

Ex Ante Evaluation of Social Programs

Petra E. Todd and Kenneth I. Wolpin¹
University of Pennsylvania

July 6, 2007

¹The authors may be contacted at petra@athena.sas.upenn.edu and wolpink@ssc.upenn.edu. We thank Jere Behrman, Andrew Foster, Jeffrey Grogger, James J. Heckman, Joseph Hotz, Hidehiko Ichimura, Susan Parker and Jeffrey Smith for helpful comments. This paper was presented in June, 2005 at the Nuremberg at the Active Labor Market conference at the IAB Institut and in December, 2005 at a CREST-INSEE conference on evaluation methods in Paris.

Abstract

This paper discusses methods for evaluating the impacts of social programs prior to their implementation. Ex ante evaluation is useful for designing programs that achieve some optimality criteria, such as maximizing impact for a given cost. This paper illustrates the use of behavioral models in predicting the impacts of hypothetical programs in a way that is not functional form dependent. Among the programs considered are wage subsidy programs, conditional cash transfer programs, and income support programs. In some cases, the behavioral model justifies a completely nonparametric estimation strategy, even when there is no direct variation in the policy instrument. In other cases, stronger assumptions are required to evaluate a program ex ante. We illustrate the application of ex ante evaluation methods using data from the PROGRESA school subsidy randomized experiment in Mexico. We assess the effectiveness of the ex ante prediction method by comparing predictions of program impacts to the impacts measured under the randomized experiment. The subsamples pertain to girls and boys aged 12-15. For the girls, the predicted impacts are fairly similar to the actual impacts, both in magnitude and in replicating the age patterns, with larger impacts observed at higher ages. For boys, the predicted impacts tend to overstate the actual impacts. The ex-ante evaluation method is also used to predict the effects of counterfactual programs that include changes to the subsidy schedule and an unconditional income transfer.

1 Introduction

Most program evaluation research focuses on the problem of *ex post* evaluation of existing programs. For example, evaluation methods such as matching or control function approaches typically require information on individuals that receive the program intervention (the treatment group) as well as on a comparison group sample that does not receive it. A limitation of these approaches is that they do not provide ways of evaluating the effects of programs prior to introducing them.

For many reasons, it is important to develop tools for *ex ante* evaluation of social programs. First, *ex ante* evaluation of a range of programs makes it possible to optimally design a program that achieves some desired impacts at a minimum cost or maximizes impacts for a given cost. Finding an optimal program design can be challenging, because it requires simulating the impacts of potentially many hypothetical programs as well as simulating program take-up rates, in order to assess costs and program coverage. The alternative experimental approach would implement alternative versions of the program and compare their impacts, but such an approach is often too costly and too time consuming to be feasible for program design purposes. A second benefit of an *ex ante* evaluation is that it may help avoid the high cost of implementing programs that are later found to be ineffective.¹ Third, *ex ante* assessment can provide some evidence on what range of impacts to expect after the program is implemented, which is useful for program placement decisions and for choosing sample sizes for any *ex post* evaluation. Fourth, in cases where there is already a program in place, *ex ante* evaluation methods can be used to study how the impacts would change if some parameters of the program were altered. As these examples illustrate, an *ex ante* evaluation is not a substitute for an *ex post* evaluation. Even if we regard *ex post* evaluations to be more reliable for estimating treatment impacts of an existing program, there is still a critical role for *ex ante* evaluation tools.

¹For example, the JTPA (Job Training Partnership Act) program was a multi-billion dollar program in the U.S. that was recently replaced, in large part because the experimental evaluation of the program showed that it was ineffective for many of the participants.

In this paper, we illustrate through several examples how to use behavioral models to predict the impacts of hypothetical programs and to justify particular estimation approaches. Among the programs considered are wage subsidy programs, conditional cash transfer programs, and income support programs. Specifying a behavioral model is usually a necessary step in developing ways of predicting the effects of a program absent any data on treated individuals. However, strong functional form assumptions are not necessarily required. As emphasized in early papers by Marschak (1953) and Hurwicz (1962) and in the more recent work of Heckman (2000,2001), Ichimura and Taber (1998, 2002) and Blomquist and Newey (2002), estimating the effect of a new policy does not necessarily require specifying the complete structure of the model governing decisions. The benefits of these kinds of methods are that they are flexible with regard to exact functional forms of behavioral models, however this benefit comes at some cost as the methods do typically require stronger independence assumptions on the distribution of observed heterogeneity and restrictions on the class of behavioral models. This paper builds on the previous literature by illustrating, using specific economic models, how to verify when the conditions for nonparametric policy evaluation are met for a variety of program interventions. As some of the examples illustrate, nonparametric estimation is sometimes feasible even when the data do not contain any direct source of variation related to the program intervention. We also provide examples where fully nonparametric estimation is not feasible and more structure is required to obtain ex ante estimates of program impacts.

This paper also suggests and implements some simple estimation strategies which are based on a modified version of the method of matching. The estimator obtains estimates of treatment effects by matching untreated individuals to other untreated individuals, where the particular set of regressors used to select the matches is implied by the economic model.

After describing the methods and the estimators, we study their performance in an application to data from the PROGRESA experiment in Mexico. PROGRESA is a conditional cash transfer program that provides cash transfers to parents conditional on their children

attending school.² The program was initially implemented as a randomized experiment, which creates a unique opportunity to use the experimental estimates to benchmark the performance of *ex ante* evaluation methods. In this paper, we compare the *ex ante* predicted program impacts, estimated using data from the randomized-out control group that did not receive the program, to the program impacts measured under the experiment. We find that the *ex ante* prediction method predicts accurately the estimated impacts for girls, but overpredicts somewhat the estimated impacts for boys. Application of the method to study counterfactual subsidy schedules indicates that older age children (age 14-15) would be highly sensitive to changes in the subsidy schedule, while younger children (age 12-13) would not be. Doubling the subsidy leads to almost a doubling of the predicted program impacts for the older children, whereas reducing the subsidy by 25% leads to roughly a halving of the predicted impacts.

2 Related Literature

The problem of forecasting the effects of hypothetical social programs is part of the more general problem of studying the effects of policy changes prior to their implementation that was described by Marschak (1953) as one of the most challenging problems facing empirical economists.³ In the early discrete choice literature, the problem took the form of the "forecast problem," in which researchers used random utility models (RUMs) to predict the demand for a new good prior to its being introduced into the choice set.⁴ Both theoretical and empirical criteria were applied to evaluate the performance of the models. Theoretically, the probabilistic choice models were compared in terms of the flexibility of the substitution patterns they allowed.⁵ Empirically, the model's performance could sometimes be assessed

²The latest incarnation of the program is called *Oportunidades*, but our dataset pertains to the initial implementation and evaluation.

³See Heckman (2000).

⁴Much of the initial empirical research was aimed at predicting the demand for transportation modes.

⁵For example, McFadden observed, with his famous Red Bus-Blue Bus example, that assuming iid Weibull errors, as in a multinomial logit model, gives unreasonable forecasts when a new good that was similar to an existing good is introduced into the choice set. (See McFadden, 1984.) More recently, Berry, Levinsohn and Pakes (1995) evaluate alternative models of automobile choice in terms of the flexibility of

by comparing the model's predictions about demands for good with the ex post realized demand.

In one of the earliest applications of this idea, McFadden (1977) uses a RUM to forecast the demand for the San Francisco BART subway system prior to its being built and then checks the accuracy of the forecasts against the actual data on subway demand. Using a similar idea, Lumsdaine, Stock and Wise (1992) study the performance of alternative models at forecasting the impact of a new pension bonus program on the retirement of workers. The program offered a bonus for workers at a large firm who were age 55 and older to retire. The authors first estimate the models using data gathered prior to the bonus program and then compare the models' forecasts to actual data on workers' departures.

There are a few empirical studies that study the performance of economic models in forecasting program effects by comparing models' forecasts of treatment effects to those obtained from randomized experiments. Wise (1985) develops and estimates a model of housing demand and uses it to forecast the effects of a housing subsidy program. He then compares his models' forecasts to the subsidy effects observed under a randomized experiment. More recently, Todd and Wolpin (2004) develop and estimate a dynamic behavioral model of schooling and fertility that they use to forecast the effects of the PROGRESA program on school and work choices and on family fertility. They evaluate the performance of the model in predicting the effect of the subsidy by structurally estimating the model on control group data and comparing the model's predictions regarding treatment effects to those estimated under the randomized experiment.⁶ In this paper, our application is to the same data and the goal of predicting the effects of the subsidy is similar. However, the *ex ante* evaluation methods studied here are much different than the methods studied in Todd and Wolpin (2004). They are based on simpler modeling structures, do not require structural estimation, and impose very weak functional form assumptions. Another recent

the substitution patterns allowed.

⁶After finding that the model forecasts well the effects of the existing subsidy program, they use the estimated model to evaluate the effects of a variety of hypothetical programs. They find an alternative subsidy schedule that would be expected to yield higher impacts on years of educational attainment at the similar cost to the existing program.

study that also uses experimental data to validate a structural model is that of Lise, Seitz and Smith (2003), which uses a calibrated search-matching model of the labor market to predict the impacts of a Canadian program that provides bonuses to long-term welfare recipients for returning to work. They validate the model by comparing its predictions against an experimental benchmark.

3 Ex Ante Evaluation Methods and Estimators

Ex ante evaluation requires extrapolating from past experience to learn about effects of hypothetical programs. In some cases, the source of extrapolation is relatively straightforward. For example, to evaluate the effect of a wage subsidy program on labor supply, we can extrapolate from the observed hours-wage variation in the data. Heckman (2000) discusses other examples pertaining to evaluating the effects of a commodity tax when there is observed price variation in the data. Ichimura and Taber (1998, 2002) have an application to evaluating the effects of a college tuition subsidy when there is observed tuition variation in the data. In other cases, however, there may be no variation in the data directly related to the policy instrument. An example we consider in this paper is the problem of evaluating the effects of a subsidy for children to attend school when we start from a situation where schooling is free for everyone.

Below, we provide examples of how to use the structure of economic models to identify program effects for different kinds of program interventions such as multiplicative wage subsidies, additive wage subsidies, income subsidies, a combination of wage and income subsidies, and school subsidy programs. For each example, we discuss estimation strategies.

3.1 Wage and income subsidy programs

A multiplicative wage subsidy program In the first example, we analyze the effect of introducing a wage subsidy on labor supply. Suppose labor supply behavior is described by a standard static model in which individuals choose the number of hours to work given their

wage rate and given their level of nonlabor asset income and total time available equal to 1.

$$\begin{aligned} & \max_{\{h\}} U(c, 1 - h, \mu) \\ & s.t. \\ & c = hw + A \end{aligned}$$

Optimal hours of work (h) can be derived as a function of wages (w) and asset income (A) and of unobserved heterogeneity in preferences (μ):

$$h^* = \varphi(w, A, \mu).$$

If we now introduce a multiplicative subsidy to wages in the amount τ , so that the budget constraint becomes

$$c = h(\tau w) + A.$$

The model with the subsidy can be viewed as a version of the model without the subsidy. That is, if $h^{**} = \eta(w, A, \tau, \mu)$ denotes the solution to the model with the subsidy, then

$$h^{**} = \eta(w, A, \tau, \mu) = \varphi(\tilde{w}, A, \mu)$$

where $\tilde{w} = w\tau$. Thus, the model without the subsidy is also the relevant one in the presence of the subsidy, suggesting that the effect of introducing a subsidy τ can be studied from ex ante wage variation in the data.

Consider a comparison between the average hours worked for persons with wages \tilde{w} and the same level of assets A to the average hours worked for a person with wages w and assets A , where the averages include individuals with zero hours:

$$h^{**}(\tilde{w}, A) - h^*(\tilde{w}, A),$$

where

$$\begin{aligned} h^{**}(\tilde{w}, A) &= \int \varphi(\tilde{w}, A, \mu) f(\mu|\tilde{w}, A) d\mu \\ h^*(w, A) &= \int \varphi(w, A, \mu) f(\mu|w, A) d\mu. \end{aligned}$$

Under the assumption that, conditional on assets, the distribution of unobserved heterogeneity does not depend on wages, i.e.

$$f(\mu|w, A) = f(\mu|A),$$

the comparison of average hours worked at wage levels \tilde{w} and w gives the average effect of introducing the wage subsidy. The functions h^{**} and h^* can be estimated nonparametrically using a method such as kernel, local linear regression or series estimation.⁷

The proposed estimation procedure can be viewed as a matching estimator.⁸ To make the analogy transparent, it is useful to transform the model into the potential outcomes notation commonly adopted in the treatment effect literature. Define $Y_1 = h^{**}$ and $Y_0 = h^*$. Also, let $D = 1$ if treated (receives the subsidy). A typical matching estimator (e.g. Rosenbaum and Rubin, 1983) would assume that there exists a set of observables Z such that

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i \mid Z_i,$$

The conventional matching approach is not useful for ex ante evaluation, because it requires data on Y_1 , which is not observed. However, a modified version of matching is possible, using the fact that the economic model along with the restriction on the distribution of μ implies

$$Y_{1i} = Y_{0j} \mid A_i = A_j, \tau w_i = w_j. \tag{1}$$

This identification assumption is inherently different from the types of assumptions typically invoked to justify matching estimators. Nonetheless, this condition motivates a matching estimator for average program effects of the form:

$$\frac{1}{n} \sum_{\substack{j=1 \\ j, i \in S_P}}^n Y_{0i}(w_i = w_j \tau, A_i = A_j) - Y_{0j}(w_j, A_j),$$

⁷There may be ranges over which the support of w and the support of \tilde{w} do not overlap. For persons whose w or \tilde{w} fall in such ranges, it is not possible to evaluate the program's impact. In that case, the average impact based on the matched samples may differ from the population impact. See Ichimura and Taber (1998) for more discussion on this point.

⁸Ichimura and Taber (1998) also draw an analogy between their proposed method of nonparametrically recovering policy impacts and matching.

where $Y_{0j}(w_j, A_j)$ denotes the hours of work choice for an individual j with set of characteristics (w_j, A_j) , $Y_{0i}(w_i = w_j\tau, A_i = A_j)$ the hours of work choice for a matched individual with characteristics $(w_j\tau, A_j)$. The matches can only be performed for the n individuals whose w values and associated $w_j\tau$ values both lie in the overlapping support region, denoted S_p , where

$$S_p = \{\tilde{w} \text{ such that } f_w(\tilde{w}) > 0\}.$$

A distinction between this approach and conventional matching approaches is that here particular functions of observables are equated, whereas conventional matching estimators equate the observables directly.

The above example shows that it is possible to estimate the impact of the policy without having to specify the functional form either of the structural equations or the reduced form hours equation. The key assumptions are that (i) the subsidy only operates through the budget constraint and does not directly affect utility and (ii) that the unobserved heterogeneity is not systematically related to wages, conditional on asset levels. In general, this approach could break down if we allowed the subsidy to affect utility directly ($U = U(c, 1 - h, \tau)$), in a way that leads to a violation of the condition that $\eta(w, A, \tau) = \varphi(\tilde{w}, A)$. Whether such a violation occurs will depend on the specific functional form of the utility function. For example, it is straightforward to show that if any affine transformation of the utility function is additively separable in τ , ($U(c, 1 - h) + v(\tau)$), then it is possible to estimate the effect of the policy nonparametrically, even if τ directly affects utility. This would allow, for example, for a "feel good" or a stigma effect from receiving the subsidy.

It is possible to relax somewhat the assumption on the distribution of the unobserved heterogeneity, for example, if one were willing to assume that

$$f(\mu|w, A, x) = f(\mu|A, x),$$

for some observables x , i.e. that unobserved heterogeneity were ignorable conditional on assets A and a set of observables x that might be assumed to affect utility or wage functions.

Finally, although we have discussed the example in terms of a wage subsidy, the same

analysis could be applied if τ were a tax instead of a subsidy. In the case of a tax, the function $v(\tau)$ might represent a psychic benefit or cost that people get from paying taxes.⁹ Also, while we have focused on hours as the main outcome of interest, the outcome of interest could also be the work decision, which is just a transformation of hours of work (i.e. $1(h^* > 0)$). In a related paper, Blomquist and Newey (2002) develop a nonparametric method that can be used to analyze the effect of changes in a nonlinear tax schedule on hours worked.

An additive wage subsidy program We next consider ex ante evaluation under alternative subsidy schemes. Consider the same set-up as before, but now assume that the subsidy to wages is additive instead of multiplicative. In this case, the constraint (with subsidy) becomes

$$c = hw + h\tau + A,$$

which we can write as

$$c = h(w + \tau) + A.$$

Thus, we have

$$h^{**} = \eta(w, A, \tau, \mu) = \varphi(\tilde{w}, A, \mu),$$

where $\tilde{w} = w + \tau$. An estimation strategy identical to that in the previous example could be used, except that now untreated individuals with wages $\tilde{w} = w + \tau$ and assets A are matched to untreated individuals with wages w and assets A . The required assumption on the distribution of observed heterogeneity is

$$f(\mu|w, A) = f(\mu|\tilde{w}, A)$$

Additional conditioning on some set observables could again be introduced.

⁹It could also represent the benefits that people derive from public goods provided by the total taxes collected, where we would have to assume that an individual does not take into account his small contribution to the total taxes collected when deciding on labor supply.

An income transfer program Now, consider a program that does not alter wages, but supplements income by an amount τ . In this case, the budget constraint becomes

$$c = hw + \tau + A,$$

which can be written as

$$c = hw + \tilde{A},$$

where $\tilde{A} = A + \tau$. Thus,

$$h^{**} = \eta(w, A, \tau, \mu) = \varphi(w, \tilde{A}, \mu).$$

In this case, the estimation strategy matches untreated individuals with wages and assets equal to w and A to other untreated individuals with wages and assets equal to w and \tilde{A} , and the required assumption on the distribution of μ is

$$f(\mu|w, A) = f(\mu|w, \tilde{A}).$$

In this case, the distribution of unobserved heterogeneity can depend on wages but not asset levels.

A combination wage subsidy and income transfer Suppose a program provides an earnings supplement in the amount τ_1 and at the same time an additive wage subsidy in the amount τ_2 . The budget constraint takes the form

$$\begin{aligned} c &= h(w + \tau_1) + A + \tau_2 \\ &= h\tilde{w} + \tilde{A}, \end{aligned}$$

where $\tilde{w} = w + \tau_1$ and $\tilde{A} = A + \tau_2$. To obtain nonparametric estimates of program impacts through matching, untreated individuals with values of wages and assets equal to (\tilde{w}, \tilde{A}) can be matched to other untreated individuals with values of wages and assets equal to (w, A) , under the assumption that the unobserved heterogeneity distribution does not depend on wages or assets

$$f(\mu|w, A) = f(\mu|\tilde{w}, \tilde{A}).$$

This condition is a stronger assumption than was required in the previous two examples. Interestingly, in this case, matching can be used to estimate program effects, but none of the observables are actually equated.

3.2 School attendance subsidy programs

In recent years, many governments in developing countries have adopted school subsidy programs and other conditional cash transfer programs as a way to alleviate poverty and stimulate human capital investment. Programs that condition cash transfers on school attendance currently exist in Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, Honduras, Mexico, Nicaragua, Peru and Uruguay.¹⁰

We next consider how to do an ex-ante evaluation of the effects of a school subsidy programs. We assume that the data contains no direct variation in the price of schooling, so that other sources of variation must be used. The model is motivated in part by a model presented in Todd and Wolpin (2006). The application in that paper was to evaluating the effect of the PROGRESA program, which was introduced in Mexico in 1997 as a means of increasing school enrollment and reducing child labor. In the example that follows, child wages play a crucial role in identifying school subsidy effects. In the first variant of the model, we assume that child wage offers are observed. In the second variant, we assume child wages are not observed.

School attendance subsidy when child wage offers are observed Consider a household making a single period decision about whether to send a single child to school or to work. Household utility depends on consumption (c) and on whether the child attends school (indicator s). A child that does not attend school is assumed to work in the labor market at wage w (below we consider an extension to allow for leisure as another option for child time). Letting y denote household income, net of the child's earnings, the household solves

¹⁰Bangladesh has adopted a similar kind of program that conditions food transfers on school attendance.

the problem:

$$\begin{aligned} & \max_{\{s\}} U(c, s, \mu) \\ & s.t. \\ & c = y + w(1 - s). \end{aligned}$$

In this example, the optimal choice $s^* = \varphi(y, w, \mu)$, where μ denotes unobservable heterogeneity.

Now consider the effects of a policy that provides a subsidy in the amount τ for school attendance, so that the problem becomes:

$$\begin{aligned} & \max_{(s)} U(c, s, \mu) \\ & s.t. \\ & c = y + w(1 - s) + \tau s. \end{aligned}$$

The constraint of the model can be rewritten as

$$c = (y + \tau) + (w - \tau)(1 - s),$$

which shows that the optimal choice of s in the presence of the subsidy is $s^{**} = \varphi(\tilde{y}, \tilde{w}, \mu)$, where $\tilde{y} = y + \tau$ and $\tilde{w} = w - \tau$. That is, the schooling choice for a family with income y , child wage w and unobserved heterogeneity μ that receives the subsidy is, under the model, the same as the schooling choice for a family with income \tilde{y} and child wage \tilde{w} .

Estimation Under the assumption that :

$$f(\mu|y, w) = f(\mu|\tilde{y}, \tilde{w}),$$

we can estimate the effect of the subsidy program on the proportion of children attending school by comparing children from families with income \tilde{y} and child wage offers \tilde{w} to children from families with income y and child wages w . The assumption is clearly a stringent condition, as family preferences for schooling are likely correlated with factors affected family

income. To make the independence assumption on unobservables more plausible, one could in addition condition also on a vector of family characteristics, denoted by x , which might include measures of family background such as parents education, and assume that:

$$f(\mu|y, w, x) = f(\mu|\tilde{y}, \tilde{w}, x).$$

A matching estimator of average program effects for those offered the program (the so-called "intent-to-treat" or ITT estimator) takes the form

$$\frac{1}{n} \sum_{\substack{j=1 \\ j, i \in S_P}}^n \{E(s_i|w_i = w_j - \tau, y_i = y_j + \tau) - s_j(w_j, y_j)\},$$

where $s_j(w_j, A_j)$ denotes the school attendance decision for a child of family j with characteristics (w_j, y_j) . The average can only be taken over the region of overlapping support S_P , which in this case is over the set of families j for which the values $w_j - \tau$ and $y_j + \tau$ lie within the observed support of w_i and y_i .

Using the same reasoning, we can evaluate the effects of a range of school subsidy programs that have both an income subsidy and a schooling subsidy component. Nonparametric reduced form policy variation is feasible in this case, even when there is no variation in the data in the policy instrument (the direct price of schooling).

In this example, not all families choose to participate in the subsidy program. Because the costs of the program will depend on how many families participate in it, a key question of interest in designing the program pertains to the coverage rates and costs of alternative hypothetical programs. In this case, the coverage rate is the probability that a family takes up the subsidy program or, in other words, sends their child to school when the subsidy program is in place:

$$\begin{aligned} & \Pr(s(w - \tau, y + \tau) = 1) \\ &= E(s(w - \tau, y + \tau)) \end{aligned}$$

This probability can be estimated by a nonparametric regression of the indicator variable s on w and y , evaluated at the points $w - \tau, y + \tau$. This estimation can only be performed

for families whose w and y values fall within the region of overlapping support, because nonparametric estimation does not provide a way of extrapolating outside the support region. Taking averages across the probability estimates for all families provides an estimate of the overall predicted take-up rate.

Using the ITT estimate and the take-up rate estimate, we can obtain an estimate of the average impact of treatment on the treated (TT). The relationship between ITT and TT for a family with characteristics (w, y) is:

$$ITT(w, y) = \Pr(\text{participates in program} | w, y)TT(w, y) + \Pr(\text{does not participate} | w, y)0.^{11}$$

Thus,

$$TT(w, y) = \frac{ITT(w, y)}{E(s(w - \tau, y + \tau))}.$$

To obtain an overall average estimate of the impact of treatment on the treated, we integrate over the distribution of w and y values that fall within the support region. Empirically, this can be done by simply averaging over the TT estimates for each of the individual families (within the support region):

$$\frac{1}{n} \sum_{\substack{j=1 \\ j, i \in SP}}^n \frac{\{E(s_i | w_i = w_j - \tau, y_i = y_j + \tau) - s_j(w_j, y_j)\}}{E(s_i | w_i = w_j - \tau, y_i = y_j + \tau)}.$$

The above model assumed that parental utility depends directly on child schooling; a modified version would allow parental utility to be a function of children's future wages (w^f), which in turn depends on schooling levels ($U(c, w^f(s))$).

Extension to Multiple Children with Exogenous Fertility The above model also assumed that parents were making decisions about one child. A straightforward modification is to allow for exogenous fertility and multiple children. For example, suppose there are two children in the family who are eligible for subsidies τ^1 and τ^2 , have wage offers w^1 and w^2 , and for which the relevant schooling indicators are s^1 and s^2 . (Children of different ages/gender

might receive different levels of subsidies). Then, the problem becomes:

$$\begin{aligned} & \max_{(s^1, s^2)} U(c, s^1, s^2) \\ & s.t. \\ c &= (y + \tau^1 + \tau^2) + (w^1 - \tau^1)(1 - s^1) + (w^2 - \tau^2)(1 - s^2). \end{aligned}$$

Estimation of the subsidy effect on enrollment requires matching families with the same configuration of children. In this case, families with income level y and child wages w^1 and w^2 are matched to other families with income level $\tilde{y} = (y + \tau^1 + \tau^2)$ and child wage offers $\tilde{w}^1 = w^1 - \tau^1$ and $\tilde{w}^2 = w^2 - \tau^2$.

Multiple Children with Endogenous Fertility The ex-ante evaluation procedure can also accommodate endogenous fertility, under the maintained assumption of no unobserved heterogeneity. Let n denote the number of children and s_i the schooling decision for child i . For simplicity, assume the potential earnings and subsidy level for each child is the same. Assuming that parents get utility over the number of children and their children's schooling levels, the model is given by

$$\begin{aligned} & \max_{(n, s^1, \dots, s^n)} U(c, n, s^1, \dots, s^n) \\ & s.t. \\ c &= (y + n\tau) + (w - \tau) \sum_{i=1}^n (1 - s^i), \end{aligned}$$

where w is the per child potential wage and τ is the subsidy. Parents decide on the number of children and on schooling decisions, and both decisions are potentially affected by the subsidy level. The expected fertility for a family with income y facing wage w in the absence of the subsidy is:

$$\sum_{j=1}^J j \Pr(n = j | \tilde{y} = y, \tilde{w} = w),$$

where j indexes the range of potential numbers of children. The expected schooling level can be written as:

$$\sum_{j=1}^J \Pr(s = 1 | \tilde{y} = y, \tilde{w} = w, \tilde{n} = j) \Pr(\tilde{n} = j | \tilde{y} = y, \tilde{w} = w) \quad (2)$$

With the subsidy, the expected number of children is

$$\sum_{j=1}^J j \Pr(\tilde{n} = j | \tilde{y} = y + j\tau, \tilde{w} = w - \tau),$$

and the expected schooling level is:

$$\sum_{j=1}^J \Pr(s = 1 | \tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j) \Pr(\tilde{n} = j | \tilde{y} = y + j\tau, \tilde{w} = w - \tau). \quad (3)$$

Noting that

$$\Pr(s = 1 | \tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j) = E(I(\tilde{s} = 1) | \tilde{y} = y + j\tau, \tilde{w} = w - \tau, \tilde{n} = j)$$

$$\Pr(\tilde{n} = j | \tilde{y} = y + j\tau, \tilde{w} = w - \tau) = E(I(\tilde{n} = j) | \tilde{y} = y + j\tau, \tilde{w} = w - \tau)$$

the probability expressions appearing in (2) and (3) can be estimated by a nonparametric regressions, where the dependent variables correspond to the indicator functions $I(\tilde{s} = 1)$ and $I(\tilde{n} = j)$. The program effect can be calculated by taking the difference between terms (2) and (3), replacing the probabilities with their corresponding estimators.

An example where nonparametric ex ante policy evaluation is not possible

Suppose we modify the model presented above to allow for an alternative use of children's time, leisure. That is, consider a model of the form:

$$\begin{aligned} & \max_{(s,l)} U(c, l, s) \\ & s.t. \\ & c = y + w(1 - l - s), \end{aligned}$$

where the optimal choice of schooling and leisure is $s^* = \varphi(y, w)$ and $l^* = \lambda(y, w)$. When the family is offered the subsidy, the constraint can be written as

$$\begin{aligned} c &= y + w(1 - l - s) + \tau s \\ &= (y + \tau) + (w - \tau)(1 - s) - (w - \tau)l + \tau l \end{aligned}$$

As seen by the last equation, it is not possible to transform the constraint into one that is solely a function of $\tilde{y} = y + \tau$ and $\tilde{w} = w - \tau$. The optimal choice of s in the presence of the subsidy is a function of \tilde{y}, \tilde{w} and of τ . Because of the dependence on τ , the policy function in the absence of the subsidy will not be the same as in the presence of the subsidy. We can still forecast the effect of the policy, but doing so requires parametric assumptions on the utility function that allow explicit derivation of the policy functions with and without the subsidy.

School attendance subsidy when only accepted child wages are observed Consider the same model, except that now assume that child wage offers are only observed for families who decide not to send their children to school. Also, assume that the cost of attending school depends on the distance to school, denoted k . The maximization problem is

$$\begin{aligned} &\max_{\{s\}} U(c, s) \\ &s.t. \\ c &= y + (1 - s)w - \delta(k)s \\ \ln w &= \mu_w + \varepsilon, \end{aligned}$$

where $\delta(k)$ is the distance cost function and the last equation is the log wage offer equation. The family chooses to send their child to school ($s = 1$) if $U(y - \delta k, 1) > U(y + \mu_w + \varepsilon, 0)$.

Below, we show that we can identify ex-ante treatment effects without having to make a distributional assumption on the utility function. However, we do need to impose a distributional assumption on log wages. Assume that ε is normally distributed with mean 0 and

variance equal to σ_ε^2 and that $f(\varepsilon|y, k) = f(\varepsilon)$. To take into account selectivity in observed wages, write the wage equation as

$$\begin{aligned} \ln w &= \mu_w + E(\varepsilon|s = 0) + \{\varepsilon - E(\varepsilon|s = 0)\} \\ &= \mu_w + E(\varepsilon|U(y + \mu_w + \varepsilon, 0) > U(y - \delta k, 1)) + u \\ &= \mu_w + E(\varepsilon|\varepsilon > \eta(y, k)) + u \end{aligned}$$

where the last equality assumes that U is monotone in s , that u has conditional mean zero by construction and that η is some function of y and k . The conditional mean function can be written as

$$E(\varepsilon|\varepsilon > \eta(y, k)) = \frac{\int_{\eta(y, k)}^{\infty} \varepsilon f(\varepsilon) d\varepsilon}{\int_{\eta(y, k)}^{\infty} f(\varepsilon) d\varepsilon},$$

where we use the assumption that $f(\varepsilon) = f(\varepsilon|y, k)$. Using the fact that

$$\begin{aligned} \Pr(s = 1|y, k) &= \Pr(\varepsilon > \eta(y, k)) \\ &= 1 - \Phi(\eta(y, k)). \end{aligned}$$

The normal cdf Φ is invertible, so we can write $\eta(y, k) = 1 - \Phi^{-1}(P) = K(P)$, where $P = \Pr(s = 1|y, k)$. We can obtain a nonparametric estimate of the conditional probability of attending school from a nonparametric regression of s on y and k . The equation for observed wages can now be written as:

$$\begin{aligned} \ln w &= \mu_w + \sigma_\varepsilon \frac{\phi\left(\frac{K(P)}{\sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{K(P)}{\sigma_\varepsilon}\right)} + u \\ &= \mu_w + \sigma_\varepsilon \lambda\left(\frac{K(P)}{\sigma_\varepsilon}\right) + u \end{aligned}$$

where $\lambda(\cdot)$ is the Mill's ratio function and K is the function defined above. Once we construct the Mill's ratio regressor, the parameters μ_w and σ_ε can be estimated using least squares. (See Heckman, 1979). Thus, we obtain estimates of μ_w and of σ_ε , the parameters of the density of the child log wage offer equation, $\phi(w)$.

To evaluate ex ante program impacts using matching, we require an estimate of $\Pr(s =$

$1|y, w, k)$ for alternative values of y , w and distance k . Use the fact that

$$\begin{aligned}
\Pr(s = 1|y, w, k) &= 1 - \Pr(s = 0|y, w, k) \\
&= 1 - \frac{f(w, y, k|s = 0) \Pr(s = 0)}{f(w, y, k)} \\
&= 1 - \frac{f(w, y, k|s = 0) \Pr(s = 0)}{g(w|y, k)g(y, k)} \\
&= 1 - \frac{f(w, y|s = 0) \Pr(s = 0)}{\tilde{\phi}(w)g(y, k)},
\end{aligned}$$

where $\tilde{\phi}(w)$ is the density of wages (log normal with parameters μ_w and σ_ε). The conditional density $f(w, y, k|s = 0)$, the joint density $g(y, k)$, and the unconditional probability $\Pr(s = 0)$ can all be nonparametrically estimated directly from the data, providing a way of estimating $\Pr(s = 1|y, w, k)$.

The matching estimator, for a subsidy of level τ , is implemented as

$$\frac{1}{n} \sum_{\substack{j=1 \\ j, i \in S_P}}^n \{\Pr(s_i = 1|w_i = w_j - \tau, y_i = y_j + \tau, k_i = k_j) - \Pr(s_j = 1|w_j, y_j, k_j)\},$$

where the probabilities are estimated by the above procedure.

Extension to a Two-Period Model Next, we consider an extension of the school subsidy example (with observed wage offers) to a two period model with perfect foresight, assuming a budget constraint that permits borrowing over time. For simplicity, we omit the unobserved heterogeneity since the treatment would be the same as in the previous examples. The price of consumption is assumed to be constant over time. The subsidy for school attendance is τ_1 in the first period and τ_2 in the second time period. y_i denotes family income net of child income and w_i denotes child wages in period i . The problem without the subsidy is

$$\begin{aligned}
&\max_{\{c_1, c_2, s_1, s_2\}} U(c_1, c_2, s_1, s_2) \\
&s.t. \\
c_1 + c_2 &\leq y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2).
\end{aligned}$$

The schooling choices in each period can be written as functions

$$\begin{aligned} s_1 &= \varphi_1(\hat{y}, w_1, w_2) \\ s_2 &= \varphi_2(\hat{y}, w_1, w_2) \end{aligned}$$

where $\hat{y} = y_1 + y_2$.

With the subsidy, the constraint becomes

$$\begin{aligned} c_1 + c_2 &= y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2) + \tau_1 s_1 + \tau_2 s_2 \\ &= (y_1 + \tau_1 + y_2 + \tau_2) + (w_1 - \tau_1)(1 - s_1) + (w_2 - \tau_2)(1 - s_2), \end{aligned}$$

so that the optimal schooling choices are

$$\begin{aligned} s_1^* &= \varphi_1(\tilde{y}, \tilde{w}_1, \tilde{w}_2) \\ s_2^* &= \varphi_2(\tilde{y}, \tilde{w}_1, \tilde{w}_2), \end{aligned}$$

where $\tilde{y} = y_1 + \tau_1 + y_2 + \tau_2$, $\tilde{w}_1 = w_1 - \tau_1$, and $\tilde{w}_2 = w_2 - \tau_2$. Estimation of program effects requires matching untreated families with two-period earnings equal to $y_1 + y_2$ to other families with two-period earnings equal to \tilde{y} . Matching would also have to be performed on the basis of the wage profile.

Extension #1: Next, consider a modification of the previous example to allow for a subsidy that is increasing in the total number of years of schooling attained. Thus, the amount of the subsidy in the second period depends on the first period schooling decision. Suppose the subsidy is τ_2 if $s_1 = 0$ and $s_2 = 1$, and it is τ_3 if $s_1 = 1$ and $s_2 = 1$. The constraint in this case is

$$\begin{aligned} c_1 + c_2 &= y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2) + \tau_1 s_1 + \tau_2 s_2(1 - s_1) + \tau_3 s_1 s_2 \\ &= y_1 + y_2 + w_1(1 - s_1) + w_2(1 - s_2) + \tau_1 s_1 + \tau_2 s_2 + (\tau_3 - \tau_2) s_1 s_2 + [\tau_1 - \tau_1 + \tau_2 - \tau_2] \\ &= \{y_1 + \tau_1 + y_2 + \tau_2\} + (w_1 - \tau_1)(1 - s_1) + (w_2 - \tau_2)(1 - s_2) + (\tau_3 - \tau_2) s_1 s_2 \end{aligned}$$

In this case, it is generally not possible to transform the constraint into the one of the original problem. However, if the wage level in the second period depended on whether the

individual attended school in period one and there was variation across families in the wage return from schooling (i.e. the value of w_2 and how it varies depending on whether attended school in period one), then it would be possible to transform the model into a version of the model without the subsidy. Although conceptually feasible, the data are unlikely to contain sufficient variation in w_1 , w_2 and in the return from schooling.¹²

Extension #2 Now, consider the model from example one, but assume that borrowing against future income is not allowed, so that the constraints (without subsidies) can be written as

$$\begin{aligned} c_1 &\leq y_1 + w_1(1 - s_1). \\ c_2 &\leq y_2 + w_2(1 - s_2). \end{aligned}$$

In this case, the optimal choice of s_1 and s_2 depends on the profile of earnings and of wages:

$$\begin{aligned} s_1 &= \varphi_1(y_1, y_2, w_1, w_2) \\ s_2 &= \varphi_2(y_1, y_2, w_1, w_2) \end{aligned}$$

It is straightforward to verify that with the subsidy, we get

$$\begin{aligned} s_1^* &= \varphi_1(\tilde{y}_1, \tilde{y}_2, \tilde{w}_1, \tilde{w}_2) \\ s_2^* &= \varphi_2(\tilde{y}_1, \tilde{y}_2, \tilde{w}_1, \tilde{w}_2), \end{aligned}$$

where $\tilde{y}_i = y_i + \tau_i$ and $\tilde{w}_i = w_i - \tau_i$. In this case, nonparametric estimation of policy effects requires matching on the earnings and wage profiles.

4 Empirical application: predicting effects of a school subsidy program

In this section, we apply the previously described methods to analyze the effects of the cash transfer program PROGRESA that was introduced in Mexico in 1997. The program

¹²The case where there is no subsidy in the first period and the final subsidy depends on the total number of years of schooling accumulated ($s_1 + s_2$) (e.g. a graduation bonus) can be viewed as a special case of this model.

provides transfers to families that are contingent upon their children regularly attending school.¹³ These transfers are intended to alter the private incentives to invest in education by offsetting the opportunity cost of not sending children to school. Similar conditional cash transfer programs now exist in Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, Honduras, Mexico, Nicaragua, Peru and Uruguay. Mexico was the first country to evaluate such a program using a randomized experimental design.¹⁴

Table 1 shows the schedule of benefits, which depends on the child's grade level and gender. In recognition of the fact that older children are more likely to engage in family or outside work, the transfer amount increases with the child's grade level and is greatest for secondary school grades. The benefit level is also slightly higher for girls, who traditionally have lower school enrollment levels.

To participate in the program, families have to satisfy some eligibility criteria, which depend on factors such as whether their home has a dirt floor, crowding indices, and ownership of assets (e.g. car). In total, the benefit levels that families receive under the program is substantial relative to their income levels, about 20-25% of total income. (Skoufias and Parker, 2000) Almost all the families that are offered the program participate in it to some extent.¹⁵ Partial participation is possible if the family only sends some children to school but not others.

The PROGRESA program was initially introduced in rural areas, has since expanded into semi-urban and urban areas, and currently covers about one quarter of all Mexican families. For purposes of evaluation, the initial phase of PROGRESA was implemented as a social experiment, in which 506 rural villages were randomly assigned to either participate in the program or serve as controls.¹⁶ Randomization, under ideal conditions, allows mean

¹³The program also provides a small transfer to the family contingent on visiting a health clinic for check-ups as well as nutritional supplements for children under the age of two. We ignore this other component of the program and focus on the school subsidies, which are by far the largest component for most families.

¹⁴The most recent incarnation of the Mexican program is called *Oportunidades*.

¹⁵In the rural villages that participated in the initial PROGRESA experiment, all the households were interviewed and informed of their program eligibility status.

¹⁶Data are available for all households located in the 320 villages assigned to the treatment group and for all households located in the 186 villages assigned to the control group.

program impacts to be assessed through simple comparisons of outcomes for the treatment and control groups. Schultz (2000a,2000b) and Behrman, Sengupta and Todd (2005) investigate the program’s experimental impacts on school enrollment and find significant impacts, particularly for children in secondary school grades.(7th-9th grade)

In this paper, we also use data from the PROGRESA experiment, but with a focus on studying the ex ante evaluation methods. As noted in the introduction, our strategy is to predict the impacts of the program only using data on the randomized-out control group, and then compare the predictions to the actual impacts estimated under the experiment.

4.1 Data sample

The data gathered as part of the PROGRESA experiment provide rich information at the individual, the household and the village level. The data include information on school attendance and grade attainment for all household members and information on employment and wages for individuals age eight and older. The data we analyze were gathered through a baseline survey administered in October, 1997 and follow-up survey administered in October, 1998. In the fall of 1998, households in the treatment group had been informed of their eligibility and began receiving subsidy checks. Control group households did not receive benefits over the course of the experiment.¹⁷

From the household survey datasets, we use information on the age and gender of the child, the child’s highest grade completed, whether the child is currently enrolled in school, and income of the mother and father. Total family income is obtained as the sum of the husband’s and the wife’s earnings, including income from main jobs as well as any additional income from second jobs. Our analysis subsample includes children age 12 to 15 in 1998, who are reported to be the son or daughter of the household head, and for whom information is available in the 1997 and 1998 surveys. In addition to the household survey datasets, supplemental data were gathered at the village level. Most importantly, for our purposes,

¹⁷The control group was incorporated two years later, but at the time of the experiment they were not told of the plans for their future incorporation.

information is available on the minimum wage paid to day laborers in each village, which we take as a measure of the potential earnings of a child laborer. This information was available for roughly half the villages in the sample.

The upper panel of Figure 1 shows a histogram of the minimum monthly laborer wages, which range from 330 to 1320 pesos per month with a median of 550 pesos.¹⁸ The lower panel of the figure shows a histogram of family income, with values ranging from 8 to 13,750 pesos (median: 660). For many families meeting the program eligibility criteria, the total monthly earnings are not much above that of a full-time worker working at the minimum laborer wage.

4.2 Estimation and empirical results

We predict the impact of the PROGRESA subsidy program on school enrollment using the methods and modeling frameworks that were described in section 3.2. The two modeling frameworks are the multiple child and single child models developed in section 3.2. The estimation method we implement here is somewhat more general, because it allows the school enrollment decision to potentially differ for girls and boys, which would accommodate, for example, difference in the utility that parents get from girls' and boys' schooling.

For the single child model, the estimator of the predicted program effect is given by

$$\hat{\alpha} = \frac{1}{n} \sum_{\substack{j=1 \\ j,i \in S_P}}^n \{E(s_i | w_i = w_j - \tau_j, y_i = y_j + \tau_j, g_i = g_j) - s_j(w_j, y_j, g_j)\},$$

where g_i denotes the child's gender, s_j is an indicator for whether child j is enrolled in school, w_j is the wage offer, and y_j is family income (net of child income). This estimator matches program eligible control group children with offered wage w_j and family income y_j to other control group children with offered wage $w_j - \tau_j$ and $y_i = y_j + \tau_j$, with the matches restricted to be between children of the same gender.¹⁹ The first term $E(s_i | w_i =$

¹⁸Approximately 10 pesos equals 1 US dollar.

¹⁹

The sum is taken over program-eligible children, but the matches are nonparametrically estimated also using children from families who are not necessarily program-eligible. Noneligibles need to be included, because

$w_j - \tau_j, y_i = y_j + \tau_j, g_i = g_j$) is each child's predicted outcome with the program and the second term $s_j(w_j, y_j, g_j)$ is the actual enrollment decision in the absence of the program (i.e. for the program-eligible control group children).

In the above equation, τ_j represents the subsidy level for which the child is eligible. Because subsidies vary by grade level, children of the same age can be eligible for different subsidy levels.²⁰ We therefore use the information in the data about each child's highest grade completed to determine the subsidy level for which the child is potentially eligible.

We estimate the matched outcomes $E(s_i | w_i = w_j - \tau_j, y_i = y_j + \tau_j, g_i = g_j)$ nonparametrically using a two dimensional kernel regression estimator. Letting $w_0 = w_j - \tau_j$ and $y_0 = y_j + \tau$, the estimator is given by

$$E(s_i | w_i = w_0, y_i = y_0, g_i = g_0) = \frac{\sum_{\substack{i=1 \\ i \in S_P}}^n s_i K\left(\frac{w_i - w_0}{h_n^w}\right) K\left(\frac{y_i - y_0}{h_n^y}\right) 1(g_i = g_0)}{\sum_{\substack{i=1 \\ i \in S_P}}^n K\left(\frac{w_i - w_0}{h_n^w}\right) K\left(\frac{y_i - y_0}{h_n^y}\right) 1(g_i = g_0)},$$

where $K(\cdot)$ denotes the kernel function and h_n^w and h_n^y are the smoothing (or bandwidth) parameters. We use a biweight kernel function:

$$\begin{aligned} K(s) &= (15/16)(s^2 - 1)^2 \text{ if } |s| \leq 1 \\ &= 0 \text{ else,} \end{aligned}$$

which satisfies the standard assumptions $\int K(s)ds = 1$, $\int K(s)sds = 0$, and $\int K(s)s^2ds < \infty$. Asymptotic consistency of this estimator requires that the smoothing parameters satisfy $nh_n^w h_n^y \rightarrow \infty$, $h_n^w \rightarrow 0$ and $h_n^y \rightarrow 0$ as $n \rightarrow \infty$.²¹

The nonparametric estimator is only defined at points where the data density is positive.

For this reason, we need restrict the estimation to points of evaluation that lie within the

augmenting family income by the level of the subsidy could change a family's eligibility status. For the PROGRESA program, eligibility was not directly based on father income, but it was based in part on assets and housing characteristics that are correlated with income.

²⁰In Mexico, it is fairly common for children of a given grade level to vary a lot by age.

²¹See, e.g., Härdle and Linton (1994), Ichimura and Todd (2007).

region S_P , where $S_P = \{(w, y) \in R^2 \text{ such that } f(w, y) > 0\}$ and $f(w, y)$ is the density. We determine empirically whether a particular point of evaluation (w_0, y_0) lies in S_P , by estimating the density at each point and checking whether it lies above a cut-off trimming level, q_α , that is small and positive. That is, we check whether

$$\hat{f}(w_0, y_0) > q_\alpha,$$

where $\hat{f}(\cdot, \cdot)$ is a nonparametric estimate of the density. The cut-off level q_α corresponds to the 2% quantile of the positive estimated density values.²²

Next, we describe how we implement the multiple child model. For the multiple child case, we consider the potential earnings of all children in the family age twelve or older (very few children under age twelve work for wages). If all the children within a family had the same subsidy levels and potential wages, the estimator for the program effect would be given by:

$$\hat{\alpha} = \frac{1}{n} \sum_{\substack{j=1 \\ j, i \in S_P}}^n \{E(s_i | w_i = w_j - \tau_j, y_i = y_j + n\tau_j, n_i = n_j, g_i = g_j) - s_j(w_j, y_j, n_j, g_j)\},$$

where n_j denotes the number of children in the family of child j who can potentially earn wages and matches are restricted to families with the same numbers of children. This estimator needs to be slightly modified to take into account that different children within the same family face different potential subsidies. Let $\bar{\tau}_j$ denote the average subsidy level offered to the children in the family of child j . The village minimum wage, which we take to represent the child's potential wages, does not vary within families, so we have to assume that the wage offer is the same for all children of working age within a family. The estimator that we use is given by:

$$\hat{\alpha} = \frac{1}{n} \sum_{\substack{j=1 \\ j, i \in S_P}}^n \{E(s_i | w_i = w_j - \bar{\tau}_j, y_i = y_j + n\bar{\tau}_j, n_i = n_j, g_i = g_j) - s_j(w_j, y_j, n_j, g_j)\}.$$

²²This procedure is similar to that used in Heckman, Ichimura and Todd (1997).

Tables 2a compares the predicted program impacts on the fraction of children enrolled in school obtained by the ex ante prediction method to the corresponding experimental impact estimates for boys and girls for the multiple child model. Impacts are estimated separately over three different age ranges and separately for boys and girls. For the estimation results that combine boys and girls or different age ranges still restrict matches to be between children of the same gender and the same age bracket. That is, a girl age 12-13 would only be matched to other girls in the same age, even for the results that aggregate across categories.²³ The sample sizes (of the eligible controls and of all controls) is shown in column three and the percentage of observations that lie within S_P is shown in column four.

For boys, the experimental impact estimates are all positive but are statistically significant only for the age 12-13 age range. For girls, the impact estimates are positive for both girls and boys and statistically significant at conventional levels for the age 14-15 and 12-15 age ranges. The ex-ante predicted impacts are also all positive, even though the estimation procedure does not constrain them to be positive, and are statistically significantly different from zero for the age 14-15 and age 12-15 age ranges. The predicted impacts tend to overstate the actual impacts for boys, but for girls they are quite close and exhibit the same pattern as the experimental impacts. The estimates that combine girls and boys tend to have lower standard errors due to larger sample sizes. Again, the predicted impacts are similar in magnitude to those observed under the experiment and exhibit a similar pattern, with larger impacts for the older age range (14-15). The overall predicted impact of 0.07 for boys and girls age 12-15 comes very close to the experimental impact of 0.06.

Table 2b reports the ex-ante predicted impacts for counterfactual subsidy levels. The first column shows the predictions if we double the subsidy schedule, the second column shows results for the original schedule, and the third column shows results for a 25% reduction in the subsidy level. The percentage of observations in the overlapping support region is given

²³We did not estimate separately by each age, because the sample sizes become too small to be reliable for nonparametric estimation. The bandwidth was set equal to 200.

in parentheses. The results suggest that doubling the subsidy would lead to a substantial increase in impacts only for the oldest age category (14-15). The estimated enrollment effect increases from 0.09 to 0.16 for boys and 0.11 to 0.15 for girls. Similarly, reducing the subsidy by 25% would roughly halve the impacts for the age 14-15 group. The estimates suggest that school enrollment of the youngest age group (12-13) is relatively insensitive to changes in the subsidy level. As seen in parentheses, the fraction of observations that lie outside of S_P decreases at higher levels of the subsidy, and increases at smaller subsidy amounts.²⁴

Tables 3(a) and 3(b) show analogous results for the single child model. The estimated predicted impacts are for the most part similar to those for the multiple child model and exhibit similar age patterns. A comparison of Table 2(b) and Table 3(b) shows that the predicted response to doubling the subsidy level is larger for the single child model than for the multiple child model. Again, in Table 3(b), we see that the method predicts that doubling the subsidy would lead to a substantial increase in the estimated impacts and that reducing the subsidy by 25% would roughly halve the impacts.

In considering the impact of any unconditional cash transfer program, it is desirable to know to what extent the conditionality makes a difference and whether similar impacts might be achieved through unconditional transfers. Therefore, in Table 4, we use the ex-ante prediction method to explore whether giving families an unconditional income transfer in the amount of 5000 pesos per year would significantly impact school enrollments. This level of transfers is almost half of family income. Table 4 gives the predicted impacts, which suggest that the unconditional income transfer would not lead to any statistically significant impacts on school enrollment. A conventional linear regression of school enrollment on the wage and on income also shows that the wage is a significant determinant of school enrollment but family income has only a negligible impact, at least in the rural villages that comprise our analysis sample.

²⁴The information on the percentage of observations in the support shows how the range of subsidies levels that can be considered is limited by the range of the data. Also, see Ichimura and Taber (1998) for detailed discussion on this point.

5 Conclusions

This paper considered methods for evaluating the impacts of social programs prior to their implementation. Through several examples, we showed how behavioral models can be used to predict impacts of hypothetical programs and to justify particular estimation strategies. In many cases, consideration of the particular structure of the model suggests a fully nonparametric estimation strategy. We illustrated when the conditions for nonparametric policy evaluation are met for different types of program interventions, including wage subsidies, income support programs and schooling subsidies of the kind that have been recently implemented in many South American countries. In some cases, the conditions for nonparametric policy evaluation were not met and stronger assumptions are required.

This paper also suggested some simple estimation strategies, which are modified versions of matching estimators, and studied their performance. The estimators compare untreated individuals to other untreated individuals, where the set of variables on which the individuals are matched is derived from the behavioral model. Our application of these methods considered ex ante evaluation of a school subsidy program, the PROGRESA program in Mexico. The availability of experimental data provides a unique opportunity to study the performance of the estimators.

A comparison of the predicted program impacts, obtained using only the control group data, to the experimentally estimated impacts show that the predictions are generally of the correct sign and come within 30% of the experimental impact. The predicted impacts for girls age 12-15 were particularly close in terms of magnitude and age patterns to the experimental impacts, while the predicted impacts for boys tended to overstate the experimental impacts.

We also used the ex-ante prediction method to explore two kinds of counterfactual programs, changing the level of the subsidies and removing the conditionality of the program. The results on changing the subsidy level revealed that school enrollment of the older age

groups (age 14-15) would be sensitive to increasing or decreasing the levels of the subsidies. The young age group (12-13) is relatively unresponsive to subsidy level changes. We also find that conditioning subsidies on schooling is important to the effectiveness of the program. A program that removes the conditionality requirement and instead provides generous unconditional subsidies would not be expected to lead to changes in enrollment.²⁵

²⁵This finding is consistent with simulation results reported in Todd and Wolpin (2006).

References

- [1] Andrews, D. and Schafgans (1998): "Semiparametric Estimation of the Intercept of a Sample Selection Model" in *Review of Economic Studies*, 65, 497-517.
- [2] Behrman, Jere, Sengupta, Piyali, and Petra Todd (2000a): "Progressing through PROGRESA: An Impact Assessment of Mexico's School Subsidy Experiment," unpublished manuscript.
- [3] Behrman, Jere and Petra Todd (2000b): "Randomness in the Experimental Samples of PROGRESA" in International Food Policy Research Institute, Washington, D.C.
- [4] Blomquist, Sören and Whitney Newey (2002): "Nonparametric Estimation with Non-linear Budget Sets" in *Econometrica*, 70 (6), 2455–2480.
- [5] Bourguignon, François, Ferreira, Francisco H. G. and Phillippe G. Leite (2002): "Ex-ante Evaluation of Conditional Cash Transfer Programs: The Case of Bolsa Escola," William Davidson Institute Working Paper.
- [6] Gomez de Leon, J. and Susan Parker (2000): "The Impact of Anti-Poverty Programs on Children's Time Use," International Food Policy Research Institute, Washington, D.C.
- [7] Härdle, Wolfgang and Oliver Linton (1994): "Applied Nonparametric Methods" in *Handbook of Econometrics, Vol. 4*, Elsevier, Amsterdam, p. 2295-2339.
- [8] Heckman, James J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete-Time Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. Manski and D. Mcfadden, 179-197.
- [9] Heckman, James J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," in *Quarterly Journal of Economics*, Vol. 115(1), p.45-97.

- [10] Heckman James, J. and Burton Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.
- [11] Heckman, James J., Hidehiko Ichimura and Petra Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 64(4), 605-654.
- [12] Heckman, James J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture" in *Journal of Political Economy*, vol. 109, no. 4.
- [13] Hurwicz, Leonid (1962): "On the Structural Form of Interdependent Systems." In *Logic, Methodology and Philosophy of Science*, edited by Ernest Nagel, Patrick Suppes and Alfred Tarski. Stanford, Calif.: Stanford University Press.
- [14] Ichimura, Hidehiko and Christopher Taber (1998): "Direct Estimation of Policy Impacts," manuscript, Northwestern University.
- [15] Ichimura, Hidehiko and Christopher Taber (2002): "Semiparametric Reduced-Form Estimation of Tuition Subsidies" in *American Economic Review*. Vol. 92 (2). p 286-92.
- [16] Ichimura, Hidehiko and Petra Todd (2007): "Implementing Semiparametric Estimators," forthcoming in *Handbook of Econometrics*, Volume 6.
- [17] Johnson, Terry R. and John H. Pencavel (1982): "Forecasting the Effects of a Negative Income Tax Program," *Industrial and Labor Relations Review*, Vol. 35, No. 2, p. 221-234.
- [18] LaLonde, Robert (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604-620.
- [19] Lise, Jeremy, Seitz, Shannon, and Jeffrey Smith (2003): "Equilibrium Policy Experiments and the Evaluation of Social Programs," working paper.

- [20] Lumsdaine, Robin L., James H. Stock and David A. Wise (1992): "Pension Plan Provisions and Retirement: Men and Women, Medicare, and Models," in D. A. Wise (ed.) *Studies in the Economics of Aging*, Chicago: University of Chicago Press.
- [21] Manski, C. F. (1988): "Identification of Binary Reponse Models," *Journal of the American Statistical Association*, 83, 403, 729-737.
- [22] Manski, C. F. (1986): "Semiparametric Analysis of Binary Response From Response-Based Samples," *Journal of Econometrics*, 31, 31-40.
- [23] Manski, C. F. and D. McFadden (1981): "Alternative Estimators and Sample Designs for Discrete Choice Analysis" in *Structural Analysis of Discrete Data with Economic Applications*, edited by C.F. Manski and D. McFadden (Cambridge, Mass.: MIT Press) 1-50.
- [24] Manski, C. F. (1975): "The Maximum Score Estimation of the Stochastic Utility Model of Choice," in *Journal of Econometrics*, 27:313-333.
- [25] Marschak, Jacob (1953): "Economic Measurements for Policy and Prediction," in William Hood and Tjalling Koopmans, eds., *Studies in Econometric Method* (New York: John Wiley, 1953), pp. 1-26.
- [26] McFadden, D. (1984): "Econometric Analysis of Qualitative Response Models," *Handbook of Econometrics*, Vol. II, edited by Z. Griliches and M.D. Intriligator.
- [27] McFadden, Daniel and A. P. Talvitie and Associates (1977): "Validation of Disaggregate Travel Demand Models: Some Tests" in Urban Demand Forecasting Project, Final Report, Volume V, Institute of Transportation Studies, University of California, Berkeley.
- [28] Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70,41-55.

- [29] Schultz, T. Paul (2000a): "Progresas's Impact on School Enrollments from 1997/98 to 1998/99," International Food Policy Research Institute, Washington, D.C.
- [30] Schultz, T. Paul (2000b): "Progresas's Impact on School on School Attendance Rates in the Sampled Population," International Food Policy Research Institute, Washington, D.C.
- [31] Skoufias, Emmanuel and Susan Parker (2000): "The Impact of PROGRESA on Work, Liesure and Time Allocation," International Food Policy Research Institute, Washington, D.C.
- [32] Todd, Petra and Kenneth I. Wolpin (2006): "Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico," *American Economic Review*, December.
- [33] Wise, David A. (1985): "A Behavioral Model Verses Experimentation: The Effects of Housing Subsidies on Rent" in *Methods of Operations Research*, 50, Verlag Anton Hain.

Table 1
Monthly Transfers for School Attendance

School Level	Grade	Gender	
		Female	Male
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	7	210	200
	8	235	210
	9	235	225

Table 2(a)
Comparison of Ex-Ante Predictions and Experimental Impacts
Multiple-child model (Bootstrap standard errors in parentheses) †

Ages	Boys			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.05** (0.02)	0.05 (0.03)	374, 610	68%
14-15	0.02 (0.03)	0.09* (0.05)	309, 569	61%
12-15	0.03 (0.02)	0.06** (0.03)	683, 1179	64%
Ages	Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.07 (0.07)	0.04 (0.04)	361, 589	67%
14-15	0.11** (0.04)	0.11* (0.06)	361, 591	68%
12-15	0.09 ** (0.02)	0.07** (0.04)	677, 1180	68%
Ages	Boys and Girls			
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.06** (0.02)	0.04 (0.03)	735, 1199	67%
14-15	0.07** (0.03)	0.10** (0.04)	625, 1160	64%
12-15	0.06** (0.02)	0.07** (0.02)	1360, 2359	66%

†Standard errors based on 500 bootstrap replications. Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

‡The first number refers to the total control sample and the second to the subset of controls that satisfy the PROGRESA eligibility criteria.

Table 2(b)
Effects of Counterfactual Subsidy Levels
Multiple-child model (% in overlap region in parentheses)

Boys			
Ages	2* Original	Original	0.75*Original
12-13	0.01 (50%)	0.05 (68%)	0.01 (92%)
14-15	0.16 (43%)	0.09 (61%)	0.04 (93%)
12-15	0.08 (47%)	0.06 (64%)	0.02 (93%)
Girls			
	2* Original	Original	0.75*Original
12-13	0.04 (48%)	0.04 (67%)	0.04 (93%)
14-15	0.15 (52%)	0.11 (68%)	0.04 (93%)
12-15	0.09 (50%)	0.07 (68%)	0.04 (93%)
Boys and Girls			
	2* Original	Original	0.75*Original
12-13	0.03 (49%)	0.04 (67%)	0.02 (93%)
14-15	0.15 (48%)	0.10** (64%)	0.04 (93%)
12-15	0.08 (49%)	0.07** (66%)	0.03 (93%)

† Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

Table 3(a)
 Comparison of Ex-Ante Predictions and Experimental Impacts
 Single-child model (Bootstrap standard errors in parentheses) †

Boys				
Ages	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.05** (0.02)	0.01 (0.03)	374, 10	87%
14-15	0.02 (0.03)	0.01* (0.04)	309, 569	83%
12-15	0.03 (0.02)	0.06 (0.03)**	683, 1179	86%
Girls				
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.07 (0.07)	0.06* (0.03)	361, 589	91%
14-15	0.11** (0.04)	0.07 (0.05)	316, 589	89%
12-15	0.09 ** (0.02)	0.06** (0.03)	677, 1180	90%
Boys and Girls				
	Experimental	Predicted	Sample-Sizes‡	% overlapping support
12-13	0.06** (0.02)	0.04* (0.02)	735, 1199	89%
14-15	0.07** (0.03)	0.09** (0.04)	625, 1160	86%
12-15	0.06** (0.02)	0.06** (0.02)	1360, 2359	88%

†Standard errors based on 500 bootstrap replications. Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

‡The first number refers to the total control sample and the second to the subset of controls that satisfy the PROGRESA eligibility criteria.

Table 3(b)
Effects of Counterfactual Subsidy Levels
Single-child model (% in overlap region in parentheses)

Boys			
Ages	2* Original	Original	0.75*Original
12-13	0.04 (59%)	0.01 (87%)	0.003 (98%)
14-15	0.24 (45%)	0.01 (83%)	0.05 (98%)
12-15	0.12 (53%)	0.06 (86%)	0.02 (98%)
Girls			
Ages	2* Original	Original	0.75*Original
12-13	0.06 (48%)	0.06 (91%)	0.05 (98%)
14-15	0.23 (51%)	0.07 (89%)	0.03 (98%)
12-15	0.14 (50%)	0.06 (90%)	0.05 (98%)
Boys and Girls			
Ages	2* Original	Original	0.75*Original
12-13	0.05 (54%)	0.04* (89%)	0.03 (98%)
14-15	0.23 (48%)	0.09 (86%)	0.04 (98%)
12-15	0.13 (52%)	0.06 (88%)	0.03 (98%)

† Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

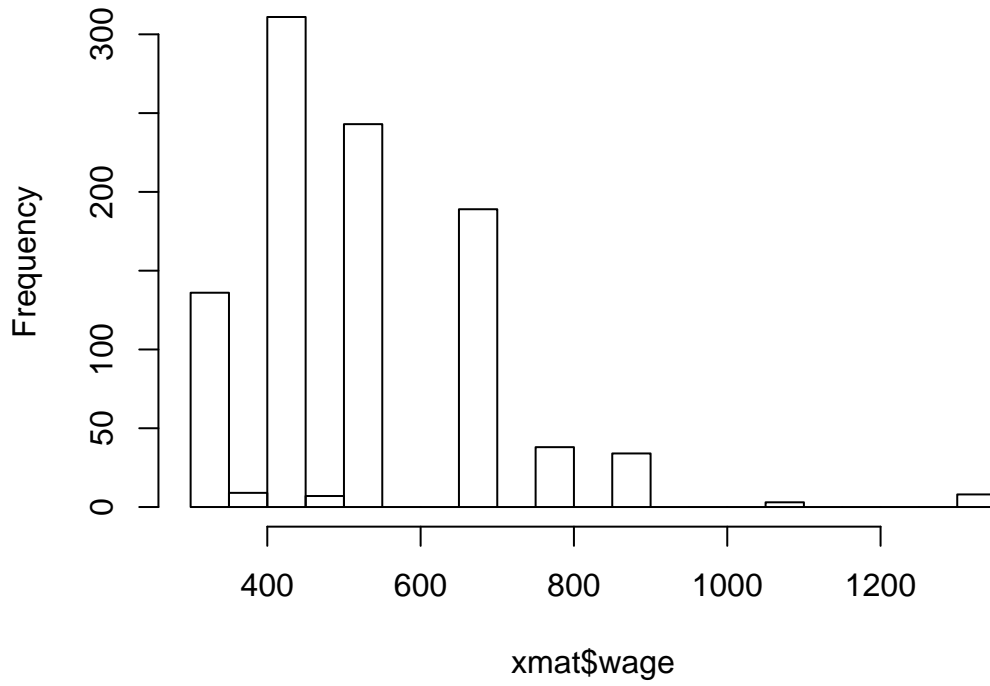
Table 4
 Predicted Impact of an Unconditional Income Transfer in the Amount of 5000 pesos/year
 Multiple-child model (Bootstrap standard errors in parentheses) †

Boys			
Ages	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.02 (0.03)	374, 610	89%
14-15	-0.06 (0.05)	309, 569	90%
12-15	-0.04 (0.03)	683, 1179	89%
Girls			
	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.03 (0.04)	361, 589	88%
14-15	0.00 (0.05)	316, 591	88%
12-15	-0.02 (0.03)	677, 1180	88%
Boys and Girls			
	Predicted	Sample-Sizes‡	% overlapping support
12-13	-0.03 (0.03)	735, 1199	88%
14-15	-0.03 (0.03)	625, 1160	89%
12-15	-0.03 (0.02)	1360, 2359	89%

†Standard errors based on 500 bootstrap replications. Bandwidth equals 200 pesos. Trimming implemented using the 2% quantile of positive density values as the cut-off point.

‡The first number refers to the total control sample and the second to the subset of controls that satisfy the PROGRESA eligibility criteria.

Histogram of Min Monthly Laborer Wage



Histogram of Total Family Income

Figure 1

