

Annex 1.9: The Impact of Mexico's Retraining Program on Employment and Wages (PROBECAT)

I. Introduction

This case is somewhat unusual in that three evaluations of the program have been carried out—first, by the World Bank using data from 1992 (Revenge, Riboud, and Tan 1994); second, by the Mexican Ministry of labor using data from 1994 (STPS 1995); and third, an update by the World Bank (Wodon and Minowa 1999). The methodologies used for the first two evaluations were quite similar, and they gave similar results. Methodological enhancements in the third evaluation led to fairly different findings and policy conclusions. The fact that the results differ substantially between the first two evaluations and the third highlights the importance of the methodology and data used, and caution in interpreting results when carrying out program evaluations.

Project Description. PROBECAT (Programa de Becas de Capacitacion para Trabajadores) is a Mexican short-term training program targeted at increasing earnings and employment for unemployed and displaced workers. PROBECAT is administered through the state employment offices. Trainees receive minimum wage during the training period, which lasts from one to six months, and the local employment office provides placement. Originally, the program was small (50,000 or so participants), but in recent years it has grown dramatically, to cover more than 500,000 persons per year.

Highlights of the Evaluations. The highlights are as follows:

- The 1994 evaluation is interesting for four reasons: the imaginative use of existing data; the construction of a matched-comparison group; the explicit recognition of the multifaceted nature of the intervention outcomes, particularly for heterogeneous groups of workers; and the explicit cost-benefit analysis. The findings of the evaluation were quite positive in terms of the impact of the program on beneficiaries.
- The 1995 evaluation is a replication of the methodology of the 1994 evaluation on a more recent data set. The findings are also favorable for the impact of the program. Because the design and findings of the 1995 evaluation match those of the 1994 evaluation, the 1995 evaluation will not be discussed below.

- The 1999 evaluation was carried out as part of the Mexico poverty assessment with the data set used for the 1995 evaluation but with a different econometric methodology. The controls used for the endogeneity of program participation showed a vanishing of the impact of the program on the probability of working and on wages after training. Although this does not imply that the program has no benefit, it suggests that it works more as a temporary safety net for the unemployed than as a job training program.

II. Research Questions and Evaluation Design

In the 1994 evaluation, the authors estimate the impact of training on (a) the probability of employment after 3, 6, and 12 months; (b) the time to exit unemployment; (c) the effect on monthly earnings, work hours per week, and hourly wages; and (d) the return on investment.

The 1999 evaluation looks at the same questions except work hours per week and hourly wages. Given that there is no impact in that evaluation on employment and monthly earnings, the return is zero, but again the program may work as a safety net.

The design of both evaluations is innovative in constructing the comparison group. In both cases, the evaluations combine an existing panel labor force survey, Encuesta Nacional de Empleo (ENEU), with a panel of trainees for the same period. That is, the program's selection criteria are used to define the control group from the ENEU. Although there is no alternative to this combination of surveys because of data limitations, the construction of the joint sample (control and treatment groups) can be critiqued, as discussed in the 1999 evaluation:

- In using the unemployed individuals in the ENEU to form the control group, it is assumed that none of the ENEU individuals have benefited from the program. This is not the case because every individual in the ENEU has some probability of having participated in PROBECAT. Fortunately, given that the program was small until 1993, only a very small minority of the individuals in the control group are likely to have participated in the program (the data for the 1999 evaluation are for 1993–94);
- The combination of two random samples (PROBECAT trainees and ENEU unemployed individuals) is not a random sample, so that in the absence of the standard properties for the residuals, the results of regressions may not yield consistent parameter estimates, especially because the models used are sensitive to the assumption of bivariate normality. In the absence of better data, not much can be done on this.

The main differences between the 1994 and 1999 evaluations are as follows;

- In the 1994 evaluation, the authors attempt to address the selection bias problems resulting from PROBECAT's nonrandom selection of trainees by estimating a probit model of the probability of participation. The comparison group is then limited to those individuals who are highly likely to participate. In the 1999 evaluation, the authors argue that this method does not eliminate the problem of endogeneity. Instead, they use an instrumental variable to control for the endogeneity of program participation.
- In the estimation of earnings in the 1994 evaluation, while participation in PROBECAT is controlled for, the sample selection bias resulting from the decision to work is not accounted for. In the 1999 study, both sample selection problems are accounted for.

III. Data

In the 1994 evaluation, data on trainees are gathered from a 1992 retrospective survey administered to 881 men and 845 women who were trained in 1990. This is supplemented with panel data on 371 men and 189 women derived from a household survey of the 16 main urban areas in Mexico. This survey was part of a regular quarterly labor force survey, ENEU, undertaken by the Mexican statistical agency. The authors exploited the rotation group structure of the survey to take workers who were unemployed in the third quarter of 1990 and then tracked those workers for a year. This was supplemented by a cohort that became unemployed in the fourth quarter of 1990 and was tracked for nine months. The same method was used in the 1999 evaluation, but for more recent data.

IV. Econometric Techniques

The key econometric techniques used are survival analysis (duration models) for the probability of working and Heckman regressions for wages. What follows is based on the 1999 evaluation. Differences with the 1994 evaluation are highlighted.

Impact of PROBECAT on the Length of Employment Search. In the survival analysis, the survivor function $S(t)$ represents the length of unemployment after training (measured in months). Given $S(t)$, the hazard function $\lambda(t)$ denoting the chance of becoming employed (or the risk of remaining unemployed) at time t among the individuals who are not yet employed at that time is $\lambda(t) = -d(\log S(t))/dt$. The survivor curve can be

specified as a function of program participation P , individual characteristics X , and state characteristics Z , so that $\lambda = \lambda(t; X, Z, P)$. In Cox's proportional hazard model, if i denotes a household and j denotes the area in which the household lives, we have

$$\lambda(t; X, Z, P1, P2) = \lambda_0(t) \exp(\gamma'X_{ij} + \delta'Z_j + \mu P_{ij}). \quad (1)$$

Cox proposed a partial maximum likelihood estimation of this model in which the baseline function $\lambda_0(t)$ does not need to be specified. If μ is positive and statistically significant, the program has a positive effect on employment. In a stylized way, the difference between the 1994 and 1996 evaluations can be described as follows:

- In the 1994 evaluation, the authors run a probit on program participation and delete from the control group those individuals with a low probability of participating in the program. They then run equation (1) without further control for endogeneity.
- In the 1999 evaluation, the authors also run a probit on program participation, but they use program availability at the local level (obtained from administrative data) as an additional determinant of participation (but not of outcome conditional on individual participation.) Then they run equation (1), not with the actual value of the participation variable but with the predicted (index) value obtained from the first stage probit. This is an instrumental variable procedure. The idea follows work on program evaluation using decentralization properties by Ravallion and Wodon (2000) and Cord and Wodon (1999). The authors compare their results with other methods, showing that other methods exhibit a bias in the value of the parameter estimates owing to insufficient control for endogeneity.

Impact of PROBECAT on Monthly Earnings. To carry out this analysis, a model with controls for sample selection in labor force and program participation is used in the 1999 evaluation (the 1994 evaluation controls only for program participation). Denote by $\log w$ the logarithm of the expected wage for an individual. This wage is nonzero if and only if it is larger than the individual's reservation wage (otherwise, the individual would choose not to work). Denote the unobserved difference between the individual's expected wage and his or her reservation wage by Δ^* . The individual's expected wage is determined by a number of individual (vector E , consisting essentially of the individual's education and past experience) and geographic variables Z , plus program participation P . The difference between the individual's expected wage and his or her reservation wage is determined by the same variables, plus the number

of children, the fact of being a household head, and the fact of being married, captured by D . The model is thus

$$\Delta_{ij}^* = \phi_{\Delta}'E_{ij} + \pi_{\Delta}'D_{ij} + \eta_{\Delta}'Z_j + \alpha_{\Delta}P_{ij} + v_{ij} \text{ with } \Delta_{ij} = 1 \text{ if } \Delta_{ij}^* > 0, \text{ and } 0 \text{ if } \Delta_{ij}^* < 0 \quad (2)$$

$$\text{Log } w_{ij}^* = \phi_w'E_{ij} + \eta_w'Z_j + \alpha_w P + \kappa_{ij} \text{ with } \text{Log } w = \text{log } w^* \text{ if } \Delta = 1 \text{ and } 0 \text{ if } \Delta = 0. \quad (3)$$

As for the survival model, in order to control for endogeneity of program participation, in the 1999 evaluation a probit for program participation is first estimated by using program availability at the local level as a determinant of individual participation. Then the above equations are estimated by using the predicted (index) value of program participation instead of its true value. In the 1994 evaluation, the model does not control for the decision to participate in the labor market given in equation (2) above. This equation is replaced by the program participation probit estimated without local availability of the program as an independent variable. Again, comparisons of various models show that bias is present when the instrumental variable technique is not used.

V. Who Carried It Out

The 1994 evaluation was conducted by Ana Revenga in the Latin America and Caribbean Country Department II of the World Bank, Michelle Riboud in the Europe and Central Asia Country Department IV of the World Bank, and Hong Tan in the Private Sector Development Department of the World Bank. The 1999 evaluation was carried out by Quentin Wodon and Mari Minowa, also at the World Bank (Latin America region).

VI. Results

The results obtained in the various evaluations are very different. The 1994 and 1995 evaluations find positive impacts of the program on employment and wages. No positive impact was found in the 1999 evaluation, which is based on the same data used for the 1995 evaluation. In terms of cost-benefit analysis, the first two evaluations are favorable but the last evaluation is not. The disappointing results in the last evaluation are not surprising. Most retraining programs in Organisation for Economic Co-operation and Development countries have been found to have limited impacts, and when programs have been found to have some impact, this impact tends to vanish after a few years (Dar and Gill 1998).

The fact that PROBECAT may not be beneficial in the medium to long run for participants according to the last evaluation does not mean that it should be suppressed. The program could be viewed as providing temporary safety nets (through the minimum wage stipend) rather than training. Or it could be improved so as to provide training with longer-lasting effects.

VII. Lessons Learned

Apart from some of the innovative features of these evaluations and their limits, the key lesson is that one should be very careful in doing program evaluations and using the results to recommend policy options. The fact that a subsequent evaluation may contradict a previous one with the use of different econometric techniques should always be kept in mind. There have been many such cases in the literature.

VIII. Sources

Revenge, Ana, Michelle Riboud, and Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages." *World Bank Economic Review* 8 (2): 247-77.

Wodon, Quentin, and Mari Minowa. "Training for the Urban Unemployed: A Reevaluation of Mexico's PROBECAT." World Bank, Government Programs and Poverty in Mexico, Report No. 19214-ME, Vol. II.