Using Human Capital Enrichment to Reduce Poverty and Inequality: the Case of *Oportunidades* in Mexico

Douglas McKee[1] and Petra Todd[2]

September 21, 2007

[1]Population Studies Center, University of Pennsylvania (mckeedm@pop.upenn.edu)
[2]Department of Economics, University of Pennsylvania (ptodd@econ.upenn.edu)

Abstract

Recent evaluations of the *Oportunidades* schooling and health subsidy program in Mexico have demonstrated statistically significant positive impacts on schooling and health outcomes. This paper adapts methods developed in Dinardo, Fortin and Lemieux (1996) for use in studying how these schooling and health impacts will affect the future earnings distributions of cohorts recently exposed to the program. The method nonparametrically simulates earnings distributions, with and without the program, and quantifies resulting changes in mean earnings and in earnings inequality. It is well recognized that the *Oportunidades* program has reduced poverty and inequality of the current generation through its targeted cash transfers. This paper finds that by enriching human capital, as measured by schooling and height, the program will also generate increases in future earnings. However, it will achieve only modest reductions in overall poverty and earnings inequality.

1 Introduction

In recent years, governments in many Latin American countries have adopted conditional cash transfer (CCT) programs as a strategy for alleviating poverty and stimulating investment in human capital. These programs typically provide cash grants to poor families if they send their age-eligible children to school as well as subsidies for regularly visiting health clinics for check-ups. CCT programs now exist in Argentina, Brazil, Chile, Colombia, Costa Rica, El Salvador, Equador, Honduras, Mexico, Nicaragua, Peru and Uruguay.[1]

The Mexican *Oportunidades* program (formerly called PROGRESA) has been rigorously evaluated using both experimental and nonexperimental evaluation designs. An experiment carried out in the first two years of its implementation (1998-1999) in rural areas demonstrated statistically significant impacts of the program on reducing child labor, improving health outcomes and increasing schooling enrollment and attainment.[2] A nonexperimental evaluation of the program in urban areas also found statistically significant impacts similar in magnitude to those found in rural areas. Today, the Mexican program provides payments to about one quarter of all families in Mexico that constitute on average 20% of those families' household income.

Previous studies of the *Oportunidades* program have documented significant impacts on education levels and on health outcomes. These impacts are estimated using comparisons between individuals who participate in the program with those who do not participate. This paper takes as a point of departure the observed shorter term impacts on education and health and uses them to evaluate the likely long-term effects of the program on the earnings distributions of these children when they become adults. *Oportunidades* has clearly reduced the poverty and inequality of the current generation of parents through its cash transfers. The question examined in this paper is how the program's impacts on human capital, as measured by years of schooling and height (as a proxy for health), will likely affect earnings inequality and poverty in the next generation.

Our approach to forecasting program impacts on earnings distributions adapts a nonparametric

[1]Programs with similar features also exist in some Asian countries, such as Bangladesh and Pakistan.

[2]See, e.g., Schultz (2000,2004), Gertler (2000), Behrman, Sengupta and Todd (2005), Parker and Skoufias (2000), Buddelmeyer and Skoufias (2003), and Todd and Wolpin (2006).

decomposition method, originally proposed in Dinardo, Fortin and Lemieux (1996), for use in simulating how the program's impacts on education and height will affect subsequent earnings, poverty and inequality of cohorts that have recently been exposed to the program. The simulation method is nonparametric in that it does not impose any functional form assumptions on the earnings-height-education-work experience relationship, other than continuity and differentiability. The flexibility with regard to model specification is important, given the evidence for nonlinearities in the relationship. We use the nonparametric simulation method to compare the earnings and employment distributions with and without the program. We also compare the results to inferences obtained using more standard parametric approaches.

The main findings from the simulations are that the program's impact on education and height generally increase mean future earnings of beneficiaries, but have little effect on future earnings inequality. A few different factors contribute to the modest overall observed impacts on inequality. First, the program targets children from poor family backgrounds, and family background is an imperfect predictor of future earnings. If program beneficiaries come from throughout the earnings distribution, then the program's effect on earnings inequality is ambiguous. Second, we find evidence for nonlinearities in how education and height affect earnings, the most notable being that the returns to education are greater for post-primary education. Such nonlinearities imply that people at the higher earnings deciles benefit more from the program.

Our empirical analysis is based on the first wave of the Mexican Family Life Survey (MXFLS-1) which was collected in 2002. The survey collected data for all members of 8,440 households and includes information about labor force participation, income for both primary and secondary jobs, education, and health. It also contains measures of family background, that we use to simulate program targeting. Our final analyses use a subsample of 5,171 individuals age 25 to 40 for which there are sampling weights and for which the required variables are reported.

This paper develops as follows. Section two describes the nonparametric simulation method that we use to study how the program affects labor force participation and the overall earnings distribution. Section three describes the Mexican Family Life Survey and our analysis samples.

Section four presents the empirical results. Section five concludes and discusses some limitations of the analysis as well as avenues for future research.

2 Methodology for Simulating Program Effects on Population Earnings Distributions

Dinardo, Fortin and Lemieux (1996) develop a semiparametric decomposition procedure to investigate the effects of institutional and labor market factors on changes in the U.S. wage distribution over time. Their approach writes the overall wage density at time t, $f_w(w|t)$, in terms of the conditional wage densities, where conditioning is on a set of labor market or institutional factors, z, whose effects on earnings they analyze:

$$f_w(w|t) = \int_z f_w(w|z,t) f_z(z|t) dz.$$

In their study, z includes variables indicating union status, industrial sector, and whether the wage falls above or below the minimum wage. Counterfactual wage densities are constructed by replacing $f_z(z|t)$ by a different hypothetical conditional density, $g_z(z|t)$.

We apply the Dinardo et. al. (1996) method to simulate earnings densities with and without a program intervention. We extend the method to account for simultaneous analysis of both labor force participation and earnings by allowing the earnings distribution to have a mass point at zero due to nonparticipation. In this section, we first describe our simulation approach in general terms, and then how it applies to our particular program evaluation.

2.1 Basic method

Denote some outcome of interest (earnings) by y and define the distribution of y in terms of its distribution conditional on some observed characteristics x and the unconditional distribution of x:

$$f(y) = \int_x f(y,x) dx = \int_x f(y|x) f(x) dx.$$

Now suppose that the program intervention changes the distribution of x from $f(x)$ to $\tilde{f}(x)$ but that the distribution of y conditional on x stays the same ($\tilde{f}(y|x) = f(y|x)$). The new unconditional distribution of y would be given by:

$$\tilde{f}(y) = \int_x f(y|x)\tilde{f}(x)dx.$$

We wish to simulate the effect of the program intervention on the outcome y as it operates through changing x. For example, suppose that the variable x represents years of schooling and height and that the program intervention increases schooling and height by some amount, i.e. $\tilde{x} = x + \Delta_x$. Suppose we have a set of n independent draws from the unconditional density, $f(x)$. If we know Δ_x we can generate for each individual $\tilde{x}_i = x_i + \Delta_{x_i}$. We can simulate the post-program earnings density $\tilde{f}(y)$ at a point y_0 by the average:

$$\begin{aligned}\hat{\tilde{f}}(y_0) &= \frac{1}{n}\sum_{x_i \in X} \hat{f}(y_0|\tilde{x}_i = x_i + \Delta_{x_i}) \\ &= \frac{1}{n}\sum_{\tilde{x}_i \in \tilde{X}} \frac{\hat{f}(y_0, \tilde{x}_i)}{\hat{f}(\tilde{x}_i)},\end{aligned}$$

where $\hat{f}(y, x_i)$ and $\hat{f}(x_i)$ are nonparametric estimators of the unconditional densities:

$$\begin{aligned}\hat{f}(y_0, x_i) &= \frac{1}{\alpha_y \alpha_x n}\sum_{j=1}^{n} K(\frac{y_j - y_0}{\alpha_y})K(\frac{x_j - x_i}{\alpha_x}) \\ \hat{f}(x_i) &= \frac{1}{\alpha_x n}\sum_{j=1}^{n} K(\frac{x_j - x_i}{\alpha_x}).\end{aligned}$$

α_y and α_x are bandwidths that are assumed to satisfy the usual requirements for consistent kernel density estimation.[3]

The MXFLS data are a stratified sample, and sampling weights are required to reweight to population proportions. Incorporating sampling weights into the simulation method is straightforward. Assume each observation has a sampling weight, ω_i, and that the weights are scaled so that $\sum \omega_i = n$. The weights can be incorporated into the estimation of $\tilde{f}(y)$ as follows:

[3]For consistency, we require $a_x \to 0, a_y \to 0$ as $n \to \infty$ and $a_y a_x n \to \infty$.

$$\widehat{\tilde{f}}(y) = \frac{1}{n}\sum_{\tilde{x}_i \in \tilde{x}} \omega_i \frac{\hat{f}(y, \tilde{x}_i)}{\hat{f}(\tilde{x}_i)},$$

and also into the estimation of the unconditional kernel densities:

$$\begin{aligned} \hat{f}(y_0, x_i) &= \frac{1}{\alpha_y \alpha_x n} \sum_{j=1}^{n} \omega_j K(\frac{y_j - y_0}{\alpha_y}) K(\frac{x_j - x_i}{\alpha_x}) \\ \hat{f}(x_i) &= \frac{1}{\alpha_x n} \sum_{j=1}^{n} \omega_j K(\frac{x_j - x_i}{\alpha_x}). \end{aligned}$$

For expositional clarity, we suppress the weights in the remainder of the discussion, although they are included in the estimation.

2.2 Accounting for probability mass at zero

Kernel density estimation can approximate well the distributions of continuous random variables, but in our data we encounter the fact that many people (especially women) report zero earnings. The program intervention might increase earnings among workers as well as change the probability of having positive earnings. We accomodate the mass point at zero in the earnings distribution by estimating the density of earnings as a mixture, where with some probability individuals earn zero and with the remaining probability they earn income drawn from the density of income conditional on its being positive, $f_{|y>0}(y)$. Both the probability of having positive earnings and the magnitude of earnings are potentially affected by the program.

Let $\tilde{y}$ be the random variable representing the distribution of income implied by the counterfactual distribution of $\tilde{x}$. Again, we assume the distribution of y conditional on x stays constant (e.g. that the relationship between earnings and education and health is stable):

$$\begin{aligned} \Pr(\tilde{y} = 0|x) &= \Pr(y = 0|x) \\ \tilde{f}_{|\tilde{y}>0}(y|x) &= f_{|y>0}(y|x) \end{aligned}$$

We can obtain the probability of zero earnings, $\Pr(\tilde{y} = 0)$, with the program intervention using the following:

$$\begin{aligned}
\Pr(\tilde{y}=0) &= \int_x \Pr(y=0|x)\tilde{f}(x)dx, \\
&\approx \frac{1}{n}\sum_{x_i \in X} \widehat{\Pr}(y=0|\tilde{x}_i = x_i + \Delta_x),
\end{aligned}$$

where X is the support of x_i and where

$$\widehat{\Pr}(y=0|x_i) = \frac{\sum_{x_j \in X} 1(y_j=0) K(\frac{x_j - \tilde{x}_i}{\alpha_x})}{\sum_{x_j \in X} K(\frac{x_j - \tilde{x}_i}{\alpha_x})}.$$

In the last equation, $1(y_i = 0)$ is an indicator that denotes whether the individual has positive earnings.

Let $\tilde{f}_{\tilde{y}>0}(y)$ be the density of income conditional on its being positive. The counterfactual distribution of y conditional on y being positive is given by:

$$\begin{aligned}
\tilde{f}_{\tilde{y}>0}(y) &= \int_x f_{y>0}(y|x)\tilde{f}(x|\tilde{y}>0) \\
&= \int_x f_{y>0}(y|x)\frac{\Pr(y>0|x)}{\Pr(\tilde{y}>0)}\tilde{f}(x) \\
\hat{\tilde{f}}_{\tilde{y}>0}(y) &= \frac{1}{n}\sum_{\tilde{x}_i \in X} \hat{f}_{y>0}(y|x_i)\frac{\widehat{\Pr}(y>0|\tilde{x}_i = x_i + \Delta_x)}{\widehat{\Pr}(y>0)} \\
&= \frac{1}{n}\sum_{\tilde{x}_i \in X} \frac{\hat{f}_{y>0}(y, x_i)}{\hat{f}_{y>0}(x_i)}\frac{\widehat{\Pr}(y>0|x_i)}{\widehat{\Pr}(\tilde{y}>0)}.
\end{aligned}$$

We estimate the conditional densities using the standard kernel density estimator applied to the subset of data for which income is positive:

$$\begin{aligned}
\hat{f}_{y>0}(y_0, x_0) &= \frac{1}{\alpha_y \alpha_x \sum_i 1(y_i>0)}\sum_{i=1}^n 1(y_i>0)K(\frac{y_i - y_0}{\alpha_y})K(\frac{x_i - x_0}{\alpha_x}) \\
\hat{f}_{y>0}(x_0) &= \frac{1}{\alpha_x \sum_i 1(y_i>0)}\sum_{i=1}^n 1(y_i>0)K(\frac{x_i - x_0}{\alpha_x}).
\end{aligned}$$

We now have all the ingredients to simulate the post-intervention earnings distribution. Earnings is drawn from the mixture

$$= \ 0 \text{ with } \Pr(\tilde{y}=0)$$

$$= \quad y \sim \widehat{\tilde{f}}_{\tilde{y}>0}(y) \text{ with } \Pr(\tilde{y} > 0)$$

2.3 Measures of Inequality

After simulating the distribution of earnings with and without actual and hypothetical program impacts, it is possible to examine the effect of the program intervention on inequality using standard measures of inequality developed in the literature. Below, we describe each of the measures considered in the empirical analysis as functions of the estimated densities, taking into account that densities may have probability mass at zero.

Coefficient of variation The coefficient of variation is a common measure of dispersion of a distribution. It is given by

$$\begin{aligned}
\text{coef of variation} &= \frac{\sqrt{\mathrm{Var}[y]}}{\mathrm{E}[y]} \\
\mathrm{E}[y] &= \Pr(y=0)\cdot 0 + (1-\Pr(y=0))\int_0^\infty y f_{|y>0}(y)dy \\
&= (1-\Pr(y=0))\int_0^\infty y f_{|y>0}(y)dy
\end{aligned}$$

$$\mathrm{Var}[y] = \Pr(y=0)\mathrm{E}[y]^2 + (1-\Pr(y=0))\int_0^\infty (y-\mathrm{E}[y])^2 f_{|y>0}(y)dy$$

Inter-quantile ranges The differences between quantiles of y are computed directly from the empirical cdf:

$$\begin{aligned}
F(y) &= \Pr(y=0) \text{ if } y=0 \\
&= \Pr(y=0) + (1-\Pr(y=0))F_{y>0} \text{ if } y>0
\end{aligned}$$

Gini Coefficient The Gini coefficient is often used as a measure of inequality of a distribution of income. Its values range between 0 and 1, with 0 corresponding to perfect equality and 1 corresponding to perfect inequality (one person has all the income).

$$G = 1 - \frac{1}{E[y]} \int_0^{\infty} (1 - F(y))^2 dy$$

Theil Entropy Coefficient The Theil entropy coefficient can be computed from a set of observations by:

$$T = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \ln \frac{y_i}{\bar{y}}$$

If everyone has the same (i.e., mean) income, then the index equals 0. If one person has all the income, then the index equals $\ln n$.

Taking the limit, we get the following formula in terms of the density, conditional on $y > 0$:

$$T_{|y>0} = \int_0^{\infty} \frac{y}{E[y]} \log \frac{y}{E[y]} f_{|y>0}(y) dy$$

Generalizing this to the case where there can be probability mass at 0 gives the following:

$$T = (1 - \Pr(y = 0)) \int_0^{\infty} \frac{y}{E[y]} \log \frac{y}{E[y]} f_{|y>0}(y) dy$$

2.4 Applying the simulation method to evaluation of *Oportunidades*

We next describe how the above methods are applied in the context of evaluating *Oportunidades*. y represents labor earnings, which is equal to hourly wages times hours worked and is modeled as a function of three covariates: e denotes years of education, h health status (measured by height in centimeters), and x years of labor market experience. The conditional density of labor market earnings is

$$f(y|e, h, x)$$

The overall income distribution integrates over the observed education, health and experience levels in the population:

$$f(y) = \int_{(e,h,x) \in A} f(y|e, h, x) dF_{e,h,x}(e, h, x).$$

Suppose the program is known to impact education levels (e) and health status (h) and that we wish to assess how these impacts translate into changes in the earnings distribution. Let Δ_e denote the average impact on education and Δ_h the average impact on health status. If everyone participates in the program, then we can nonparametrically simulate the effect of the program on the income distribution by increasing the education and health status values to the expected post-program values:

$$\tilde{f}(y) = \int_{(e,h,x)\in A} f(y|e,h,x)dF_{e,h,x}(e+\Delta_e, h+\Delta_h, x|(e,h,x)\in S).$$

Because nonparametric estimation methods do not extrapolate beyond the observed support (A), this simulation can only be performed for the subset of people for whom $(e+\Delta_e, h+\Delta_h, x)\in A$, which we denote by S.

The above equation assumes that everyone experiences a program effect of the magnitude (Δ_e, Δ_h). However, subsidy programs are usually targeted at only a fraction of the population on the basis of poverty-related criteria. They may be targeted, for example, at children from families with low parental education levels, which was an important criterion used for the *Oportunidades* program. Let $D = 1$ for the subset of individuals targeted by the program.

The overall income distribution, $g(y)$, reflects that of the targeted and nontargeted subgroups:

$$\begin{aligned} g(y) &= \Pr(D=0)f(y|D=0) \\ &\quad + \Pr(D=1)\tilde{f}(y|D=1) \end{aligned}$$

Suppose the nontargeted subgroup experiences no effect of the program.[4] The larger is the subgroup targeted by the program ($\Pr(D = 1)$), the larger is the potential effect on the overall earnings distribution.

Using this methodology, we can explore the relative contribution of education and health impacts in changing the overall income distribution, by considering the case where we set $\Delta_e = 0$ and the only effect is through Δ_h, and also the case where $\Delta_h = 0$ and the only effect comes through

[4] This assumption rules out general equilibrium effects, which are discussed in the concluding section of the paper.

Δ_e. Implementing the simulation estimator of the previous section requires nonparametrically estimating the conditional density $f(y|e,h,x)$ and the density $f_{e,h,x}(e,h,x)$. We estimate the latter using three dimensional kernel density estimators:

$$f_{e,h,x}(e_0,h_0,x_0) = \frac{1}{na_e a_h a_x}\sum_{i=1}^{n} K\left(\frac{e_i-e_0}{a_e}\right) K\left(\frac{h_i-h_0}{a_h}\right) K\left(\frac{x_i-x_0}{a_x}\right),$$

where a_e, a_h, and a_x are the bandwidth choices. To estimate the conditional density $f(y|e,h,x)$, observe that the conditional density can be expressed as the ratio of two densities:

$$f(y|e,h,x) = \frac{f(y,e,h,x)}{f(e,h,x)},$$

each of which can be nonparametrically estimated by standard kernel density estimators.

3 Description of the Analysis Subsamples

In this paper, we analyze data from the Mexican Family Life Survey (MxFLS-1), which conducted interviews with 8,440 households in 150 communities in 2002. Every household member age 15 or older was interviewed, yielding about 38,000 individual interviews. 16 of Mexico's 32 states/districts are represented including 70% of the population and weights are provided to make the sample nationally representative. The survey includes comprehensive information on labor force participation and income for both primary and secondary jobs in the formal and informal sectors. The survey also includes information on household structure, education, and health. The key variables used in simulating counterfactual outcomes are income, labor force participation, education level, height and labor market experience. Appendix A describes how each of these variables is constructed from the data.

Table 1 presents descriptive statistics on our two main analysis samples: Adult men and women age 25 to 40. About 10% of men and 64% of women report zero labor income. Mean monthly earnings for males are 3,945 pesos and for women 1,140 pesos, where the means include zeros for nonworkers.[5] The average education level for men is 8.8 years, which is about one year higher

[5]In 2002 the average daily exchange rate was 1 USD equals 9.68 pesos. Because a small number of the the earnings values seemed to be outliers, we implemented a trimming procedure and omitted all individuals who reported income higher than 40,000 pesos/month. This corresponded to 9 of 5,180 observations or the top 0.2%.

than the average for women of 7.7 years. Men are on average 166 centimeters tall, and women are on average 153 centimeters tall. The Gini coefficient for men is 0.483 and for women is 0.819. The higher coefficient reflects the fact that a large fraction of women do not work, so the earnings distribution for women is more unequal than that for men.[6]

4 Empirical Results

In this section, we use the methods described above to simulate how the *Oportunidades* program is likely to affect the earnings distribution of program participants. In particular, we use data on earnings currently observed for the age 25 to 40 population and draw inferences on how increases in schooling and height would affect earnings distributions. The experimental evaluations of the *Oportunidades* program (as well as the previous PROGRESA program) found that the program increases schooling levels by 0.6 years on average and adds about 1cm to height for both men and women.[7] We consider the following hypothetical combinations of impacts and their effect on the earnings outcome distribution: (a) an increase in education of 0.6 years and a corresponding decrease in years of potential labor market experience, (b) an increase in height of one cm, (c) a combined increase in education and height in the magnitudes specified in (a) and (b), (d) an increase in education of 0.6 years with no change in corresponding labor market experience, and (e) an increase in education by three years and a corresponding decline in experience. An increase of three years of education is a very large impact that is much greater than what was observed under the program, but we include this hypothetical impact simply for purposes of comparison.

4.1 Targeting

The first step in simulating the effects of the program on future earnings is to identify the sub-population targeted by the program. Our goal is to simulate the medium and longer term effects

[6]As a point of reference, most developed European nations tend to have Gini coefficients for household income between 0.24 and 0.36. For household income, the United States Gini coefficient is around 0.45 and for Mexico is 0.55 (in 2003).

[7]See Behrman and Hoddinott (2005) for discussion of the impacts of PROGRESA on height, and Schultz (2000, 2004), Behrman, Sengupta and Todd (2005) and Todd and Wolpin (2006) for discussion of schooling impacts.

of *Oportunidades* on earnings inequality and poverty. Ideally, we would compare two groups: The "treatment" group would be the current population after they have experienced 20 years of exposure to the *Oportunidades* program and the "control" group would be the same people in a world where the program did not exist. Unfortunately, we cannot currently observe either group. Also, while we can observe which families are currently participating in the program, it is likely that children from today's *Oportunidades* households may not themselves meet the eligibility criteria when they are adults. Indeed, one of the primary goals of the program is to reduce the intergenerational transmission of poverty.

We therefore adopt a synthetic cohort approach for the simulation, that assumes stability in earnings relationships for neighboring cohorts. In particular, it assumes that individuals age 25 to 40 can be used to represent the future earnings of children currently participating in the program. These individuals are too old to have been exposed to *Oportunidades* when they were children, and the vast majority have completed their education. We simulate the effects of *Oportunidades* by identifying the 40% of them that would have been most likely to be targeted when they were young using observed family background characteristics. We analyze the effects of the program by changing this group's observed characteristics (education, height, and potential experience) in a way that is consistent with the impacts that have been measured in recent program evaluation studies.

The MXFLS-1 dataset does not contain information on all the criteria used to determine eligibility for *Oportunidades*, and in fact the exact eligibility criteria are not made public. However, from interactions with program officials, we know the approximate criteria and use the most closely related variables from the MXFLS-1 dataset to approximate eligibility. In particular, we estimate a probit model for program participation using data on children (age 9 to 12) who are currently participating in *Oportunidades* as a nonlinear function of several variables: mother's education, father's education, whether the household has indoor plumbing, and the number of children age 0-10 in the household. Table 2 presents descriptive statistics for these variables. In the sample, 37% of children participate in *Oportunidades*. The program is most active in the poorer southern states (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla), where 31% of the children live. On average, the children in the sample have mothers with 4.7 years of education and fathers with 5.2 years of education. Only 46% of these children live in households with indoor plumbing. Table 3 shows the estimated coefficients from the probit model for program participation.[8] As expected, parental education, indoor plumbing, and the presence of young children in the household are highly significant determinants of program participation.

Next, we compute a propensity score for each adult age 25 to 40 using the estimated coefficients and measures of their family background (parental education, characteristics of the household when they were age 12, and an approximation of the number of children age 0 to 10 in the household at that time). Although the actual targeting of *Oportunidades* is based on several additional variables, we have to restrict the analysis to the subset available in the dataset for both children and adults, which fortunately includes the major determinants of program eligibility. We classify the 40% with the highest predicted probabilities of participation as the target group and the remaining 60% as the non-target group.

Table 4 compares the characteristics of the target and non-target groups, separately for men and women. For both men and women, the target group has much lower maternal and paternal education levels. Individuals in the target groups also grew up with more young children in households that weremuch less likely to have indoor plumbing. For both men and women, there is a two year educational gap between the target and non-target groups as well as a two cm difference in height. The labor market experience measure we use is Mincer potential experience, which equals age minus years of education minus six. The target group has more experience under this measure, mainly because of having less education.[9]

The mean levels in Table 4 clearly show that the target group is less advantaged than the non-target group. In particular, mean monthly earnings are 3,300 pesos per month for targeted

[8]participation model is estimated only for children in rural and semi-urban areas, because in 2002 (the time of our data collection) the program had not been significantly extended to urban areas. The data contain information pertaining to interviews about the child with the parents of 1,970 children age 9-12 in rural areas. After dropping the children with missing variables, we are left with 1,699 observations.

[9]The MXFLS data do not include years of actual labor market experience.

men and 4,300 pesos per month for non-targeted men. Targeted women can expect about half (700 pesos per month) the labor income of non-targeted women (1,500 pesos per month). But there is still substantial overlap in the two income distributions, as shown in Figure 1. The top panel describes men's labor income while the bottom panel describes women's. The solid line in each panel is a nonparametric estimate of the density of positive earnings, while the two dashed lines correspond to the densities of positive earnings in the target and nontarget groups. The target density has been scaled by a factor of 0.4 and the non-target by a factor of 0.6 so that together they add up to equal the total population density. Again, the mean of the the target subsample is clearly lower than that of the nontarget, but a significant proportion of the target group can expect to receive earnings above the population mean and a large proportion of the nontarget group receives very little income.[10]

4.2 Simulating Counterfactual Distributions

We next compute counterfactual distributions of future labor income to show the medium-term effects of the *Oportunidades* program. It is useful to start by looking at plots of the conditional density of labor income with respect to two of our measures of human capital: education and height. Figures 2 and 3 illustrate the nonlinear relationship between distributions of income, education and height, and thus the benefit of estimating it nonparametrically. Figure 2 graphs the conditional density of male non-zero labor income, conditional on education and height and Figure 3 does the same for women. It is evident from the figures that some levels of schooling attainment have a much larger marginal benefit than other levels. The pattern is more homogeneous with respect to height, but there is still evidence of nonlinearity at the upper end of the height distribution.

Table 5a and 5b show the results of our main simulation experiments for men and women. The first column displays characteristics of the null counterfactual distribution and should be used as

[10]The fraction of men receiving no labor income differs very little between the target (11.4%) and nontarget (0.6%) groups, but the difference is actually quite large among women where 71% of targeted women receive no labor income compared to 58% of the non-target group.

a point of comparison. This income distribution is computed using an unchanged counterfactual distribution of human capital and is approximately equal to the original income distribution except for the error introduced by nonparametric smoothing. The other columns of Table 5a and 5b each represent a different set of impacts, given by (a)-(e), where we give the stated program impact to each individual in the target group. For example, in the case of (a), each individual's education level is augmented by 0.6 years, which implies a corresponding decrease in the Mincer measure of labor market experience. We use the nonparametric simulation method developed above to generate a counterfactual earnings distribution whose features can be compared to the pre-program earnings distribution. As previously noted, we simulate changes in labor force participation along with changes in the earnings distribution. That is, the earnings distribution includes a mass point at zero for nonworkers and the fraction of nonworkers can be affected by the program. As above, monthly earnings are measured in thousands of pesos.

Table 5a indicates that the program would not significantly affect the fraction of men participating in the labor market, which remains around 90% across all the simulations. Also, impacts (a)-(d) have modest effects on mean men's earnings and almost no effect on men's earnings inequality, regardless of the measure. The effect of a 0.6 year impact on education (in columns (b) and (d)) is larger for women than it is for men; however, we continue to observe relatively minor changes in income inequality. The hypothetical large three year increase in education, shown in column (e), leads to substantially higher mean earnings. It does not, however, lead to substantial changes in income inequality for men. A one cm increase in height leads to about a 30 peso increase in mean monthly earnings for men but no substantial difference for women. The height impact leads to a slight increase in earnings inequality for both men and women, according to the Theil measure. The estimates in Tables 5a and 5b suggest that the program's impacts on education and on height increase earnings in levels, but do not have much effect in reducing earnings inequality. One factor that can explain this finding is that targeting children from poor backgrounds is not the same as targeting future low-earning adults, because of intergenerational mobility. Another factor that explains the modest effects on inequality is the increasing returns to education on earnings. For example, a year of education in secondary school has a higher monetary return than an additional year of primary school. Those individuals in the target group that already have substantial amounts of education are experiencing bigger increases in salary than those in the target group that start with low amounts of schooling. That said, it does seem that the large increase in education reduces inequality among women. This is driven almost entirely by an increase of about 5% in the fraction of women earning positive labor income.

Tables 6a and 6b present the simulation results for individuals living in five southern states in Mexico that are considered to be among the poorest and where the *Oportunidades* program has been particularly active. The estimated reductions in inequality are a little larger than those shown in Table 5a and 5b for the whole country, although, with the exception of the large increase in education for women, they are still modest. Inequality declines for men in these states because of the moderate (at least for the three year education intervention) increase in their labor force participation. The *Oportunidades* program seems likely to increase earnings levels of the next generation by augmenting human capital, but is not likely to lead to any large reduction in earnings inequality, other than through its transfer mechanism.

In addition to implementing the nonparametric simulations, we also used a parametric model to simulate effects on inequality to see how sensitive our results are to the particular simulation method. Table 7 presents estimated coefficients from a parametric log earnings regression that we use to predict the earnings that each individual would get with any given increase in education or height. Only workers are used in the regression. The simulation assumes that the program does not affect the proportion working, so nonworkers receive zero earnings with or without the program. To simulate earnings with the program, we augment the regressand (education, height, and/or experience), keep the estimated residual associated with each worker the same and predict earnings.

The regression specification allows the monetary return to an additional year of education to differ within three major schooling categories: primary school (1-6), secondary school and high school (7-12) and college (post 12). The pattern of estimated coefficients mirrors reinforce our previous finding that the monetary return to education is higher at higher levels of education. The coefficients on height and height squared are strongly jointly significant for both men and women and the marginal effect of an additional centimenter at the mean height is about 2% for both groups.[11] Similarly, the coefficients on experience and experience squared are also jointly significant. The marginal effect of an additional year of experience at the mean (1% for men and 2% for women) is much lower than the return to any year of education. Tables 8a and 8b show the simulation results that are based on the parametric earnings model. A comparison of Table 8a with Table 5a shows that for men, the parametric simulation approach tends to predict slight reductions in inequality relative to the slight increases in inequality predicted by the nonparametric approach.

The difference in how the two methods evaluate the impact on inequality for the women is sometimes stark. For example, the nonparametric method predicts a sizeable decline in inequality for women in response to the three year increase in education, but the parametric method predicts almost no change. This is mainly because the parametric model does not account for the large increase in participation predicted in the nonparametric method.[12] Figure 4 plots the original and counterfactual densities predicted by each method for the three year education increase. It shows graphically that conditional on participating, the predicted income densities are quite similar under the parametric and nonparametric methods.

The final set of parametric results are shown in Tables 9a and 9b. These tables show how *Oportunidades* affects the distribution of future labor income in the poor southern states according to the parametric method described above. The positive effects on mean levels of earnings are very similar to those found using the nonparametric methods, but the moderate reductions in inequality are nonexistent due to lack of adjustment in the labor force participation rate.

[11] Our estimate is somewhat larger than one reported in a study by Strauss and Thomas (1977) for Brazil that finds a 1% increase in height leads to a 2.4% increase in adult male earnings, in a regression of log hourly wages on height and years of education.

[12] A more refined parametric approach that incorporates a model of program participation could also be implemented.

5 Conclusions

The *Oportunidades* program aims to reduce poverty of the current generation through transfers and to alleviate poverty of the next generation through human capital investment. A number of experimental and nonexperimental evaluation studies have documented that the program significantly impacts educational attainment and health over the short-term. This paper develops and applies a nonparametric simulation method for the purpose of studying how increases in education and health will likely affect the distribution of earnings in the next generation. This new simulation method builds on techniques previously developed in DiNardo, Fortin, Lemieux (1996) by applying the methods in a program evaluation context and by incorporating effects on both labor market participation and earnings.

The empirical findings suggest that the human capital investment in today's youth will increase their mean earnings levels, but will not have a large effect on earnings inequality. Behrman (2006) comes to a similar conclusion in a survey of human capital policies and from an empirical study of how increasing education affects earnings inequality in Chile. The key factors underlying the modest effects on inequality that we observe are the difficulty in predicting which children will become future low earning adults and nonlinearities in how height and education are priced in the labor market. With regard to the first factor, childhood poverty is a strong predictor of future low earnings, but there is also substantial intergenerational mobility that makes it difficult to target low adult earners on the basis of childhood characteristics. With regard to the second factor, we found evidence of important nonlinearities in how height and education influence earnings. Because of these nonlinearities, people at the upper deciles on the targeted population tend to benefit more from the program intervention. Most notably, an additional year of secondary school has a higher monetary return than an additional year of primary school.

We conclude by considering some limitations of the simulation method studied in this paper. First, the method assumes that the observed relationship between earnings and the covareates of education, height, and work experience is causal. This raises concern about potential bias due to unobserved ability, which is the subject of a large U. S. labor literature. Previous attempts to control for ability bias have relied mainly on instrumental variables or natural experiments (e.g. twins with different levels of schooling).[13] Although there is variation in reported estimates, most estimates of the rate of return to education that purport to control for ability bias through the use of instrumental variables often exceed those obtained by ordinary least squares. The variation in estimates is partly accounted for by hetergeneity in returns to education on earnings that requires a LATE (local average treatment effect) interpretation of the instrumental variables estimates.[14] Estimates that account for ability bias using variation in twin pairs, on the other hand, tend to be somewhat lower than cross-sectional OLS estimates. Because the literature finds that OLS estimates do not necessarily overstate instrumental variables estimates, we have no reason to believe that our nonparametric procedure necessarily overstates the return to education. However, further exploration of how the simulation method could be modified to account for unobserved ability would be useful.

A second critical assumption of the simulation method is the usual synthetic cohort assumption, namely the characteristics of today's 25 to 40 year olds of 2002 are representative of the future adulthood of today's children. Extrapolating out from current time trends, children today will likely attain more education than current 25 to 40 year olds. Our estimates indicate that the marginal effect of education is increasing in years of education, so rising education levels could lead the simulation to understate somewhat the impact of *Oportunidades* on earnings. Third, the simulation method does not account for the general equilbrium effects of increasing the education levels of a large fraction of future labor force, which would tend to decrease returns to education. Any decline, though, is mitigated some by the fact that Mexico is an open economy. Fourth, this study focused on individual level earnings for men and women, although household-level earnings inequality may be more relevant to policy makers. It is also not clear how to interpret high income inequality in a group (like women) where a large proportion choose not to work, because they have a partner who provides enough money for the household. The simulation

[13] e.g., Behrman, Rosenzweig and Taubman (1994), Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), Card (1995, 1999).

[14] Card (1999).

method could be modified to incorporate a marriage outcome, where marriage opportunities and outcomes potentially also depend on variables influenced by the program.

Lastly, improvements in future earnings are only one of the long-term benefits expected from the program. For example, there is a substantial literature documenting that upgrading mother's education increases child test scores. Female program beneficiaries who choose not to work may be more effective mothers and may choose to have fewer children and to invest more in them.

References

[1] Ashenfelter, O. and A. B. Krueger (1994): "Estimates of the Economic Return to Schooling from a New Sample of Twins," *Quarterly Journal of Economics*, 84(5):1157-73.

[2] Ashenfelter, O. and C. Rouse (1998): "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins," *Quarterly Journal of Economics*, 113(1):253-84.

[3] Behrman, Jere R. and John Hoddinott (2005): "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition," Oxford Bulletin of Economics and Statistics, Vol. 67, No. 4, pp. 547–569.

[4] Behrman, Jere R., Piyali Sengupta and Petra Todd (2004): "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment," Philadelphia: University of Pennsylvania, forthcoming in *Economic Development and Cultural Change.*

[5] Behrman, Jere R. and Emmanuel Skoufias (2004): "Evaluation of PROGRESA/Oportunidades: Mexico's Anti-Poverty and Human Resource Investment Program," in Jere R. Behrman, Douglas Massey and Magaly Sanchez R, eds., *The Social Consequences of Structural Adjustment in Latin America*, book manuscript.

[6] Behrman, Jere R. (2006): "How Much Might Human Capital Policies Affect Earnings Inequalities and Poverty?," working paper.

[7] Behrman, Jere R., Rosenzweig, Mark R. and Paul Taubman (1994): "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," in *Journal of Political Economy*, volume 102, no. 6, 1131-1174.

[8] Buddelmeyer, Hilke, and Emmanuel Skoufias (2003) "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," IZA Discussion Paper No. 827 (July), Institute for the Study of Labor (IZA), Bonn, Germany.

[9] Card, David (1999): “The Causal Effect of Education on Earnings.” In *Handbook of Labor Economics, Vol. 3A*, edited by Orley Ashenfelter and David Card. (Amsterdam: North Holland).

[10] Card, David (1995): “Earnings, Schooling, and Ability Revisited,” in *Research in Labor Economics,* Vol. 14, edited by S. Polachek, Greenwich, Conn: JAI Press, 23-48.

[11] Chakravarty, S. R. (1990). *Ethical Social Index Numbers.* New York: Springer-Verlag.

[12] DiNardo, John, Nicole M. Fortin, Thomas Lemieux (1996): “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,” Econometrica, Vol. 64, No. 5, pp. 1001-1044.

[13] Gertler, Paul (2000): ”Final Report: The Impact of PROGRESA on Health,” *International Food Policy Research Institute*, Washington, D. C.

[14] Parker, Susan W. and Emmanuel Skoufias, 2000, “The impact of PROGRESA on work, leisure and time allocation,” October. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C. <http://www.ifpri.org/themes/progresa.htm>

[15] Persico, Nicola, Andrew Postlewaite, and Daniel Silverman (2004): “The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height,” *Journal of Political Economy.*

[16] Schultz, T. Paul (2000): “Impact of PROGRESA on school attendance rates in the sampled population,” February. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C.

[17] Schultz, T. Paul (2004): “School subsidies for the poor: Evaluating a Mexican strategy for reducing poverty,” *Journal of Development Economics*, (Revision of June 2000 Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C. http://www.ifpri.org/themes/progresa.htm

[18] Skoufias, E. and B. McClafferty (2001): "Is PROGRESA working? Summary of the results of an Evaluation by IFPRI," Report submitted to PROGRESA. Washington, D.C.: International Food Policy Research Institute, <http://www.ifpri.org/themes/progresa.htm>

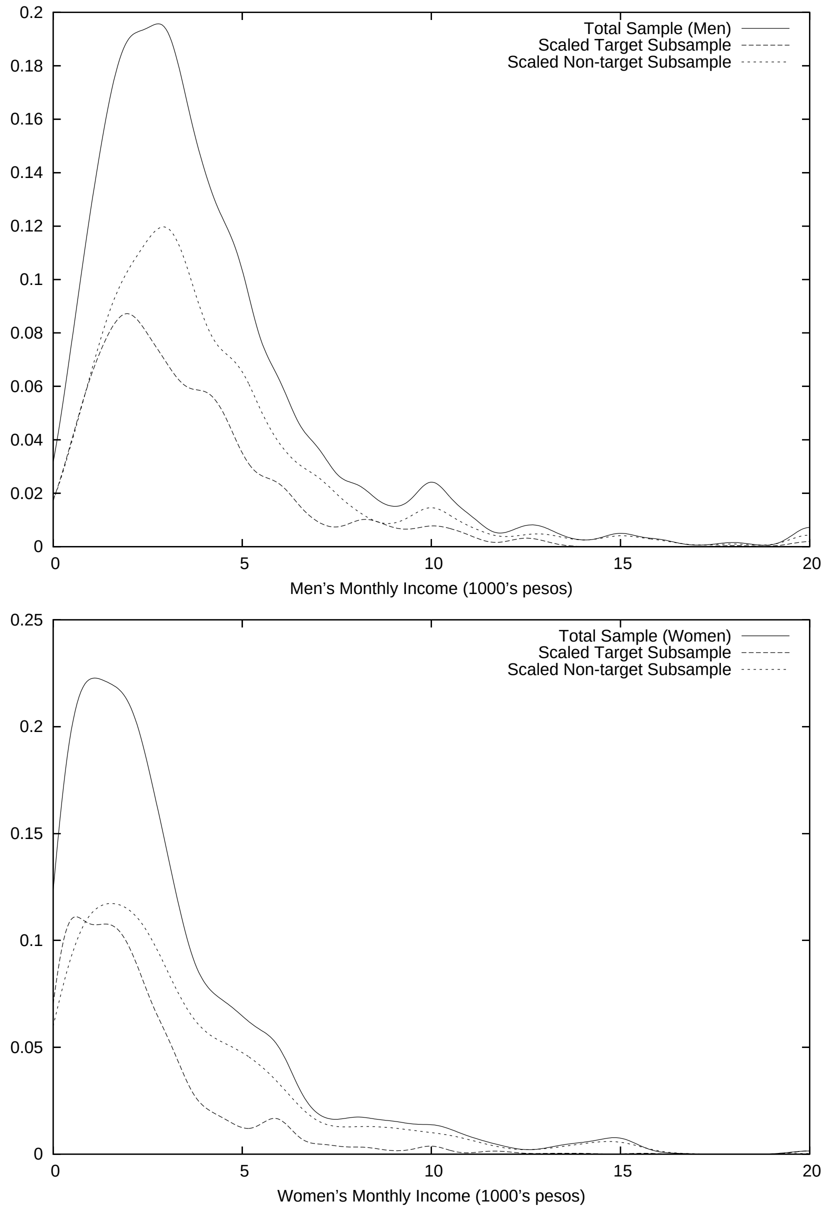
[19] Strauss, John (1986): "Does Better Nutrition Raise Farm Productivity?" *Journal of Political Economy,* 94, April, p. 297-320.

[20] Strauss, John and Duncan Thomas (1998): "Health, Nutrition and Economic Development," *Journal of Economic Literature*, Vol. 36 (2), 766-817.

[21] Todd, Petra and Kenneth Wolpin (2006): "Assessing the Impact of a School Subsidy Program in Mexico," with Kenneth I. Wolpin, American Economic Review, 2006, December.

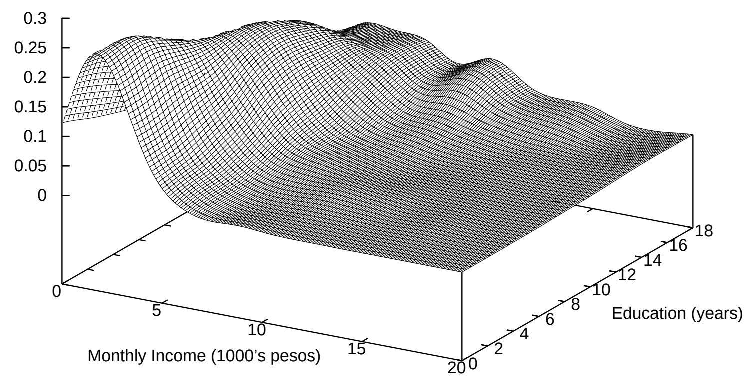
[22] Todd, Petra A., 2004, "Technical Note on Using Matching Estimators to Evaluate the Opportunidades Program For Six Year Follow-up Evaluation of Oportunidades in Rural Areas," Philadelphia: University of Pennsylvania, mimeo.

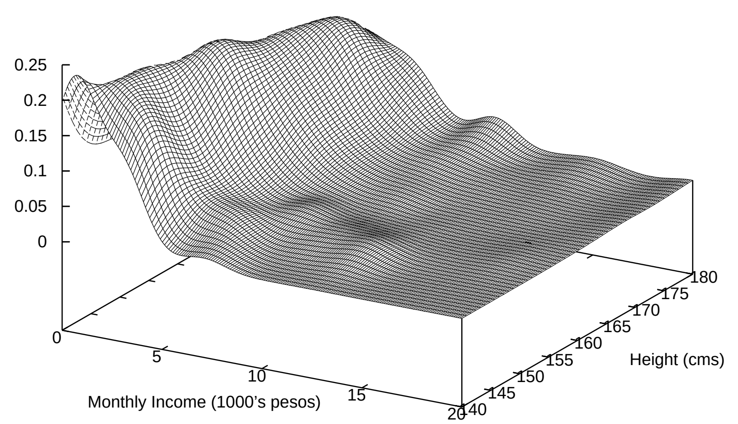
Figure 1: Densities of Income for Men and Women



Source: MXFLS 2002
All densities are nonparametrically estimated using non-zero values of income.

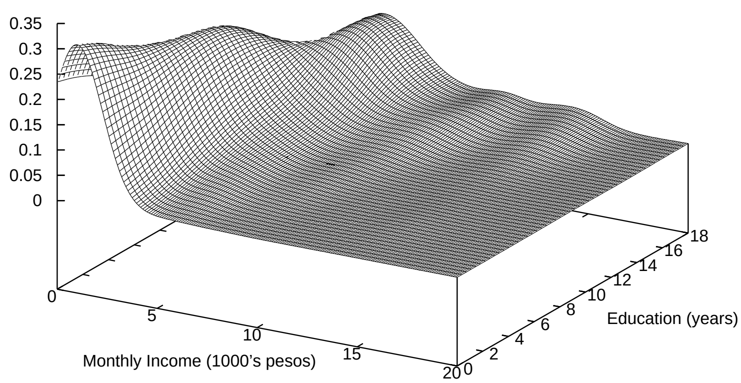
Figure 2: Density of Men's Income Conditional on Education and Height

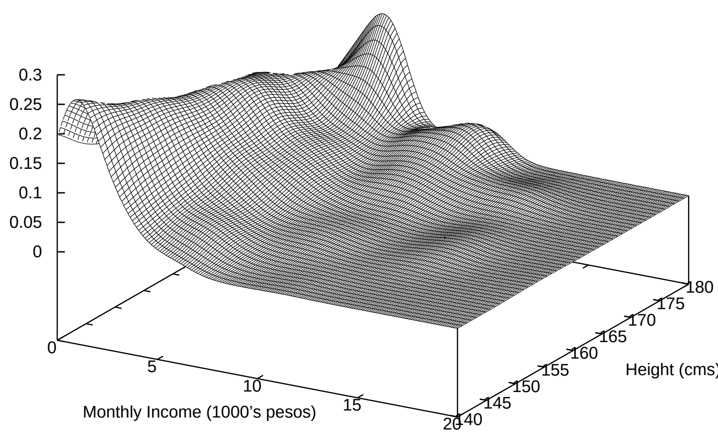




Source: MXFLS 2002
All densities are nonparametrically estimated using non-zero values of income.

Figure 3: Density of Women's Income Conditional on Education and Height

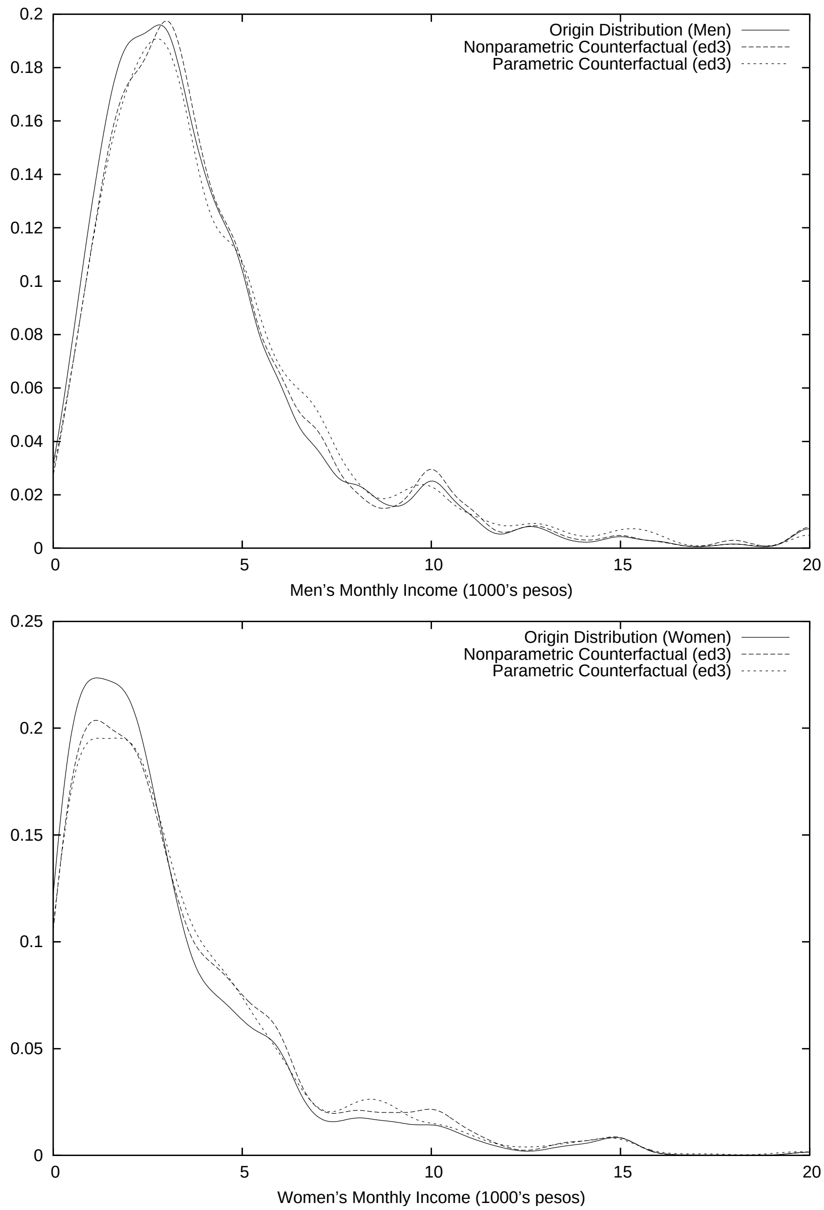




Source: MXFLS 2002

All densities are nonparametrically estimated using non-zero values of income.

Figure 4: Origin and Counterfactual (ed3) Densities of Income



Source: MXFLS 2002
All densities are nonparametrically estimated using non-zero values of income.

Table 1
Descriptive Statistics
Men and Women, age 25-40

	Men	Women
Proportion with zero earnings	0.0999	0.638
Mean monthly earnings, measured in pesos	3.945 (0.187)	1.140 (0.127)
Median earnings	3.000	0.000
Interquartile range of earnings	3.300	3.600
Coefficient of Variation	1.123	2.276
Gini Coefficient	0.483	0.819
Theil Index	0.443	1.459
Fraction of those with positive earnings with earnings below 25 percentile	0.095	0.273
Mean education level, measured in years	8.8 (0.27)	7.7 (0.20)
Mean height, measured in cm	166 (0.52)	153 (0.41)
Mean potential labor market experience	17.3 (0.36)	18.5 (0.26)
Sample Size	1950	3221

Table 2
Descriptive Statistics for Variables Used in Targeting Analysis
Children Age 9-12

	Children age 9-12
Participates in PROGRESA	0.37 (0.05)
Mother's education	4.7 (0.28)
Father's education	5.2 (0.21)
Maximum of parents' education	6.1 (0.22)
Household has indoor plumbing	0.46 (0.05)
Number of children age 0-10 in household	2.1 (0.08)
Lives in Poor Southern State‡	0.31 (0.07)
Sample Size	1699

‡ Chiapas, Oaxaca, Guerrero, Michoacan, or Puebla

Table 3
Estimated Probit Model for Probability of Participating in PROGRESA

Variable	Coefficient	p-value
Mother’s education 6 years of less	-0.624	0.000
Mother’s education 7 to 9 years	-0.914	0.001
Mother’s education 10 to 12 years	-1.286	0.006
Mother’s education 13 or more years	-0.652	0.264
Father’s education 6 years of less	-0.592	0.000
Father’s education 7 to 9 years	-0.836	0.000
Father’s education 10 to 12 years	-1.284	0.015
Father’s education 13 or more years	0.317	0.453
Max parent’s education 6 years of less	0.750	0.000
Max parent’s education 7 to 9 years	0.916	0.000
Max parent’s education 10 to 12 years	1.211	0.034
Max parent’s education 13 or more years	-0.370	0.551
Indoor plumbing	-0.291	0.029
2 to 4 young childrens in household	0.139	0.160
5 young children in household	0.466	0.037
6 or more young children in household	1.159	0.023
Living in poor southern state‡	0.257	0.201
Constant term	-0.214	0.302
Sample Size	1699	
Pseudo R-squared	0.11	

‡ Chiapas, Oaxaca, Guerrero, Michoacan, or Puebla

Table 4
Descriptive Statistics for Men and Women, age 25-40,
by Projected PROGRESA Participation

	Men		Women	
	40% Target	60% Non-target	40% Target	60% Non-target
Mother’s education	1.9	4.4	1.9	4.1
	(0.14)	(0.22)	(0.12)	(0.21)
Father’s education	2.7	5.1	2.6	4.6
	(0.17)	(0.26)	(0.13)	(0.25)
Max Parental education	3.2	5.7	3.2	5.3
	(0.19)	(0.24)	(0.14)	(0.24)
Indoor plumbing	0.18	0.85	0.18	0.80
	(0.03)	(0.03)	(0.03)	(0.03)
# children age 0-10 in household	2.5	1.1	2.5	1.2
	(0.12)	(0.07)	(0.10)	(0.06)
Living in poor southern state	0.34	0.11	0.35	0.11
	(0.06)	(0.05)	(0.06)	(0.04)
Mean monthy earnings (in 1000s of pesos)	3.3	4.3	0.7	1.5
	(0.28)	(0.25)	(0.07)	(0.20)
Education	7.4	9.6	6.3	8.7
	(0.25)	(0.30)	(0.20)	(0.22)
Height	164.4	166.9	152.1	154.3
	(0.56)	(0.55)	(0.50)	(0.43)
Experience	19.5	16.1	20.7	16.9
	(0.40)	(0.42)	(0.26)	(0.33)
Sample Size	867	1083	1629	1592

Table 5a
Simulated Effects of PROGRESA Impacts on Income Distribution
Based on Nonparametric Earnings Density Estimations
Men, Age 25-40

	Original	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Proportion with zero earnings	0.098	0.099	0.099	0.099	0.098	0.097
Mean earnings	3.931	3.944	3.961	3.974	3.948	4.255
Std. Dev. earnings	4.382	4.382	4.455	4.456	4.373	4.857
Median earnings	3.003	3.013	3.012	3.021	3.024	3.180
Interquartile Range	3.331	3.337	3.338	3.339	3.342	3.443
Coefficient of Variation	1.115	1.111	1.125	1.121	1.108	1.142
Gini Coefficient	0.485	0.485	0.487	0.487	0.484	0.491
Theil Index	0.446	0.445	0.451	0.450	0.443	0.458
Fraction of those with positive earnings with earnings below 25 percentile	0.114	0.114	0.113	0.112	0.113	0.100

Sample size is 1950

Table 5b
Simulated Effects of PROGRESA Impacts on Income Distribution
Based on Nonparametric Earnings Density Estimations
Women, Age 25-40

	Original	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Proportion with zero earnings	0.637	0.632	0.639	0.633	0.632	0.589
Mean earnings	1.147	1.173	1.149	1.177	1.176	1.448
Std. Dev. earnings	2.618	2.644	2.618	2.650	2.653	2.954
Median earnings	0.000	0.000	0.000	0.000	0.000	0.000
Interquartile Range	1.335	1.394	1.337	1.399	1.395	1.909
Coefficient of Variation	2.283	2.254	2.2280	2.251	2.256	2.040
Gini Coefficient	0.819	0.816	0.819	0.816	0.816	0.791
Theil Index	1.478	1.460	1.479	1.460	1.461	1.325
Fraction of those with positive earnings with earnings below 25 percentile	0.282	0.278	0.278	0.275	0.277	0.250

Sample size is 3221

Table 6a
Simulated Effects of PROGRESA Impacts on Income Distribution
In Poor Southern States (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla)
Based on Nonparametric Earnings Density Estimations
Men, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Proportion with zero earnings	0.135	0.132	0.137	0.133	0.133	0.103
Mean earnings	3.441	3.463	3.445	3.471	3.473	3.831
Std. Dev. earnings	4.063	3.980	4.104	4.014	3.988	4.032
Median earnings	2.633	2.693	2.632	2.695	2.708	3.142
Interquartile Range	3.367	3.339	3.393	3.362	3.364	3.256
Coefficient of Variation	1.181	1.149	1.191	1.157	1.148	1.053
Gini Coefficient	0.515	0.506	0.516	0.507	0.506	0.470
Theil Index	0.508	0.490	0.513	0.493	0.491	0.425
Fraction of those with positive earnings with earnings below 25 percentile	0.152	0.145	0.152	0.145	0.145	0.114

Sample size is 334

Table 6b
Simulated Effects of PROGRESA Impacts on Income Distribution
Southern States (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla)
Based on Nonparametric Earnings Density Estimations
Women, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Proportion with zero earnings	0.671	0.664	0.673	0.666	0.666	0.600
Mean earnings	0.776	0.793	0.784	0.803	0.801	1.070
Std. Dev. earnings	1.895	1.904	1.920	1.927	1.927	2.284
Median earnings	0.000	0.000	0.000	0.000	0.000	0.000
Interquartile Range	0.728	0.786	0.723	0.784	0.781	1.358
Coefficient of Variation	2.442	2.400	2.448	2.401	2.405	2.136
Gini Coefficient	0.833	0.830	0.834	0.831	0.831	0.799
Theil Index	1.595	1.570	1.599	1.573	1.574	1.389
Fraction of those with positive earnings with earnings below 25 percentile	0.365	0.366	0.360	0.360	0.362	0.338

Sample size is 591

Table 7a
Estimated Coefficients from Parametric Earnings Models

Variables	Men	Women
Years of Primary Education	0.051	0.095
	(0.026)	(0.035)
Years of Secondary and High School Education	0.074	0.135
	(0.025)	(0.032)
Years of college education	0.153	0.242
	(0.026)	(0.033)
Height	-0.061	0.109
	(0.086)	(0.138)
Height squared	0.00025	-0.00029
	(0.00025)	(0.00044)
Experience	0.063	0.047
	(0.023)	(0.036)
Experience squared	-0.0016	-0.00079
	(0.0006)	(0.00097)
Constant term	3.338	-10.810
	(7.355)	(10.667)
Sample Size	1720	1044
R-squared	0.205	.281

Table 7b
Estimated Coefficients from Parametric Earnings Models
Southern States (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla)

Variables	Men	Women
Years of Primary Education	0.081	0.130
	(0.045)	(0.064)
Years of Secondary and High School Education	0.087	0.192
	(0.043)	(0.059)
Years of college education	0.088	0.163
	(0.029)	(0.055)
Height	-0.068	-0.556
	(0.113)	(0.408)
Height squared	0.0003	0.002
	(0.0003)	(0.001)
Experience	0.034	0.076
	(0.056)	(0.050)
Experience squared	-0.0009	-0.0009
	(0.001)	(0.002)
Constant term	4.270	37.45
	(8.882)	(31.10)
Sample Size	275	192
R-squared	0.201	0.315

Table 8a
Simulated Effects of PROGRESA Impacts on Income Distribution
Based on Parametric Earnings Model
Men, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Mean earnings	3.945	4.003	3.969	4.028	4.006	4.252
Std. Dev. earnings	4.432	4.475	4.450	4.497	4.476	4.711
Median earnings	3.000	3.000	3.000	3.000	3.000	3.171
Interquartile Range	3.300	3.440	3.365	3.406	3.453	3.510
Coefficient of Variation	1.123	1.118	1.121	1.116	1.117	1.108
Gini Coefficient	0.483	0.483	0.483	0.482	0.483	0.483
Theil Index	0.443	0.441	0.442	0.441	0.441	0.439
Fraction of those with positive earnings with earnings below 25 percentile	0.095	0.095	0.095	0.095	0.095	0.087

Sample size is 1950

Table 8b
Simulated Effects of PROGRESA Impacts on Income Distribution
Based on Parametric Earnings Model
Women, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Mean earnings	1.140	1.166	1.146	1.172	1.169	1.284
Std. Dev. earnings	2.595	2.643	2.605	2.654	2.649	2.918
Median earnings	0.000	0.000	0.000	0.000	0.000	0.000
Interquartile Range	1.200	1.295	1.235	1.328	1.302	1.500
Coefficient of Variation	2.276	2.268	2.273	2.265	2.266	2.272
Gini Coefficient	0.819	0.818	0.819	0.818	0.818	0.817
Theil Index	1.459	1.455	1.457	1.453	1.454	1.451
Fraction of those with positive earnings with earnings below 25 percentile	0.273	0.271	0.273	0.271	0.271	0.252

Sample size is 3221

Table 9a
Simulated Effects of PROGRESA Impacts on Income Distribution
Southern States (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla)
Based on Parametric Earnings Model
Men, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Mean earnings	3.397	3.497	3.429	3.531	3.496	3.925
Std. Dev. earnings	4.020	4.134	4.060	4.178	4.125	4.682
Median earnings	2.600	2.667	2.600	2.709	2.645	2.952
Interquartile Range	3.100	3.213	3.172	3.272	3.227	3.727
Coefficient of Variation	1.182	1.180	1.1172	1.181	1.178	1.191
Gini Coefficient	0.510	0.509	0.510	0.509	0.509	0.507
Theil Index	0.494	0.492	0.494	0.492	0.492	0.490
Fraction of those with positive earnings with earnings below 25 percentile	0.129	0.129	0.129	0.129	0.129	0.092

Sample size is 334

Table 9b
Simulated Effects of PROGRESA Impacts on Income Distribution
Southern States (Chiapas, Oaxaca, Guerrero, Michoacan, and Puebla)
Based on Parametric Earnings Model
Women, Age 25-40

	Orig	Ed	Height	Ed and height	Ed, no change in exp	Ed, 3 years
		(a)	(b)	(c)	(d)	(e)
Mean earnings	0.791	0.829	0.811	0.851	0.843	1.002
Std. Dev. earnings	1.908	2.009	1.960	2.067	2.043	2.511
Median earnings	0.000	0.000	0.000	0.000	0.000	0.000
Interquartile Range	0.720	0.720	0.720	0.720	0.720	0.800
Coefficient of Variation	2.410	2.420	2.415	2.428	2.421	2.504
Gini Coefficient	0.833	0.833	0.833	0.833	0.833	0.834
Theil Index	1.545	1.546	1.546	1.549	1.546	1.562
Fraction of those with positive earnings with earnings below 25 percentile	0.348	0.348	0.354	0.340	0.340	0.298

Sample size is 591

Appendix A: Construction of Samples and Variables

This appendix describes how each of the variables for the empirical analysis was constructed. The data analysis has three parts. First, we estimate a probability of participating in the Progresa/Oportunidades program and use the estimated model to simulate program targeting for men and women between age 25 and 40. Second, we estimate nonparametrically the relationship between income, education, height, and work experience for men and women between age 25 and 40. Third, we compute the counterfactual income distribution under assumptions of how the program affects education, height, and work experience that are consistent with recent evaluations of short-term prorgram impacts.

Sample Construction

The initial sample of individuals who filled out book 3a of the survey (and can be generalized to the national population) is 6,564. When we drop the individuals who worked but did not report their income, the number goes down to 5,871. It drops further to 5,180 (79% of the original sample) when we drop those individuals who did not report their education or whose height was not measured. Finally, we drop an additional 9 outlier observations for individuals who report receiving more than 40,000 pesos in the previous month. This leaves a final sample size of 5,171.

Construction of Variables

Income Income is measured as total labor income earned in the previous month. It includes zeros for those individuals who don't work. About 6% of individuals who reported working in the previous week are "peasants on their plot." 40% of these individuals report zero income in the last month. This seems plausible for subsistence farmers. Only 2% of other individuals who report working report zero income. Income is measured in thousands of pesos and in 2002 the average daily exchange rate was 1 USD = 9.68 pesos.

We do not use proxy reports on income, because it is not clear how to combine this data with the first-person reports and weight the data correctly. The proxy reports also have more missing data.

Education The MxFLS collects the type of the last school attended and, for most individuals, the number of years that the individual completed at that level. We do not include years of "technical education" in our measure, because wage returns to technical education (based on linear regressions) are much lower than the returns of conventional schooling.

Height Height is measured in centimeters.

Experience The survey did not collect information on actual labor force experience, so we use the standard Mincer measure of potential experience equal to age minus education minus six.