Do Workfare Participants Recover Quickly from Retrenchment?

Martin Ravallion, Emanuela Galasso, Teodoro Lazo, Ernesto Philipp[1]

Development Research Group
World Bank

Trabajar Project Office
Ministry of Labor
Government of Argentina

September 2001

Abstract: The incomes of continuing participants in a workfare program in Argentina are compared to those of a matched comparison group of non-participants and with a matched group of past participants who leave the program. We find partial income replacement, amounting to one-quarter of the gross workfare wage within six months of leaving the program, rising to one half in 12 months. The estimates are unbiased in the presence of a time-invariant errors due to miss-matching. Fully removing selection bias would probably yield lower income replacement. Test results based on a second follow-up survey suggest that valid inferences can be drawn about program impacts from our measures of income replacement.

Keywords: workfare; propensity-score matching; double-difference; Argentina
JEL classifications: H43, I38

---

[1] The work reported in this paper is part of the ex-post evaluation of the World Bank's Social Protection III Project in Argentina. The authors' thanks go to staff of the Trabajar project office in the Ministry of Labor, Government of Argentina, who have helped in countless ways, and to the Bank's Manager for the project, Polly Jones, for her continuing support of the evaluation effort and many useful discussions. We also benefited from discussions with Jyotsna Jalan and the comments of Guillermo Perry and seminar participants at the Ministry of Labor. Support from the Evaluation Thematic group of the World Bank's Poverty Reduction and Economic Management Network is gratefully acknowledged. These are the views of the authors, and need not reflect those of the Government of Argentina or the World Bank. Correspondence: Martin Ravallion, World Bank, 1818 H Street NW, Washington DC, 20433 USA; mravallion@worldbank.org.

1. Introduction

The welfare outcomes of cutting a workfare program—which imposes work requirements on welfare recipients—will depend in part on labor market conditions facing the participants. High income replacement after retrenchment might suggest that unemployment is not a serious poverty problem. But even when there is high unemployment, there are other ways that retrenched workers might recover the lost income. Possibly the work experience on the program will help them find work, including self-employment. Or possibly private transfers will help make up for the loss of public support. Tracking ex-participants after their retrenchment and measuring their income replacement may thus provide important clues to understanding the true impact of a workfare program.

This paper tires to learn about the impact of a workfare program by studying income replacement for those observed to leave the program after its contraction. The analytic problem we face is the usual one in causal studies of missing data on the counter-factual. It is well recognized that "single difference" comparisons of income levels between participants and non-participants can be highly misleading given the existence of (observable and unobservable) heterogeneity in characteristics that jointly influence participation and incomes in the absence of the program. Simulations and comparisons with actual experiments have suggested that careful matching in terms of observable covariates can greatly reduce the bias in observational studies.[2] Amongst the various matching methods available, Propensity Score Matching (PSM) has attracted recent interest given its theoretical properties, notably that exact matching by this

[2] For evidence based on simulations see Rubin (1979) and Rubin and Thomas (2000). For evidence based on an actual evaluation see Dehejia and Wahba (1999) who find that single-difference matching based on propensity scores gives a good approximation to results of a randomized evaluation of a US training program — much better than the non-experimental methods that Lalonde (1986) assessed for the same program. However, Smith and Todd (2000) question the robustness of Dehejia and Wahba's findings to model specification.

method is the observational equivalent of randomization (Rosenbaum and Rubin, 1983, 1985). PSM gives unbiased estimates if (inter alia) the conditional independence ("strong ignorability") assumption holds, whereby pre-intervention outcomes are independent of participation given the variables used for matching (Rosenbaum and Rubin, 1983). Conditional dependence will leave a bias, which will depend on the amount of relevant data available for matching.

Another approach in the literature is the popular double difference (DD) estimate, obtained by comparing treatment and comparison groups in terms of outcome changes over time relative to a pre-intervention baseline. DD allows for conditional dependence arising from additive time-invariant latent heterogeneity. Since PSM optimally balances observed covariates between the treatment and comparison groups, it is the obvious method for selecting the comparison group in double-difference studies. The results of Heckman, Ichimura and Todd (1997), Heckman et al., (1998), Heckman and Smith (1998) and Smith and Todd (2000) suggest that a hybrid method, combining PSM for selecting the comparison group with DD to eliminate time-invariant errors, can greatly reduce (but not eliminate) the bias found in other evaluation methods, including single-difference matching.

DD estimators have their limitations. In some circumstances it is implausible that the selection-bias is time invariant. For example, there is a potential bias in DD estimators when the changes over time are a function of initial conditions that also influence program placement.[4] There is also the well-known bias for inferring long-term impacts that can arise when there is a

[3] For recent evidence based on simulations see Rubin and Thomas (2000). For evidence based on actual evaluation see Dehejia and Wahba (1999) who find that single-difference PSM given a good approximation to results of a randomized evaluation of a US training program — much better than the non-experimental methods that Lalonde (1986) assessed for the same program. Smith and Todd (2000) question the robustness of Dehejia and Wahba's findings to model specification.

[4] Jalan and Ravallion (1998) show that this can seriously bias evaluations of poor-area development programs that are targeted on the basis of initial geographic characteristics that also influence the growth process.

pre-program earnings dip (known as "Ashenfelter's dip" following Ashenfelter, 1978). In assessing short-term impact — a common concern of safety-net interventions —one would not normally want to ignore this dip, though it remains relevant to assessing the time profile of gains from the safety net.

What if one does not have a pre-intervention baseline? This is common for safety-net interventions, such as workfare programs, that have to be set up quickly, in response to a macroeconomic or agro-climatic crisis. There is no time to do a baseline survey of (probable) participants and non-participants. Nor is randomization usually feasible in such settings. Suppose instead that we follow up samples of participants and non-participants over time, post-intervention, and that some participants become non-participants. What can we then learn about the program's impacts?

The approach we propose here is to examine what happens to participants' incomes when they leave a workfare program, and to compare this with the incomes of continuing participants, after netting out economy-wide changes, as revealed by a matched comparison group of non-participants. While this approach is feasible without a baseline survey, it brings its own problems. Firstly, while differencing over time can eliminate bias due to latent (time-invariant) matching errors, there remains a potential bias due to any selective retrenchment from the program based on unobservables. We argue that the direction of bias can be determined under plausible assumptions.

Secondly, while we are not concerned with any pre-program "Ashenfelter's dip," there may well be a post-program version of the same phenomenon, namely when earnings drop sharply at retrenchment, but then recover. As in the pre-program dip, this need not be a source of bias in assessing the impact of a safety-net intervention (to the extent that the pre-program dip entails a welfare change); nonetheless, the post-program dip is clearly of interest in assessing the dynamics of recovery from retrenchment. To help address this issue we follow up initial participants over multiple survey rounds.

We are also interested in seeing whether this type of follow-up study of participants can identify the gains to current participants from a program — the classic "treatment effect on the treated" as it is called in the evaluation literature. There are concerns about selection bias, and there is the problem that past participation may bring current gains to those who leave the program. Assuming these lagged gains are positive, the net loss from leaving the program will be less than the gain from participation relative to the counter-factual of never participating. We derive a test for the joint conditions needed to identify the mean gains to participants from this type of study, also exploiting further follow-up surveys of past participants.

We study Argentina's Trabajar Program. This government program aims to provide work to poor unemployed workers on approved sub-projects of direct value to poor communities. The sub-projects cannot last more than six months, though a worker is not prevented from joining a new project, if available. In earlier research on the same program, Jalan and Ravallion (1999) estimated the counter-factual income of current participants if they had not participated using the mean income of a comparison group of non-participants, obtained by PSM. For the purpose of the present study, we designed a survey of a random sample of current participants, and returned to the same households six months later, and then 12 months later. In addition to natural rotation, there was a very sharp contraction in the program's aggregate outlays after the first survey.

The following section describes the program and the data for its evaluation. Section 3 describes our evaluation method in theoretical terms. Section 4 presents our results, while some conclusions can be found in Section 5.

2. The program and data

In response to a sharp increase in the measured unemployment rate, the Government of Argentina greatly expanded and redesigned its Trabajar Program in May 1997, with financial and technical support from the World Bank. The Trabajar Program aims to provide short-term work at relatively low wages on socially useful projects in poor areas. The projects are proposed by local (governmental and non-governmental) organizations with priority given to proposals that are likely to benefit poor areas, according to ex-ante assessments. Workers cannot join the program unless they are recruited to an approved project. The projects last a maximum of six months, but a worker is not prevented from switching to a new project on the same basis.

The wage rate was initially set at a maximum of $200 per month, which was cut to $160 in 1999 at the time of an overall contraction in outlays. (Undercutting of the wage rate is allowed, but it is uncommon.) The wage rate 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. By way of comparison, the average monthly earnings for workers in the poorest 10% of households (ranked by total income per person) in Greater Buenos Aires (GBA) in May 1996 was $263 (calculated from the Permanent Household Survey, discussed further below). (As expected, the poorest decile also received the lowest average wage, and average wages rose monotonically with household income per person.)

The data collection for this study began with a survey in May/June 1999 of Trabajar participants in the main urban areas of three provinces — Chaco (Gran Resistencia), Mendoza (Gran Mendoza) and Tucuman (Gran Tucaman – Tafi Viejo). These provinces were chosen as representing the range of labor markets found in Argentina. The families of 1500 randomly chosen Trabajar workers were interviewed, spread evenly between the three provinces. The sampled beneficiary households were a simple random sample from the list of all beneficiaries at the time. The households of participants were the units for interviewing.

The survey of participants was chosen to coincide with the twice-yearly Permanent Household Survey (PHS). This is an urban survey focusing on employment and incomes, though it also includes questions on education and demographics. We calculate individual income from questions on income from work (wages, bonuses, self-employment income, Trabajar earnings) and from non-labor sources (pension, rents, dividends, fellowships, food coupons, private transfers). All provincial capitals or other urban centers with at least 100,000 inhabitants are included in the PHS.[5] The survey is conducted twice a year, around May and October. The PHS sample size is set to achieve (with 95% confidence) an error of 1% in the unemployment rate within each urban conglomerate. In large conglomerates, a random sample of geographic units is chosen, within which a fixed number of households is sampled. In smaller conglomerates, a one-stage random sample is used. The PHS sample includes 27,000 households.

The PHS is our source of the comparison group for initial participants, to be selected by propensity-score matching, as described in more detail in the next section. For program participants, the same interview questionnaire was used as for the PHS, with PHS interviewers. This avoids the matching bias that can arise when the surveys of participants and non-participants are not comparable (Heckman, Ichimura and Todd, 1997; Heckman et al., 1998). Extra questions were added for the survey of Trabajar participants. Moreover, miss-matching can be reduced by selecting the comparison group separately from each geographic area, to make sure that the individuals belong to the same local labor markets.[6]

5 An exception is Viedma, capital of Rio Negro, that was replaced for the urban-rural conglomerate of Alto Valle del Rio Negro.

6 Heckman et al (1998) find that the mismatch due to different questionnaire and different labor markets amounts to half of the selection bias in their analysis.

A follow-up survey of the same Trabajar participants was done in October/November 1999, to coincide with the next round of the PHS, and similarly in May/June 2000.[7] The PHS has a rotating panel design with one quarter replaced each round, so it was possible to form a panel for the comparison group. Our matches were constrained to only include those who would be followed up. Naturally this limits the matching options — particularly so by the second follow-up survey, by which time only half of the original sample is re-surveyed.

The PHS is a far shorter survey instrument than that used by Jalan and Ravallion (1999) for their single difference estimate of the impact on incomes of Trabajar participation. Since there are fewer observables in the data, the matching is unlikely to be as good. Results in the literature suggest that single-difference PSM estimates can be unreliable when the data available do not include important determinants of participation (Heckman et al., 1997, 1998; Smith and Todd, 2000). However, here we have the advantage that we can follow up participants over time, exploiting the rotating panel design of the PHS. Thus, although we cannot expect that our single difference PSM estimates will be as reliable as in Jalan and Ravallion, we can eliminate the time-invariant errors due to miss-matching arising from violations of the conditional independence assumption.

There was a sharp contraction is Trabajar participation after the first two surveys. 49% of the Trabajar workers interviewed in the baseline survey were no longer employed under the program in the first follow-up survey (Table 1). Only 16% of the original Trabajar workers were employed on the program by the second follow-up survey. This contraction in employment on the program did not appear to stem from a "pull" effect from the rest of the economy. There was little sign of economic recovery between the surveys. The overall unemployment rate increased

[7] A fourth survey was done six months later. Over 90% of the initial Trabajar workers had left the program by the fourth wave. There were too few continuing participants to facilitate further analysis.

in one of the provinces (Chaco) and fell, but not greatly, in the other two; see Table 2, which also gives unemployment rates for the second follow up survey, six months later, and for six months prior to the first survey.

The large number of participants leaving the program appears instead to be due to a normal process of rotation arising from the fact that projects do not last longer than six months. When a project ends, its beneficiaries are not incorporated automatically in another project. The responsible organizations are the ones that select the participants. In the country as a whole, 45% of Trabajar workers participate in only one project (46% in Chaco, 52% in Mendoza and 51% in Tucuman).

On top of this designed rotation, there was a severe contraction in aggregate outlays on the program starting at the end of 1999. This was an outcome of overall fiscal austerity, to keep Argentina within macroeconomic targets. Aggregate spending on the program by the center in the first five months of 2000 was only 29% of its level in the last five months of 1999. Existing projects were completed, but the number of new projects approved shrank sharply in the latter part of 1999, to bring down the center's outlays. As already noted, the wage rate was also cut; Table 3 gives the sample mean wages by survey round. The aggregate cuts to the program made it less likely that past participants would find another project to join. A large new workfare program, the Emergency Employment program took up some of the slack in 2000. This was not in operation by the time of the second survey (first follow-up survey), but it was by the third survey. While our impact estimates using the first and second surveys are not likely to be affected by this new program, this is not true of the results using the third survey.

3. Estimation methods

Our strategy is to compare income changes between those who stay in the program and those who leave, after netting out the income changes for an observationally similar comparison group of non-participants. This is an example of what has been called in the literature a "difference-in-difference-in-difference" or "triple-difference" estimate.[8] We first discuss our method of selecting comparison groups, both of initial participants and for continuing participants. We then describe our version of the triple-difference estimator.

3.1 Controlling for observed heterogeneity

We use PSM to balance observed covariates at two stages. Firstly we form a matched comparison group for initial participants and secondly we match those who continue to participate over time ("stayers") with those that drop out ("leavers"). The second stage matching deals with the observed differences between subsequent leavers and stayers.

PSM balances the distributions of covariates between participants and a comparison group based on similarity of their predicted probabilities of participation (their "propensity scores"). Rosenbaum and Rubin (1983) show that exact matching on the basis of propensity scores eliminates the bias in identifying the causal effect due to covariates. PSM is thus the observational analog of an experiment in which participation is independent of outcomes; the difference is that a pure experiment does not require the untestable assumption of independence conditional on observables.

---

[8] The triple-difference method appears to have been first used by Gruber (1994) who included interaction effects between time and location (as well as separate time and location effects) in modeling the earnings effects of labor laws in the US.

Two groups are identified: those that participate (D_i =1) and those that do not (D_i=0). We rule out interference between units under the assumption that the gain to a worker from participation in a program such as Trabajar does not spillover to nonparticipants.[9] Participants are matched to individuals who did not participate on the basis of the propensity score, defined as $P(x_i) = \Pr(D_i = 1 | x_i)$ where x_i is a vector of pre-exposure control variables. Rosenbaum and Rubin (1983) prove that if the D_i's are independent over all i, and outcomes are independent of participation given x_i (i.e. unobserved differences do not influence whether or not i participates), then outcomes are also independent of participation given $P(x_i)$, just as they would be if participation was assigned randomly.[10] The value of $P(x)$ is used to select control subjects for each of those treated. This eliminates bias in estimated treatment effects due to differences in the covariates.

In practice the propensity score must be estimated. Here we follow the standard practice in PSM applications of using the predicted values from standard logit models to estimate the propensity score for each observation in the participant and the comparison-group samples. Using the estimated propensity score, matched-pairs are constructed 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 $[\hat{P}(x_i) - \hat{P}(x_j)]^2$ over all j in the set of non-participants, where $\hat{P}(x_k)$ is the predicted propensity score for observation k. Matches are only accepted if $[\hat{P}(x_i) - \hat{P}(x_j)]^2$ is less than 0.00001 (an absolute difference less than 0.0032.) We only include those observations on non-participants that share a common range of values of the propensity

9 In the matching literature, this is the stable unit-treatment value assumption (Rosenbaum and Rubin, 1983).

10 The assumption that outcomes are independent of participation given x_i is variously referred to in the literature as the "conditional independence," "strong ignorability," or "selection on observables."

scores calculated for the participants (i.e., the two groups share common support in the predicted propensity scores). In the following analysis, the comparison group for each participant is defined as the set of five nearest neighbors amongst non-participants in terms of the predicted propensity scores.

3.2 Latent heterogeneity

PSM gives unbiased estimates of program impact if selection into the program is based solely on observables; selection bias due to (non-ignorable) latent heterogeneity will remain in single difference comparisons using PSM. With access to a pre-intervention baseline, one can eliminate time-invariant selection bias due to unobservables, by differencing over time. One can also deal with time-invariant selection bias with only post-intervention data, using follow-up surveys. However, in doing so one must recognize that the gains to current non-participants need not be symmetric before and after participation; while one may be happy to assume that baseline units are unaffected by the program, it is far less plausible that drop outs gain nothing currently from their past participation. So there are two distinct sources of selection bias in our set-up. One is in the existence of latent heterogeneity in who participates in the program, leading to miss-matching in determining the comparison group, and hence a systematic error in measuring the counter-factual income. Secondly, latent heterogeneity may affect the decision to stay in the program or drop out. Participants with high (unobserved) gains from participation may well be less likely to drop out of the program. This source of bias does not disappear.

The double difference ("difference-in-difference" or *DD*) estimate is the difference in the income gains over time between a treatment group of program participants and a matched comparison group of non-participants. Our triple difference estimate (*DDD*) is defined as the difference between the value of the double difference for stayers and leavers.

Without loss of generality we can write the observed income of a Trabajar participant at date t as:

$$Y_{it}^{T} = Y_{it}^{*} + G_{it} \qquad (t \geq 1) \tag{1}$$

where Y_{it}^{*} is the counter-factual income of the Trabajar participant if the program had not existed, and G_{it} is the income gain from participation (either participation at that date or previously).

An indicator of the counter-factual income is available for a matched comparison group and is given by Y_{it}^{C}. This is a noisy indicator due to miss-matching arising from latent heterogeneity. Thus the single difference estimator is potentially biased. We make the standard assumption in double-difference studies that the selection bias is time invariant, and so it is swept away by taking differences over time. More precisely, the first difference of Y_{it}^{C} is assumed to provide an unbiased estimate of the first difference of Y_{it}^{*}:

$$E(\Delta Y_{it}^{C}) = \Delta Y_{it}^{*} \tag{2}$$

where Δ refers to the difference between the value at t and t-1. From (1) and (2) it is evident that:

$$E[\Delta(Y_{it}^{T} - Y_{it}^{C})] = \Delta G_{it} \tag{3}$$

In the usual double difference set up, period 1 precedes the intervention, and it is assumed that $G_{i1} = 0$ for all i. However, in our case, the program is in operation in period 1. The scope for identification arises from the fact that some participants at date 1 drop out of the program at date 2. Let $D_{it} = 1$ if individual i stays in the program, and let $D_{it} = 0$ if she does not. Our triple-difference estimator for t=2 is then:

$$DDD \equiv E[\Delta(Y_{i2}^T - Y_{i2}^C)|D_{i2} = 1] - E[\Delta(Y_{i2}^T - Y_{i2}^C)|D_{i2} = 0] =$$

$$[E(G_{i2}|D_{i2} = 1) - E(G_{i2}|D_{i2} = 0)] - [E(G_{i1}|D_{i2} = 1) - E(G_{i1}|D_{i2} = 0)] \qquad (4)$$

The first term in square brackets on the far RHS of (4) is the net gain to continued participation in the program, given by the difference between the gain to participants in period 2 and the gain to those who dropped out. This provides our measure of income replacement. Notice that there may be some gain from past participation for those who drop out ($E(G_{i2}|D_{i2} = 0) \neq 0$). For example, current participants may learn a skill that raises future income. Even so, $E(G_{i2}|D_{i2} = 1) - E(G_{i2}|D_{i2} = 0)$ gives the income loss to those who leave the program, allowing for the possibility that leavers may benefit from past participation.

The second term on the RHS of (4) is the selection bias arising from any effect of the gains at date 1 on participation at date 2. Under the conditional independence assumption of PSM, this term equals zero and *DDD* then gives the mean net gain to stayers in the program. It is readily verified from (4) that *DDD* gives the current gain to participants ($E(G_{i2}|D_{i2} = 1)$) if (in addition) there are no current gains to non-participants i.e., one can set $E(G_{i2}|D_{i2} = 0)$.

Unlike our matching of non-participants in the first survey (to form the comparison group), we cannot relax the conditional independence assumption in drawing inferences from our matching of stayers with leavers. The most plausible way in which this assumption would not hold is that leavers tend to be those with lower gains in the first period. Then $E(G_{i1}|D_{i2} = 1) > E(G_{i1}|D_{i2} = 0)$ and so our *DDD* estimator based on (4) will underestimate the net gain to continued participation in the program.

Notice also that if there is no latent heterogeneity then we can estimate mean gains using only the single difference, by comparing mean income of participants with that of the matched control group; this is the estimator used by Jalan and Ravallion (1999). Similarly, to obtain $E(G_{i2}|D_{i2}=1)$ one would simply take the difference in mean incomes of those who stayed in the program and the comparison group, and to obtain $E(G_{i2}|D_{i2}=0)$ one would make the same calculation for leavers. By comparing the two estimates of $E(G_{i2}|D_{i2}=1)$ - $E(G_{i2}|D_{i2}=0)$ we determine how much bias there is in the matching estimator due to unobserved heterogeneity.

So far we have focused on t=2. A third round allows a joint test of the conditions required for interpreting *DDD* as an estimate of the gains to current participants in period 2. Recall that those conditions are that $E(G_{i2}|D_{i2}=0)$ (no current gain to ex-participants) and that $E(G_{i1}|D_{i2}=1)=E(G_{i1}|D_{i2}=0)$ (no selection bias in terms of who leaves the program). Suppose one decomposes the aggregate estimate of *DDD* according to whether or not a person leaves the program in the third survey round. If there is no selection bias then this decomposition gives:

$$\begin{aligned} DDD = {} & [E(G_{i2}|D_{i2}=1, D_{i3}=1) - E(G_{i2}|D_{i2}=0, D_{i3}=1)]\Pr(D_{i3}=1) + \\ & [E(G_{i2}|D_{i2}=1, D_{i3}=0) - E(G_{i2}|D_{i2}=0, D_{i3}=0)]\Pr(D_{i3}=0) \end{aligned} \tag{5}$$

If in addition there are no current gains to non-participants then this simplifies further to:

$$DDD = E(G_{i2}|D_{i2}=1, D_{i3}=1)\Pr(D_{i3}=1) + E(G_{i2}|D_{i2}=1, D_{i3}=0)\Pr(D_{i3}=0) \tag{6}$$

However, in the absence of selection bias the two terms in this decomposition will be equal, i.e.,

$$DDD = E(G_{i2}|D_{i2}=1, D_{i3}=1) = E(G_{i2}|D_{i2}=1, D_{i3}=0) \tag{7}$$

We will test this implication. If it holds in the data then we can interpret *DDD* as the gain to current participants (the “mean effect of the treatment on the treated”). If it fails then either or both of the two conditions above fail, though we do not know which it is.

4. Results

Recall that there are two matches that need to be done, one for selecting the comparison group of non-participants in the first survey and one for balancing observed covariates between subsequent leavers and stayers. Table 4 gives the logit regressions used in constructing the propensity scores for these two stages of matching. In modeling whether an initial participant drops out in the second round we can make use of a richer set of questions available only for the Trajabar sample. Extra questions on participation in neighborhood associations and indicators of whether the selection into the program was due to personal contacts (with various actors) allow us to measure the importance of social networks to program participation, found to be important in Jalan and Ravallion (1999). Moreover, we have information on the workers' labor force histories (prior to joining Trabajar). The evaluation literature has recently emphasized changes in labor force status as an important determinant of participation in training programs.[11]

In the second-stage matching we find that the additional variables for the Trabajar sample are jointly significant in explaining who drops out of the program. However, the first logit regression (used to determine the comparison group of non-participants from the national sample) still has far higher predictive ability than the second-stage logit, as indicated by the pseudo R^2. The lack of observable correlates of which individuals dropped out adds weight to the *a priori* arguments in section 2 that this was not due to "pull" factors, which one would expect to be correlated with observables (such as education). The significant determinants of program participation in the first regression seem plausible.

[11] See for example Dehejia and Wahba (1998, 1999), Heckman and Smith (1998), Heckman, Lalonde and Smith (1999).

Of the original sample of 1459 Trabajar participants, we restrict the sample to those workers aged 15-65. 264 observations had to be dropped because satisfactory matches were not available in the PHS for both survey rounds, or key data were missing.[12] The 1195 Trabajar participants were then matched with 1868 distinct individuals in the PHS (allowing up to five matches, and with replacement). After forming the panel across the two surveys, we ended up with a sample of 1018 Trabajar participants who could be matched satisfactorily and followed up in the second round. Figure 1 gives the distribution of the log of the odds ratios for Trabajar samples and the PHS samples by province. The vertical lines give the regions of common support. Note that we are mainly losing observations from the PHS sample with low probabilities of participating in Trabajar.

In the second survey, 520 of the Trabajar participants from the first round dropped out of the program. After matching on the basis of the propensity scores based on Table 4 (column 2), we had 419 stayers matched with 400 leavers.

Table 5 gives the calculations of *DDD*. Our estimate of the income gain to stayers from participation in the program, net of the income gain attributed to past participation, is $140 per month, about three quarters of the gross wage. Table 6 repeats the calculations of Table 5, but this time no matching is done in the second survey. The results are very similar, consistent with our expectation that the bulk of the people dropping out of the program were doing so involuntarily. For if withdrawal from the program had been voluntary then we would expect it to be correlated with observed correlates and (hence) that the second-stage matching would make a difference to our results.

[12] 36 were outside the region of common support, 137 did not satisfy the maximum absolute difference in propensity scores of 0.00001, and 64 did not have at least one match in both survey rounds.

It is notable how poorly the single difference estimator performs in the first survey, with no significant positive impact indicated (Tables 5 and 6). There is clearly a large bias due to latent heterogeneity in the single difference estimator in our data. However, the single difference estimator comparing stayers and leavers in the second round does better; indeed, it gets closer to the *DDD* estimate than the double difference estimate for program leavers.

In Table 7 we give a breakdown of the results by province. The results suggest that the losses from retrenchment are smaller for areas with less tight local labor markets: our estimate of *DDD* is lowest for Mendoza, where the unemployment rate is also lowest (Table 2).

Table 8 gives the analogous results to Table 6, at household level. The "bottom line" is that we find no sign of a spillover effect on the earnings or non-labor incomes of other household members. Measured impacts are confined to the incomes of Trabajar workers. (Of course there are sure to be consumption gains to others within the household, via some degree of pooling.)

Table 9 gives the results when we look at income replacement over 12 months. Naturally we cannot do this over the same samples as before, and (in particular) the sample of continuous stayers is greatly reduced. The third survey round does indicate a sizable recovery of income for leavers from the second round; the treatment group's mean income dips from $228 per month in the first survey to $86 in the second, but rebounds to $138 in the third (Table 9). This possibly reflects in part income gains from the new Emergency Employment Program introduced in 2000.[13] However, there is also evidence of a rise in private employment. Amongst those who left the program in round 2, 49% had found jobs in that round, and the work was deemed to be "permanent employment" in 34% of cases; by round 3, the proportion employed had risen to

[13] The special Trabajar module has a question on whether individuals participated in another temporary program (other than Trabajar); 5% of the Trabajar sample report doing so in the third survey, as opposed to 2% in the second survey.

58% and 59% were permanent jobs. The *DDD* estimate for round three indicates that about half of the second round loss is recovered. On aggregating over the two rounds, the net income gain to stayers is $73 per month (=130-58), representing slightly less than one half of the mean gross wage in round 3 (Table 2a).

What can we conclude about net income gains from the program? Recall that our *DDD* estimator also gives the net gain to current participation if there is no selection bias and no current gains to non-participants who had previously participated (section 3). Given that there was a large contraction imposed on the program, selection bias might not be considered an important concern (section 2). What about lagged effects? One source of evidence can be found in qualitative questions added to the third round of the survey, on whether current or past Trabajar participants felt that the program had improved their earning opportunities outside the program. Table 10 summarizes the results. A high proportion of respondents felt that the program improved their chances of getting a job; roughly half felt that it gave them a marketable skill; about one quarter felt that it expanded their contacts. These results are suggestive of income gains from the program to ex-participants. However, there are two caveats. The expected gains may take some time to materialize, depending on the aggregate labor market conditions. Secondly, there may well be biases in answering qualitative questions of this sort. Cognitive dissonance may lead participants to prefer to believe that there will be future gains beyond their current participation in the program.

Table 11 gives our results in testing the joint hypothesis of no current gains from past participation and no selection bias (equation (7)). Under the null, the gains from participating in the program in period 2 should not depend on participation in period 3. We are not able to reject the null hypothesis (p-value 0.47). This provides a justification for interpreting *DDD* as the mean gain to current participants.

It is of interest to compare our results with those of Jalan and Ravallion (1999). The latter paper used single-difference matching on a richer data set. Their estimated mean net gain to Trabajar participation was $103 per month, rising to $157 using nearest-neighbor matching. Our single-difference estimate gives an implausible results that deviates greatly from Jalan and Ravallion. This is not true of our *DDD* estimate in Table 5 of $140. Aggregating over the three rounds, our estimated income gain to participants of $73 is less than Jalan and Ravallion obtained.[14] But this is what we would expect as long as participants are able in time to recover a greater amount of the income lost from retrenchment than initially observed.

From the point of view of evaluation design, our results suggest a trade off between the resources devoted to cross-sectional data collection for the purpose of single-difference matching, versus collecting longitudinal data with a lighter survey instrument. The lighter instrument we have used here was not able to deliver plausible single-difference estimates using PSM, when compared to prior estimates using richer data for the same program. However, it would appear that we have been able to satisfactorily address this problem by tracking households over time, even using the lighter instrument.

5. Conclusions

To see what happens to workfare participants after they leave the program, we have matched a random sample of participants in Argentina's Trabajar Program with a group of non-

[14] The estimated forgone income of participants as a proportion of the gross wage is higher than Datt and Ravallion's (1994) estimate for a workfare program in India, though the latter setting was arguably one in which unemployment was higher.

participants drawn from a strictly comparable national survey. We then followed both samples over time, during a period of aggregate program contraction on top of designed rotation of program beneficiaries. We have used propensity score matching methods to balance observed covariates at two stages: between initial participants and the comparison group, and between those who left the program and those who stayed. Since we track outcomes over time, we can eliminate any time-invariant selection bias in the first matching. Selection bias remains in the second stage matching, though we can sign the bias under plausible assumptions. The fact that there was a large centrally-imposed contraction in program outlays as well as designed rotation helps reduce concerns about selection bias at the second-stage matching.

We find that the estimated income losses to those who left the program were sizable, representing about three-quarters of the gross wage within the first six months, though falling to slightly less than half over 12 months, indicating existence of a post-program version of “Ashenfelter’s dip”. Fully removing selection bias would probably yield even lower estimates of income replacement.

Interpreting our triple-difference estimate as a measure of the gains from the program requires two conditions: that there is no selection bias in leaving the program, and that there are no lagged income effects from past participation. On a priori grounds we find the selection bias argument implausible in this setting. On the other hand, the existence of lagged income effects is supported by qualitative questions in the survey. We have proposed a joint test of these conditions, based on comparing the triple difference estimate for those who left versus stayed in a third round of the survey. Statistically, we cannot reject the conditions required for using our triple-difference measure as an estimate of the gains to current participants. We conclude that the program generated sizable net income gains to participants.

While our results point to losses from retrenchment, one should be cautious in drawing conclusions for other settings. A key factor is likely to be the level of unemployment (notably amongst the poor) at the time the program is cut. If one cuts disbursements at a time of sufficiently rapid economic recovery, or in regions where recovery is underway, then the loss to workers is likely to be smaller than we have found.

References

Ashenfelter, Orley, 1978, "Estimating the Effect of Training Programs on Earnings," *Review of Economic Studies* 60: 47-57.

Datt, Gaurav and Martin Ravallion, 1994, "Transfer Benefits from Public Works Employment", *Economic Journal*, 104: 1346-1369.

Dehejia, Rajeev H., and Sadek Wahba, 1999, "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94: 1053-1062.

Gruber, Jonathan, 1994, "The Incidence of Mandated Maternity Benefits," *American Economic Review*, 84(3): 622-641.

Heckman, J., H. Ichimura, and P. Todd, 1997, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies* 64(4): 605-654.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1998, "Characterizing Selection Bias using Experimental Data", *Econometrica*, 66: 1017-1099.

Heckman, J., R. Lalonde and J. Smith, 1999, "The Economics and Econometrics of Training Programs", In *Handbook of Labor Economics*, volumes 3 and 4 (O. Ashenfelter and D. Card eds.), forthcoming. Amsterdam: North-Holland.

Heckman, James and Jeffrey Smith, 1998, "The Pre-Program Earning Dip and the Determinants of Participation in a Social Program: Implication for Simple Program Evaluation Strategies", University of Chicago and University of Western Ontario, mimeo.

Jalan, Jyotsna and Martin Ravallion, 1998, "Are There Dynamic Gains from a Poor-Area Development Program?", *Journal of Public Economics*, Vol.67, No.1, Jan., 1998, pp. 65-86.

Jalan, Jyotsna and Martin Ravallion, 1999, "Income Gains to the Poor from Workfare: Estimates for Argentina's Trabajar Program," Policy Research Working Paper 2149, World Bank.

Lalonde, R., 1986, "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review* 76: 604-620.

Ravallion, Martin, 2000, "Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved," *World Bank Economic Review*, 14(2): 331-346.

Rosenbaum, Paul R., and Donald B. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70: 41-55.

____________ and _______, 1985, "Constructing a Control Group using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician* 39: 35-39.

Rubin, Donald B., "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association* 74: 318-328.

Rubin, Donald B., and Neal Thomas (2000), "Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates," *Journal of the American Statistical Association* 95: 573-585.

Smith, Jeffrey and Petra Todd, 2000, "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" mimeo, University of Western Ontario and University of Pennsylvania.

Table 1: Trabajar participation rates across survey rounds

	May 1999 (baseline survey)	October 1999 (first follow-up survey)	May 2000 (second follow-up survey)
Total interviewed	1459	1332	1291
Participants	1459	632	212
(% of total interviewed)	(100%)	(47.4%)	(16%)
Chaco	504	149 (34%)	17 (4%)
Mendoza	474	285 (63%)	146 (32%)
Tucuman	481	198 (44%)	49 (11%)
% non-participant who are employed		49.0%	60.5%

Source: Authors' calculations from the Trabajar sample.

Table 2: Unemployment rates in the selected provinces and nationally

% of the labor force unemployed (urban areas)	October 1998	May 1999 (baseline survey)	October 1999 (first follow-up survey)	May 2000 (second follow-up survey)
Chaco	11.3	9.5	12.4	10.4
Mendoza	5.7	7.6	6.8	9.8
Tucuman	14.9	19.2	15.9	19.9
All urban areas	12.4	14.5	13.8	15.4

Source: Trabajar Project Office, Ministry of Labor.

Table 3: Average wage rate for Trabajar projects in the selected provinces

Wage rate ($ per month)	May 1999 (baseline survey)	October 1999 (first follow-up survey)	May 2000 (second follow-up survey)
Chaco	194.9	183.8	165.5
Mendoza	200.0	192.8	168.9
Tucuman	195.9	195.4	165.5

Source: Trabajar Project Office, Ministry of Labor.

Table 4: Logit regressions for program participation

	Chaco				Mendoza				Tucuman			
	Propensity score 1st wave		*Dropout 2nd wave*		*Propensity score 1st wave*		*Dropout 2nd wave*		*Propensity score 1st wave*		*Dropout 2nd wave*	
	Coeff	Std error	Coeff	Std error	Coeff	Std error	Coeff	Std error	Coeff	Std error	Coeff	Std error
Common variables												
age 25_29	0.205	0.20	-0.444	0.38	0.024	0.23	0.285	0.41	0.584	0.21**	-0.410	0.42
age 30_39	-0.571	0.23**	-0.277	0.44	-0.638	0.24**	-0.427	0.48	-0.080	0.24	0.000	0.48
age 40_49	-1.117	0.28**	0.815	0.63	-0.649	0.27**	0.904	0.53*	0.220	0.26	0.134	0.57
age 50_54	-1.089	0.32**	0.577	0.60	-1.165	0.29**	0.502	0.56	-0.265	0.29	1.071	0.69
male	1.687	0.20**	-0.745	0.48	1.329	0.17**	0.832	0.50*	0.838	0.18**	0.553	0.42
head of the household	-0.089	0.27	0.079	0.51	0.373	0.29	-0.124	0.49	-0.229	0.28	-0.463	0.54
spouse of the head	-0.438	0.36	-1.405	0.76*	-0.170	0.36	0.807	0.99	-0.438	0.34	0.068	0.80
married	0.352	0.22	0.290	0.38	-0.277	0.23	-0.112	0.41	-0.256	0.22	0.057	0.43
PS not completed	-0.524	0.30*	0.735	0.70	1.000	0.28**	0.263	0.77	0.024	0.26	0.393	0.62
PS completed	-0.537	0.27**	0.526	0.65	0.741	0.25**	1.005	0.71	-0.238	0.22	0.507	0.51
SS not completed	-1.158	0.27**	0.505	0.61	-0.232	0.26	0.928	0.68	-0.795	0.22**	1.012	0.48**
SS completed	-0.955	0.30**	0.590	0.69	-0.636	0.31**	1.080	0.80	-0.677	0.28**	0.720	0.62
house is a villa	-0.322	0.26	-2.310	0.81**	0.468	0.28*	-1.658	1.36	0.442	0.38	0.183	1.18
house is an apt.	-0.765	0.44*	-0.088	0.60	-0.535	0.23**	0.288	0.46	0.420	0.21**	0.592	0.48
1 room	1.717	0.30**			2.187	0.32**	0.046	0.70	1.570	0.34**	-1.347	0.70*
2 rooms	1.141	0.25**	0.532	0.52	1.657	0.29**	-0.389	0.60	1.634	0.26**	-0.465	0.60
3 rooms	0.315	0.22	0.381	0.45	1.137	0.28**	0.029	0.57	0.932	0.24**	-0.627	0.57
4 rooms	0.270	0.23	0.578	0.48	0.714	0.29**	0.110	0.61	0.337	0.26	0.174	0.56
bathroom in the hh	-0.170	0.26	0.110	0.70	1.115	0.28**	1.408	0.61**	1.285	0.31**	-0.262	0.62
own only land	-0.407	0.20**	-0.244	0.33	0.675	0.28**	0.077	1.17	-1.810	0.26**	0.084	0.50
renting	-1.415	0.28**	0.351	1.11	-1.089	0.24**	-0.723	0.55			-1.001	0.61
walls - de Mampostería	-1.183	0.32**	-2.043	1.16*	0.350	0.18*	-0.545	0.43	-1.891	0.19**	-0.563	0.44
fraction non-migrants	1.390	0.24**	1.289	0.55**	-0.163	0.22	-0.087	0.51	0.465	0.25*	-1.472	0.66**
extended family	0.411	0.18**	-0.375	0.37	0.305	0.18*	-0.158	0.36	-0.060	0.19	0.082	0.44
fraction children 6-12 attending school	-1.685	0.40**	0.143	1.02	-0.807	0.46**	0.535	0.97	0.632	0.43	-0.365	0.98

fraction children 13-18 attending school	-0.330	0.18*	0.222	0.36	-0.227	0.19	-0.239	0.35	-0.423	0.18**	0.000	0.40
fraction members 0-5	-5.579	1.12**	0.490	1.73	-1.687	0.87*	0.760	1.97	-6.302	0.99**	2.318	1.86
fraction members 6-14	-3.019	1.13**	-1.849	1.85	-0.975	0.95	0.569	2.18	-6.662	0.99**	2.110	1.96
fraction members 15-64	-3.878	1.01**	-0.748	1.47	-1.850	0.78**	0.387	1.70	-5.194	0.88**	1.420	1.61
household size	0.166	0.04**	-0.018	0.08	0.032	0.04	-0.100	0.08	0.113	0.04**	-0.091	0.08
constant	2.314	1.14**	2.567	2.08	-2.750	0.91*	-3.536	2.08	3.020	0.96	1.298	1.97
<u>Extra variables for Trabajar participants</u>												
participated to neighborhood associations			-0.714	0.46			-0.686	0.40*			-0.245	0.43
entered in Trabajar due to personal contacts w/:												
- municipality officials			-0.939	0.48**			-0.006	0.34			1.298	0.62**
- union leaders			0.059	0.47			0.692	0.39*			-0.327	0.42
- former Trabajar workers			-0.800	0.46*			-0.502	0.45			-1.055	0.51**
- dirigentes barriales/others			0.154	0.41			0.280	0.38			-1.047	0.47**
previously employed as:												
- temporary worker			0.391	0.38			0.849	0.49*			-0.254	0.44
- permanent worker			-0.448	0.36			0.440	0.52			-0.504	0.48
Number of obs	2023		359		2615		352		1827		302	
log likelihood	-824.8		-197.0		-952.1		-197.6		-795.8		-175.4	
pseudo R2	0.268		0.155		0.221		0.128		0.238		0.157	
F-test joint significance basic specification (p-value)			*0.064*				*0.284*				*0.253*	
F-test joint significance new variables (p-value)			*0.008*				*0.044*				*0.002*	

Note: (1) 1st stage matching of participants with non-participants using Trabajar and PHS samples; (2) 2nd stage matching of leavers and stayers using Trabajar sample. Robust standard errors in parentheses; * significant at 10%; ** significant at 5%.

Table 5: Triple-difference estimates of income replacement

	Stayers in round 2 ($D_{i2}=1$) N=419			Matched leavers in round 2 ($D_{i2}=0$) N=400		
	Trabajar Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single Difference	Trabajar Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single difference
	$\overline{Y}_t^T =$	$\overline{Y}_t^C =$		$\overline{Y}_t^T =$	$\overline{Y}_t^C =$	
t=1	228.9	282.7	-53.8	223.6	294.4	-70.8
	(3.8)	(13.2)	(14.0)	(2.9)	(12.6)	(12.9)
t=2	228.4	277.3	-48.8	83.0	288.8	-205.8
	(4.1)	(13.3)	(13.9)	(6.2)	(12.0)	(13.9)
Single difference	$\overline{\Delta Y}_t^T =$	$\overline{\Delta Y}_t^C =$		$\overline{\Delta Y}_t^T =$	$\overline{\Delta Y}_t^C =$	
	-0.5	-5.4		-140.7	-5.6	
	(4.5)	(7.4)		(6.2)	(8.1)	
Double difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_{i2}=1] =$			$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_{i2}=0] =$		
	4.9			-135.1		
	(8.3)			(10.6)		
Triple difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_2=1] - [\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_2=0] =$					
	140.0					
	(13.4)					

Note: Standard errors in parentheses.

Table 6: Triple-difference estimates of income replacement with only first-stage matching

	Stayers in round 2 ($D_{i2}=1$) *N*=498			Leavers in round 2 ($D_{i2}=0$) *N*=520		
	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single Difference	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single difference
	$\overline{Y}_t^T=$	$\overline{Y}_t^C=$		$\overline{Y}_t^T=$	$\overline{Y}_t^C=$	
t=1	228.1 (3.4)	286.7 (12.4)	-58.4 (13.0)	225.4 (2.9)	286.1 (10.6)	-60.6 (10.9)
t=2	228.2 (3.7)	278.4 (12.0)	-50.2 (12.5)	83.4 (5.6)	281.6 (9.9)	-198.2 (11.8)
Single difference	$\overline{\Delta Y}_t^T=$ 0.10 (4.7)	$\overline{\Delta Y}_t^C=$ -8.2 (7.2)		$\overline{\Delta Y}_t^T=$ -142.0 (5.5)	$\overline{\Delta Y}_t^C=$ -4.4 (6.8)	
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ 8.3 (7.9)			$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -137.6 (9.1)		
Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 145.9 (12.1)					

Note: Standard errors in parentheses.

Table 7: Disaggregation by province

	Stayers in round 2 ($D_{i2}=1$)		Leavers in round 2 ($D_{i2}=0$)	
	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)
Chaco	*N=128*		*N=233*	
Single difference	$\overline{\Delta Y}_t^T=$ -26.7 (8.2)	$\overline{\Delta Y}_t^C=$ -12.6 (7.2)	$\overline{\Delta Y}_t^T=$ -167.6 (6.3)	$\overline{\Delta Y}_t^C=$ -3.8 (6.8)
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ -14.1 (11.1)		$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -171.5 (10.6)	
Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 157.5 (16.6)			
Mendoza	*N=231*		*N=122*	
Single difference	$\overline{\Delta Y}_t^T=$ 7.4 (5.9)	$\overline{\Delta Y}_t^C=$ -8.7 (13.2)	$\overline{\Delta Y}_t^T=$ -112.5 (14.7)	$\overline{\Delta Y}_t^C=$ -27.2 (20.2)
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ 16.1 (14.1)		$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -85.2 (10.6)	
Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 101.3 (27.0)			
Tucuman	*N=139*		*N=165*	
Single difference	$\overline{\Delta Y}_t^T=$ 12.7 (6.7)	$\overline{\Delta Y}_t^C=$ -3.4 (10.8)	$\overline{\Delta Y}_t^T=$ -127.6 (9.8)	$\overline{\Delta Y}_t^C=$ -0.8 (10.0)
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ 16.1 (12.8)		$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -128.4 (14.4)	
Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 144.5 (19.6)			

Table 8: Household income effects

	Stayers in round 2 ($D_{i2}=1$) *N=498*			Leavers in round 2 ($D_{i2}=0$) *N =520*		
	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single difference	Participants in round 1 ($D_{i1}=1$)	Matched non-participants round 1 ($D_{i1}=0$)	Single difference
	$\overline{Y}_t^T =$	$\overline{Y}_t^C =$		$\overline{Y}_t^T =$	$\overline{Y}_t^C =$	
t=1 Total hh income	537.2	706.4	-169.2	536.4	691.8	-155.4
	(15.8)	(11.0)	(23.6)	(16.8)	(18.4)	(23.9)
Trabajar workers' income	261.0			257.5		
	(5.6)			(5.7)		
Earnings other members	236.1	630.8		233.6	602.3	
(13.7)	(13.7)	(17.3)		(15.2)	(15.7)	
Other income	40.1	75.6	-35.5	45.3	89.5	-44.2
	(6.2)	(6.7	(8.8	(5.7)	(7.5)	(9.5)
t=2 Total hh income	507.3	689.6	-182.3	354.2	670.2	-315.9
	(18.4)	(13.3)	(22.8)	(14.9)	(17.0)	(22.9)
Trabajar workers' income	232.5			83.5		
	(4.0)			(5.6)		
Earnings other members	232.9	614.4		225.3	583.7	
	(13.3)	(16.5)		(13.1)	(15.5)	
Other income	41.9	75.2	-33.3	43.7	86.4	-42.7
	(6.3)	(5.4)	(8.4)	(4.9)	(5.3)	(7.1)
Single difference	$\overline{\Delta Y}_2^T =$	$\overline{\Delta Y}_2^C =$		$\overline{\Delta Y}_2^T =$	$\overline{\Delta Y}_2^C =$	
Total hh income	-29.8	-16.8		-182.2	-21.5	
	(11.4)	(12.3)		(14.6)	(12.6)	
Trabajar workers' income	-28.5			-172.4		
	(5.7)			(7.5)		
Earnings other members	-3.2	-16.4		-8.2	-18.5	
	(9.6)	(11.0)		(11.5)	(10.9)	
Other income	1.8	-0.3		-1.6	-3.1	
	(3.8)	(5.1)		(4.9)	(6.2)	
Double difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \| D_2 = 1] =$			$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \| D_2 = 0] =$		
Total hh income	-13.1			-160.5		
	(16.2)			(19.7)		
Trabajar workers' income	-28.5			-172.4		
	(5.7)			(7.5)		
Earnings other members	13.2			10.3		
	(14.0)			(16.2)		
Other income	2.2			1.5		
	(6.2)			(8.0)		
Triple difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \| D_2 = 1] - [\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \| D_2 = 0] =$					
Total hh income	147.4 (25.6)					
Trabajar workers' income	144.8 (9.5)					
Earnings other members	2.9 (21.5)					
Other income	1.8 (5.1)					

Table 9: Triple-difference estimates of income replacement over 12 months (two follow-up rounds)

	Stayers in rounds 2 and 3 ($D_{i2}=1; D_{i3}=1$) N=118			Leavers in rounds 2 and 3 ($D_{i2}=0; D_{i3}=0$) N=424		
	Treatment	Comparison	Single difference	Treatment	Comparison	Single difference
	$\overline{Y}_t^T =$	$\overline{Y}_t^C =$		$\overline{Y}_t^T =$	$\overline{Y}_t^C =$	
T=1	228.0	285.7	-57.7	227.5	272.2	-44.6
	(8.1)	(27.9)	(29.6)	(3.4)	(13.7)	(13.8)
T=2	241.3	314.9	-73.7	85.5	276.4	-190.9
	(8.4)	(32.9)	(33.0)	(6.2)	(12.3)	(14.1)
T=3	219.5	289.0	-69.5	138.4	267.3	-128.9
	(6.8)	(25.3)	(25.8)	(7.0)	(18.6)	(14.5)
Single difference	$\overline{\Delta Y}_2^T =$	$\overline{\Delta Y}_2^C =$		$\overline{\Delta Y}_2^T =$	$\overline{\Delta Y}_2^C =$	
	13.3	29.3		-142.0	4.2	
	(9.2)	(20.8)		(6.2)	(9.3)	
	$\overline{\Delta Y}_3^T =$	$\overline{\Delta Y}_3^C =$		$\overline{\Delta Y}_3^T =$	$\overline{\Delta Y}_3^C =$	
	-21.7	-25.9		52.9	-9.0	
	(9.1)	(18.0)		(7.4)	(16.5)	
Double difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_{i2}=1, D_{i3}=1] =$			$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_{i2}=0, D_{i3}=0] =$		
	-16.0			-146.3		
	(22.0)			(11.9)		
	$[\Delta(\overline{Y}_3^T - \overline{Y}_3^C) \mid D_{i2}=1, D_{i3}=1] =$			$[\Delta(\overline{Y}_3^T - \overline{Y}_3^C) \mid D_{i2}=0, D_{i3}=0] =$		
	4.2			62.0		
	(20.0)			(16.5)		
Cumulative gain	-11.8			-84.3		
	(22.4)			(15.7)		
Triple difference	$[\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_2=1, D_3=1] - [\Delta(\overline{Y}_2^T - \overline{Y}_2^C) \mid D_2=0, D_3=0] =$					
	130.3					
	(25.3)					
	$[\Delta(\overline{Y}_3^T - \overline{Y}_3^C) \mid D_2=1, D_3=1] - [\Delta(\overline{Y}_3^T - \overline{Y}_3^C) \mid D_2=0, D_3=0] =$					
	-57.8					
	(33.1)					
Cumulative gain	72.5					
	(32.0)					

Note: Standard errors in parentheses; both matched comparison groups and matched leavers (as in Table 3).

Table 10: Perceived gains from past participation from the program

Length of exposure to the program:	One round	Two rounds	Three rounds
	$D_{i2}=1; D_{i2}=0; D_{i3}=0$	$D_{i2}=1; D_{i2}=1; D_{i3}=0$	$D_{i2}=1; D_{i2}=1; D_{i3}=1$
N=962	*N* = 464 (48.2%)	*N*= 339 (35.2%)	*N*= 136 (14.1%)
	% of respondents replying "yes"		
Expanded job opportunities:			
Expected gains in t=2	35.7	52.6	51.2
Expected gains in t=3 all	33.7	37.9	48.7
Employed in t=3	35.4	42.7	
Unemployed in t=3	31.1	31.1	
Learned skills for other jobs:			
Expected gains in t=2	51.8	66.1	64.7
Expected gains in t=3 all	45.0	53.5	61.5
Employed in t=3	38.6	51.0	70.0
Unemployed in t=3	55.0	57.1	
Expanded contacts for future:			
Expected gains in t=2	25.7	35.9	40.0
Expected gains in t=3 all	21.9	22.9	32.6
Employed in t=3	24.5	22.6	30.0
Unemployed in t=3	17.8	23.3	

Note: Sample of Trabajar workers - special module in period 3

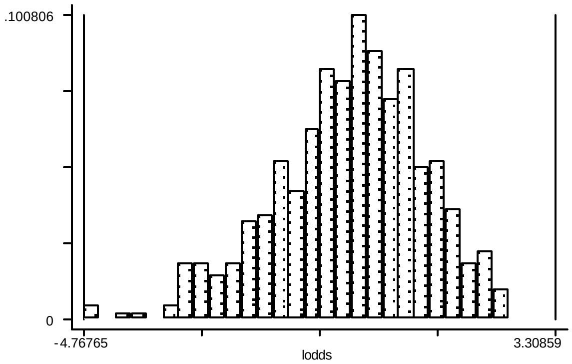
Table 11: Joint test of no lagged gains from past participation and no selection bias: Triple-difference estimates of income replacement over 6 months

Stayers round 3 ($D_{i3}=1$)				
	Participants round 2 ($D_{i2}=1$)		Non-participants round 2 ($D_{i2}=0$)	
	N=118		N=19	
Single difference	$\overline{\Delta Y}_t^T=$ 13.3 (9.2)	$\overline{\Delta Y}_t^C=$ 29.3 (20.9)	$\overline{\Delta Y}_t^T=$ -151.7 (22.8)	$\overline{\Delta Y}_t^C=$ 54.6 (23.0)
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ -16.0 (22.1)		$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -206.3 (34.3)	
(A) Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 190.3 (56.8)			
Leavers round 3 ($D_{i3}=0$)				
	Participants round 2 ($D_{i2}=1$)		Non-participants round 2 ($D_{i2}=0$)	
	N=292		N = 424	
Single difference	$\overline{\Delta Y}_t^T=$ -3.2 (4.8)	$\overline{\Delta Y}_t^C=$ -4.8 (11.2)	$\overline{\Delta Y}_t^T=$ -142.1 (6.2)	$\overline{\Delta Y}_t^C=$ 4.2 (7.6)
Double difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=1]=$ 1.6 (11.7)		$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_{i2}=0]=$ -146.3 (11.9)	
(B) Triple difference	$[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=1]-[\Delta(\overline{Y}_2^T-\overline{Y}_2^C)\vert D_2=0]=$ 147.9 (17.3)			
t-test of equality	H_0: (A)=(B) Difference: 42.4 (59.4) t-statistic: 0.71 p-value: 0.47			

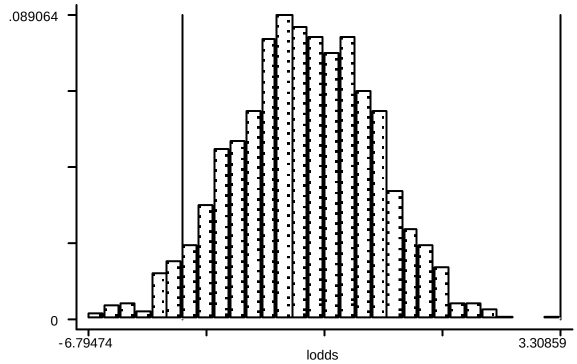
Note: Standard errors in parentheses. Under the joint null of 'no lagged gains from past participation' and 'no selection bias', the triple differences should be the same (see equation (7))

Figure 1: Log of odds ratio – matching first wave

Chaco

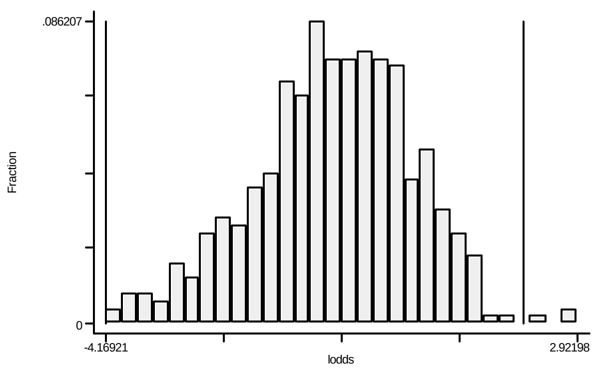


Trabajar sample

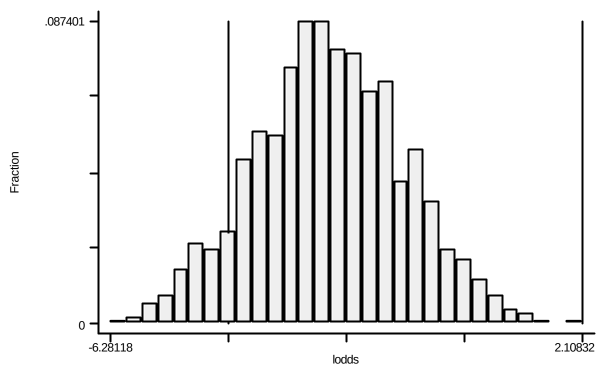


PHS sample

Mendoza

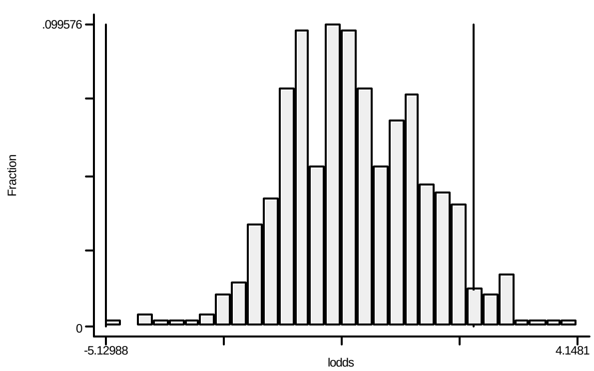


Trabajar sample

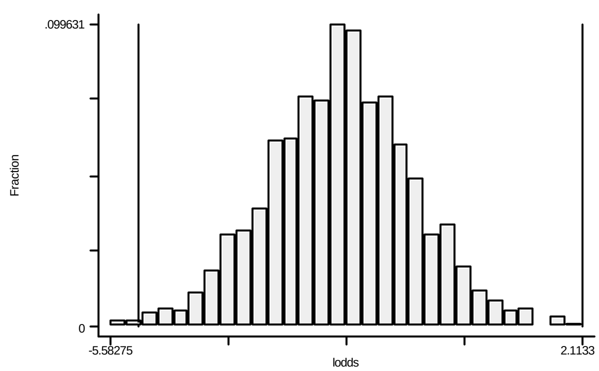


PHS sample

Tucuman



Trabajar sample



PHS sample

Note: The vertical bars delimit the region of common support