.340)	22.818
Ventile 1	3.289	5.645	351.300 (428.177)	158.240 (409.963)	82.298
Ventile 2	25.000	31.948	424.370 (320.742)	101.360 (281.681)	30.767
Decile 2	32.895	34.000	520.800 (286.501)	87.490 (202.641)	18.400
Decile 3	16.447	15.261	718.660 (493.045)	136.284 (420.507)	21.166
Decile 4	12.500	8.251	655.579 (322.183)	92.353 (196.851)	14.123
Decile 5	4.605	2.605	696.143 (224.638)	79.000 (126.926)	12.558
Decile 6 – 9	5.263	2.295	963.663 (473.150)	132.663 (248.887)	14.006
Decile 10	No participants in this decile				

Notes: These numbers correspond to the nearest five estimator for the sub-group of female participants. Standard errors in parentheses.

Table 8: Income gains for those 15-24 years of age

Groups	% of participants in decile	Persons of participant households	H'hold income of Trabajar participants	Net income gain due to the program	Net gain as % of pre-intervention income
Full sample	100.000	100.000	618.789 (401.990)	125.241 (255.903)	25.592
Decile 1	30.214	37.012	434.619 (332.660)	121.500 (261.500)	35.287
Decile 2	31.567	34.431	636.060 (353.555)	143.657 (272.418)	28.629
Decile 3	16.234	14.776	738.666 (383.006)	133.560 (275.162)	19.921
Decile 4	11.838	8.313	620.135 (378.544)	73.146 (169.706)	10.559
Decile 5	10.034	3.618	886.735 (422.0520)	152.898 (262.636)	17.400
Decile 6 - 9	3.495	1.850	1,069.600 (608.221)	102.142 (176.652)	9.550
Decile 10	No participants in this decile				

Notes: These numbers correspond to the nearest five estimator for the sub-group of 15-24 year participants. Standard errors in parentheses.

Table 9: Income gains for those 25-64 years of age

Groups	% of participants in decile	Persons of participant households	H'hold income of Trabajar participants	Net income gain due to the program	Net gain as % of pre-intervention income
Full sample	100.000	100.000	443.443 (328.253)	85.820 (231.032)	22.241
Bottom 5%	7.423	13.062	307.386 (251.260)	97.474 (221.489)	38.564
Next 5%	39.451	45.767	342.499 (252.305)	71.809 (205.962)	22.207
Decile 2	26.651	24.938	487.939 (251.477)	86.833 (180.047)	21.204
Decile 3	11.046	8.851	625.097 (395.1020)	122.505 (334.238)	25.578
Decile 4	10.812	5.046	476.941 (410.221)	74.724 (271.968)	13.594
Decile 5	2.221	1.343	755.921 (561.663)	123.176 (331.995)	13.996
Decile 6-9	2.396	0.993	753.736 (437.869)	115.478 (224.021)	15.834
Decile 10	No participants in this decile				

Notes: These numbers correspond to the nearest five estimator for the sub-group of 25-64 year participants. Standard errors in parentheses.

Figure 1: Mean Income Gain Plotted Against Pre-Intervention Income

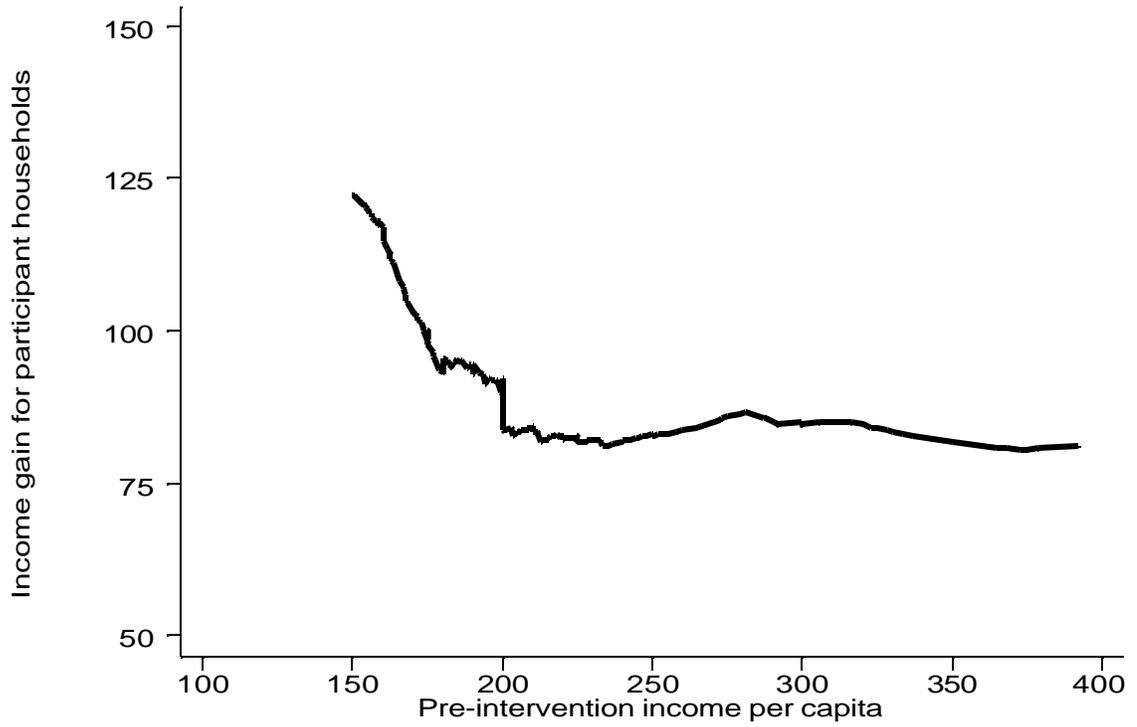


Figure 2: Empirical Distribution Functions

