Ex Ante Evaluation of Social Programs

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

June 1, 2005

The authors may be contacted at petra@athena.sas.upenn.edu and wolpink@ssc.upenn.edu.
We than Andrew Foster and Susan Parker for helpful comments.
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 through several examples the use of behavioral models in predicting the impacts of hypothetical programs. 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 modeling and/or functional form 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 experiment in Mexico. We assess the effectiveness of the method by comparing ex ante predictions of program impacts to the impacts measured under the randomized experiment.
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 data on individuals that receive the program intervention (the treatment group) as well as data on a comparison group sample that does not receive the intervention. A limitation of these approaches is that they do not provide a way 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, which would implement alternative versions of the program and compare their impacts, is often too costly 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 an idea of 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 an existing 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 that make use of data on the treated group to be more reliable for estimating treatment impacts of an existing program, there is still a critical role for ex ante evaluation tools.

\(^1\)For 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 an economic model is crucial to finding ways of predicting the effects of a program absent any data on treated individuals. However, strong functional form assumptions do not necessarily need to be imposed. In many of the cases we consider, the structure of the model justifies a fully nonparametric approach to estimating program effects.

In this regard, our work is closely related to recent work by Ichimura and Taber (1998, 2002) that provides a general set of conditions necessary for reduced form, nonparametric estimation of policy effects, and proposes and implements a method for analyzing the effects of a college tuition subsidy. Our paper builds on their work by illustrating, using simple economic models, how to verify when the conditions 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. One is a modified version of a matching estimator that 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. The program was initially implemented as a randomized experiment, which creates a unique opportunity to study the performance of ex ante evaluation methods. In particular, our strategy is to 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.

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 Marshak (1953) as one of the most challenging problems facing empirical economists. 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. 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. Empirically, the model’s performance could sometimes be assessed 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

---

\(^2\)See Heckman (2000).

\(^3\)Much of the initial empirical research was aimed at predicting the demand for transportation modes.

\(^4\)For 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 the substitution patterns allowed.
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. 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 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 the 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 structural models to identify program effects for different kinds of program interventions including multiplicative wage subsidies, additive wage subsidies, income subsidies, a combination of wage and income subsidies, and school subsidy programs. For each example, we also discuss estimation strategies.

3.1 Wage and income subsidy programs

Example #1: A multiplicative wage subsidy program  In this 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.

\[
\max_{\{h\}} U(c, 1-h) \\
s.t. \\
c = hw + A
\]

In such a model, optimal hours of work \(h\) can be derived as a function of wages \(w\) and asset income \(A\):

\[h^* = \varphi(w, A)\]

If we now introduce a multiplicative subsidy to wages in the amount \(\tau\), the budget constraint becomes

\[c = h(\tau w) + A.\]
Note that the model with the subsidy can be viewed as a version if the model without the subsidy. That is, if $h^{**} = \eta(w, A, \tau)$ denotes the solution to the model with the subsidy, then we have

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

where $\tilde{w} = w\tau$. This shows that the structural model without the subsidy is also the relevant one in the presence of the subsidy, so we can study the effect of introducing a subsidy $\tau$ from ex ante wage variation in the data. As discussed in Ichimura and Taber (1998), when the reduced form relationship is the same under the old and new policy, it is sometimes possible to do a nonparametric, reduced form evaluation of the policy. In this case, we can assess the policy effect on each person’s labor supply nonparametrically as follows. First use ex ante data to estimate the $\varphi$ function that describes the relationship between hours, wages and assets. The function can be estimated nonparametrically using a method such as kernel, local linear regression or series estimation.\(^6\) For each individual, evaluate the function at the value $w$ and at the new post-policy value $\tilde{w}$ to determine the impact that the wage subsidy has on that person’s labor supply. Taking averages across people within subgroups of interest provides the average policy effect for that subgroup.

We can view the proposed estimation procedure as a matching estimator.\(^7\) 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,\(^6\)\(^7\)

\(^6\)There 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. See Ichimura and Taber (1998) for more discussion on this point.

\(^7\)Ichimura and Taber (1998) also draw an analogy between their proposed method of nonparametrically recovering policy impacts and matching.
using the fact that the economic model implies

\[ Y_{1i} = Y_{0j} \mid A_i = A_j, \tau w_i = w_j \]  

(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_{j,i}^{n} Y_{0i}(w_i = w_j \tau, A_i = A_j) - Y_{0j}(w_j, A_j)) \right] \}

where \( Y_{0j}(w_j, A_j) \) denotes the hours of work choice for an individual \( j \) with set of characteristics \( (w_j, A_j) \) and \( Y_{0i}(w_i = w_j \tau, A_i = A_j) \) the hours of work choice for a matched individual with characteristics \( (w_j \tau, A_j) \). The matches can only be performed in the region \( S_p \) where the support of \( \tilde{w} = w_j \tau \) lies within the support of \( w_j \).\footnote{\( S_p = \{ \tilde{w} \text{ such that } f_w(\tilde{w}) > 0 \} \), where \( f_w(\tilde{w}) \) is the density of \( w \) evaluated at \( \tilde{w} \).} An interesting 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 under weak assumptions, notably, without having to specify the functional form of the utility function. The main assumption is that the subsidy only operates through the budget constraint and does not directly affect utility. 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 \( \tau \) \( (U(c, 1 - h) + v(\tau)) \), then it is possible to estimate the effect of the policy nonparametrically, even if \( \tau \) directly affects utility. This would allow, for example, for a "feel good" effect from receiving the subsidy.
Finally, although we have discussed the example in terms of a wage subsidy, the analysis would also hold if \( \tau \) were a tax instead of a subsidy. In the case of a tax, the function \( v(\tau) \) might represent a psychic benefit or a psychic cost that people get from paying taxes.\(^9\) Also, while we have focused on hours work as the outcome of interest, the same analysis would apply if the outcome of interest were the decision to work, which is just a transformation of hours of work (i.e. \( 1(h^* > 0) \)).

**Allowing for unobserved heterogeneity**  In the above model, there is no unobserved heterogeneity, so that all individuals with the same asset and wage values make the same decision. To make the model more realistic, we can incorporate an unobserved heterogeneity term, \( \mu \), that is assumed to affect preferences for leisure or consumption:

\[
\max_{\{h\}} U(c, 1 - h, \mu)
\]

\[
s.t.
\]

\[
c = hw + A
\]

Now, the optimal choices for hours worked will also be a function of the unobserved heterogeneity term, \( \mu \) : \( h^* = \varphi(w, A, \mu) \). In the presence of the subsidy, the optimal choice is given by \( h^{**} = \varphi(\tilde{w}, A, \mu) \).

Since \( \mu \) is unobserved, it is not possible to match individuals based on their values of \( \mu \). To justify the application of the matching in the presence of unobserved heterogeneity, we require that:

\[
E(h^{**}|A = A_i, w = w_i) = E(h^*|A = A_i, w = \tau w_i) \tag{2}
\]

or

\[
E(Y_1|A = A_i, w = w_i) = E(Y_0|A = A_i, w = \tau w_i)
\]

\(^9\)It 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.
This assumption is equivalent to the condition that
\[
\int \varphi(w, A, \mu) f(\mu|w, A) d\mu = \int \varphi(\tau w, A, \mu) f(\mu|\tau w, A) d\mu \text{ for } w \in S_P.
\]

If we assume in addition that the policy function \( \varphi \) is additive in the unobservables, \( \varphi(w, A, \mu) = \varphi_1(w, A) + \varphi_2(\mu) \), then a sufficient condition to apply the matching method is
\[
f(\mu|w, A) = f(\mu|\tau w, A) \text{ for } w \in S_P,
\]
which is a strong assumption. Also, the additivity assumption places restrictions on the class of utility functions that can be considered.

Under assumption (1) and (2) (or (1), additivity and (3)), the average policy effect can be estimated by
\[
\frac{1}{n} \sum_{j=1}^{n} E(Y_{0i}|w_i = w_j, A_i = A_j) - Y_{0j}(w_j, A_j),
\]
where \( Y_0 \) denotes \( h^* \) as before. \( E(Y_{0i}|w_i = w_j, A_i = A_j) \) can be estimated nonparametrically by nearest neighbor, kernel or local linear matching.\(^{10}\) (See Rosenbaum and Rubin (1983) for discussion of nearest neighbor methods and Heckman, Ichimura and Todd, 1997, for discussion of other nonparametric methods).\(^{10}\)

**Example #2: An additive wage subsidy program**  
We next consider ex ante evaluation under alternative subsidy schemes. For simplicity, we ignore unobserved heterogeneity, since the treatment of it would be the same as in the previous example. 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
\[
c = h(w + \tau) + A.
\]
\(^{10}\)Here, the matching has to be performed on two variables.
Thus, we have
\[ h^{**} = \eta(w, A, \tau) = \varphi(\tilde{w}, A), \]
where \( \tilde{w} = w + \tau \). This justifies using an estimation strategy identical to that in the previous example, except that now we match untreated individuals with wages \( \tilde{w} = w + \tau \) and assets \( A \) to untreated individuals with wages \( w \) and assets \( A \).

**Example #3: An income transfer program**  Next consider a program that does not alter wages, but supplements income by an amount \( \tau \). 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,
\[ h^{**} = \eta(w, A, \tau) = \varphi(w, \tilde{A}). \]
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 \( \tilde{A} \).

**Example #4: A combination wage subsidy and income transfer**  Suppose a program provides an earnings supplement in the amount \( \tau_1 \) and an additive wage subsidy in the amount \( \tau_2 \). The budget constraint takes the form
\[ c = h(w + \tau_1) + A + \tau_2, \]
which can be written as
\[ c = h\tilde{w} + \tilde{A}, \]
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 \( (\tilde{w}, \tilde{A}) \) are matched to other untreated individuals with values of wages and assets equal to \( (w, A) \).
Interestingly, matching is 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 investment in human capital. Programs that condition cash transfers on school attendance currently exist in Brazil, Colombia, Costa Rica, Mexico, and Nicaragua.\textsuperscript{11}

We next consider how to evaluate the effects of a school subsidy programs, under the assumption that there is no direct variation in the data in the price of schooling. This example and the next one is based on a model presented in Todd and Wolpin (2004). The application in that paper was to evaluating the effect of the PROGRESA program that was introduced in Mexico in 1997 as a means of increasing school enrollment and reducing child labor. In this example, child wages play a crucial role in identifying school subsidy effects. In the first variant of the model (example #5), we assume that child wage offers are observed. Later, in example, #6, we assume child wages are not observed.

Example #5: 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 an indicator for whether the child attends school \((s)\). A child that does not attend school is assume to work in the labor market at wage \(w\) (below we consider an extension to allow for leisure as another option). Letting \(y\) denote household income, net of the child’s earnings, the problem solved

\textsuperscript{11}Bangladesh has adopted a similar kind of program that conditions food transfers on school attendance. (cite Ravaillon’s paper)
by the household is

$$\max_{\{s\}} U(c, s)$$

s.t.

$$c = y + w(1 - s).$$

In this example, the optimal choice $$s^* = \varphi(y, w)$$. Now consider the effects of a policy that provides a subsidy in the amount $$\tau$$ for school attendance, so that the problem becomes:

$$\max_{\{s\}} U(c, s)$$

s.t.

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

We can rewrite the constraint of the model as

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

which shows that the optimal choice of $$s$$ in the presence of the subsidy is $$s^{**} = \varphi(\tilde{y}, \tilde{w})$$, where $$\tilde{y} = y + \tau$$ and $$\tilde{w} = w - \tau$$. That is, the schooling choice for a family with income $$y$$ and child wage $$w$$ 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** We can estimate the effect of the subsidy program on the proportion of children attending school by matching children from families with income $$\tilde{y}$$ and child wage offers $$\tilde{w}$$ to children from families with income $$y$$ and child wages $$w$$. A matching estimator of average program effects for those offered the program (the so-called "intent-to-treat" or ITT estimator) takes the form

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

where $$s_j(w_j, A_j)$$ denotes the school attendance decision for a child of family $$j$$ with characteristics $$(w_j, y_j)$$. As before, 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. Thus, nonparametric reduced form policy variation is feasible in this case, even when there is no variation in the data in the policy instrument (the price of schooling).

In this example, not all families choose to participate in the subsidy program. Since the costs of the program will depend on how many families participate in it, a key question of interest is the coverage rates of the 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:

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

We can estimate this probability 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, since nonparametric estimation does not provide a way of extrapolating outside the support region. Taking averages across the probability estimates for all families then 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:

$$ITT(w, y) = \Pr(\text{participates in program}| w, y)TT(w, y) + \Pr(\text{does not participate}| w, y)0.$$ \(12\)

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):

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

**Some Extensions**  The above model assumed that parental utility depends directly on child schooling. The model could easily be modified to allow parental utility to be a function of children’s future wages ($w^f$), which in turn depends on schooling levels ($U(c, w^f(s))$).

The above model also assumed that parents were making decisions about one child. If we were willing to assume that fertility is exogenous with respect to the subsidy, then the model could easily be modified to allow for multiple children. For example, suppose there are two children in the family who are eligible for subsidies $\tau_1$ and $\tau_2$, 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 subsidies). Then, the problem becomes:

$$\max_{(s^1, s^2)} U(c, s^1, s^2)$$

s.t.

$$c = (y + \tau^1 + \tau^2) + (w^1 - \tau^1)(1 - s^1) + (w^2 - \tau^2)(1 - s^2).$$

Estimation of the subsidy effect on enrollment requires that we match 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 $\tilde{w}^2 = w^2 - \tau^1$.

**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:

$$\max_{(s,l)} U(c, l, s)$$

s.t.

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

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

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

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

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 $\tau$. Because of the dependence on $\tau$, 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.

**Example # 6: School attendance subsidy when child wages are not observed**

Consider a household making a single period decision about whether to send a single child to school or to work. Let the utility of the household be separable in consumption ($c$) and school attendance ($s$), namely $u = c + (\alpha + \varepsilon)s$, where $s = 1$ if the child attends school, $= 0$ otherwise and $\varepsilon$ is represents heterogeneity across families in preferences for schooling. Assume that the cost of attending school depends on distance to the school, denoted $k$. Children who work contribute to family income, so the family’s income is $y + w(1 - s) - \delta ks$, where $y$ is parent’s income, $w$ is the child’s earnings, and $\delta ks$ is the distance cost that is only incurred if the child attends school. Under utility maximization, the family chooses to have the child attend school ($s = 1$) if $\varepsilon > w - \alpha + \delta k$.  

15
If we assume that \( \varepsilon \) has conditional median equal to zero \( (F_{\varepsilon|w,k}(0) = 1/2) \) and that the wage offers and distances are observed for all children, then \( \alpha \) and \( \delta \) are nonparametrically identified. (See Manski, 1988)\(^{13} \) Suppose instead that wages are only observed for children who work and specify a child wage offer equation:

\[
w = z\gamma + v
\]

where \( z \) are other determinants of wage offers (such as regional labor market characteristics) and are observed for all children. The equation governing whether a child attends school or works is

\[
s = 1 \text{ if } \alpha - \delta k + \varepsilon > z\gamma + v, \text{ else } s = 0.
\]

The probability that a child attends school can be written as

\[
\Pr(s = 1|z, k) = \Pr(z\gamma - \alpha + \delta k < \varepsilon - \nu) = F_{\varepsilon - \nu|z,k}(z\gamma - \alpha + \delta k),
\]

where \( F_{\varepsilon - \nu|z,k}(\cdot) \) is the cdf of \( \varepsilon - \nu \). Under an assumption that the conditional median of \( \varepsilon - \nu \) is 0, the parameters \( \gamma, \alpha \) and \( \delta \) can be estimated up to scale by either a parametric or semiparametric discrete choice estimation method.\(^{14} \)

Next, consider estimation of the child wage offer equation using only data on children who work \((s = 0)\) for whom wages are observed. We can write the wage equation as

\[
w = z\gamma + E(\nu|z, s = 0) + \{\nu - E(\nu|z, s = 0)\},
\]

where the error term in brackets \( (\eta = \nu - E(\nu|z, s = 0)) \) has conditional mean zero by construction.

As shown by Heckman (1980), we can consistently estimate the parameter \( \gamma \) by including a control function to capture \( E(\nu|z, s = 0) \). Under the assumptions that (i) \( \nu \) and \( \varepsilon \) are jointly

\(^{13}\)They could be estimated by a nonparametric binary choice estimator, such as a maximum score estimator. (Manski, 1975)

\(^{14}\)See Manski (1988).
distributed with density \( f(\nu, \varepsilon | z, k) \) and (ii) the conditional density equals the unconditional density, \( f(\nu, \varepsilon | z, k) = f(\nu, \varepsilon) \), we obtain
\[
\begin{align*}
  w &= z\gamma + K(\Pr(s = 1 | z, k)) + \eta
\end{align*}
\]

If there is a continuous exclusion restriction that affects the school-going decision but not the wage offer equation (in this example, the continuous exclusion restriction is assumed to be distance to school, \( k \)), then the parameter \( \gamma \) can be identified even when the form of the \( K \) function is unrestricted.\(^{15}\) To see why, note that we can hold constant \( z \) at some value and vary \( k \) and, in doing so, nonparametrically trace out the \( K \) function. Fixing \( K \) at some value, variation in \( z \) provides the source of identification of \( \gamma \).\(^{16}\) After \( \gamma \) is obtained, we can use the results of the initial discrete choice estimation, to obtain the scale of \( \alpha \) and \( \delta \).

Next, consider how we would use this model to estimate the effects of a government program that aims to increase school attendance of children through the introduction of a subsidy to parents in the amount \( b \) if they send their child to school. Under such a program, the probability that a child attends school will increase by \( F_{\varepsilon-\nu}(z\hat{\gamma} - \hat{\alpha} - b + \hat{\delta}k) - F_{\varepsilon-\nu}(z\gamma - \alpha + \delta k) \). The function \( F_{\varepsilon-\nu}(s) \) can be estimated nonparametrically by a nonparametric regression of \( s \) on \( z\hat{\gamma} - \hat{\alpha} + \hat{\delta}k \) (here, we use the fact that \( E(s|z\hat{\gamma} - \hat{\alpha} + \hat{\delta}k = \tau) = \Pr(s = 1|z\hat{\gamma} - \hat{\alpha} + \hat{\delta}k = \tau) \)). To assess the effect on the probability of attending for each person, given the incentivewith the subsidy, we simply evaluate the function at the point \( \tau - b \).

\section{Empirical application: predicting effects of a school subsidy program}

In this section, we apply the methods described previously to analyze the effects of the cash transfer program PROGRESA that was introduced in Mexico in 1997. The program provides transfers to families that are contingent upon their children regularly attending

---

\(^{15}\)Only weak assumptions on the continuity of the \( K \) function are required. See Heckman and Robb (1980).

\(^{16}\)The intercept of the wage equation will, in general, not be separately identified from the \( K \) function unless there is a subset of the data for which \( \Pr(s = 0 | z, k) = 1 \). On this point, known in the literature as identification at infinity, see Heckman (1980) and Andrews and Schafgans (1998).
school. These transfers are intended to alter the private incentives to invest in education by offsetting the opportunity cost of not sending children to school. 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 offered the program participate in it to some extent. Partial participation is possible, for example, if the family can send some children to school but not others.

Table 1: Monthly Transfers for School Attendance

<table>
<thead>
<tr>
<th>School level</th>
<th>Grade</th>
<th>Monthly Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>Primary</td>
<td>3</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>135</td>
</tr>
<tr>
<td>Secondary</td>
<td>7</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>255</td>
</tr>
</tbody>
</table>

The PROGRESA program was initially introduced in rural areas, has since expanded into semi-urban and urban areas, and currently has a coverage of about ten million families. For purposes of evaluation, the initial phase of PROGRESA was implemented as a

---

17 The 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.

18 In the rural villages that participated in the initial PROGRESA experiment, all the households were interviewed and informed of their program eligibility status.
social experiment, in which 506 rural villages were randomly assigned to either participate in the program or serve as controls.\footnote{Data 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.} Randomization, under ideal conditions, allows mean 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 effectiveness of 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 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.\footnote{The control group was also incorporated two years later, but they were not told of the plans for their future incorporation during the time of the experiment.}

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 from program eligible families, who are 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. The sample excludes children from families where the husband or the wife reports being self-employed, which was necessary because the data are not detailed enough to determine their income. In addition to the household survey datasets, supplemental data were gathered at the village level. Most importantly, for our purposes, 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.

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. 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, according to the procedure described in section 3.2, example #5. The estimator we use is given by

$$\hat{\alpha} = \frac{1}{n} \sum_{j=1}^{n} \{ E(s_j| w_i = w_j - \tau_j, y_i = y_j + \tau_j) - s_j(w_j, y_j) \},$$

where $s_j$ is an indicator for whether child $j$ is enrolled in school, $w_j$ is the wage offer for child $j$, and $y_j$ is family income (net of any child income). This estimator matches 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 $y_i = y_j + \tau_j$. Here, $\tau_j$ represents the subsidy level for which the child is eligible. Since subsidies vary by grade level, children of the same age can be

21 Approximately 10 pesos equals 1 US dollar.
eligible for different subsidy levels if they attend school.\textsuperscript{22}

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

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

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:

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

which satisfies the standard assumptions $\int K(s)ds = 1$, $\int sK(s)ds = 0$, and $\int K(s)s^2ds < \infty$. Asymptotic consistency of this estimator requires that the smoothing parameters satisfy $nh_n^w h_n^y \to \infty$, $h_n^w \to 0$ and $h_n^y \to 0$ as $n \to \infty$.\textsuperscript{23}

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 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, $q$, that is small and positive. That is, we check whether

$$\hat{f}(\hat{w}_0, \hat{y}_0) > q,$$

where $\hat{f}(\cdot, \cdot)$ is a nonparametric estimate of the density.\textsuperscript{24}

\textsuperscript{22}In Mexico, it is fairly common for children of a given grade level to vary a lot by age.
\textsuperscript{23}See, e.g., Härdle and Linton (1994).
\textsuperscript{24}This procedure is similar to that used in Heckman, Ichimura and Todd (1997).
Tables 2a and 2b compare the predicted program impacts obtained by the method described above to the experimental impacts for boys and girls. The table gives the impacts on enrollment in percentage points. Impacts are estimated separately over various age ranges. The sample size in each cell is shown in parentheses along with the percentage of observations that lie outside of $S_P$. For all the age/gender groups, the experimentally estimated program impacts are positive. The predicted impacts are also all positive, even though the estimation procedure does not constrain them to be positive. For boys, the predicted impact understates the actual impact for boys age 12-13 (0 vs. 4.9 percentage points), but then overstates it for boys age 14-15 (5.7 vs. 1.6). The predicted impact over the entire range, age 12-15, is fairly close to the experimentally estimated impact (2.8 vs. 2.1). For girls, the predicted program impacts tend to underestimate the actual program impacts - by 0.9 percentage points for ages 12-13, 4.9 for ages 14-15 and and 2.4 for the overall age range 12-15.

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages</td>
<td>Experimental</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>12-13</td>
<td>4.9</td>
<td>0.0</td>
<td>(232,11%)</td>
</tr>
<tr>
<td>14-15</td>
<td>1.6</td>
<td>5.7</td>
<td>(197,11%)</td>
</tr>
<tr>
<td>12-15</td>
<td>2.8</td>
<td>2.1</td>
<td>(429,11%)</td>
</tr>
</tbody>
</table>

25 We 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 for wages and equal to 400 for income. The cut-off used for determining $S_P$ was set equal to $1e-08$. 
Table 2b
Comparison of Ex-Ante Predictions and Experimental Impacts

<table>
<thead>
<tr>
<th>Ages</th>
<th>Experimental</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12-13</td>
<td>7.7</td>
<td>6.8</td>
<td>(221,10%)</td>
</tr>
<tr>
<td>14-15</td>
<td>14.8</td>
<td>9.9</td>
<td>(179,11%)</td>
</tr>
<tr>
<td>12-15</td>
<td>11.3</td>
<td>8.9</td>
<td>(400,10%)</td>
</tr>
</tbody>
</table>

In addition to predicting the effect of the existing subsidy program, we can also use the same estimator to study the effects of other hypothetical programs, such as changes in the subsidy schedule. Tables 3a and 3b consider an increase of the subsidy to 1.5 times the original subsidy schedule, as well as a decrease to one half of the original subsidy amounts. As seen in parentheses, the fraction of observations that lie outside of $S_P$ increases at higher levels of the subsidy, and decreases at smaller subsidy amounts. This shows clearly how the range of subsidies levels that can be considered is limited by the range of the data.\(^{26}\)

With one exception for boys, the predicted impacts suggest that enrollment levels would either stay the same or increase as the level of subsidy increases. For half the original subsidy, the impacts for girls and boys are roughly comparable. When we increase the subsidy, the predictions indicate sizeable effects for both boys and girls, but the effect size is much larger for girls.

Table 3a
Effects of Counterfactual Subsidy Levels

<table>
<thead>
<tr>
<th>Ages</th>
<th>1.5*Original</th>
<th>Original</th>
<th>0.5*Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-13</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(232,63%)</td>
<td>(232,11%)</td>
<td>(232,0)</td>
</tr>
<tr>
<td>14-15</td>
<td>9.7</td>
<td>5.7</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>(197,57%)</td>
<td>(197,11%)</td>
<td>(197,0)</td>
</tr>
<tr>
<td>12-15</td>
<td>3.1</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>(262,64%)</td>
<td>(429,11%)</td>
<td>(429,0)</td>
</tr>
</tbody>
</table>

\(^{26}\) Also, see Ichinura 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 suggested a fully non-parametric estimation strategy. Our work builds on Ichimura and Taber (1998, 2002) by illustrating when the conditions for nonparametric policy evaluation are met for a variety of program interventions, including wage subsidies and income support programs.

This paper also suggested some simple estimators, which are modified versions of matching estimators. The estimators compare untreated individuals to other untreated individuals, where the set of variables on which the matching is based is implied by the behavioral model. We study the performance of the estimators using data from the Mexican PROGRESA experiment. 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 usually come within 30% of the experimental impact.

### Table 3b

Effects of Counterfactual Subsidy Levels

<table>
<thead>
<tr>
<th>Ages</th>
<th>1.5*Original</th>
<th>Original</th>
<th>0.5*Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-13</td>
<td>9.4</td>
<td>6.8</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>(221,37%)</td>
<td>(221,10%)</td>
<td>(221,0)</td>
</tr>
<tr>
<td>14-15</td>
<td>19.2</td>
<td>9.9</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>(179,39%)</td>
<td>(179,11%)</td>
<td>(179,0)</td>
</tr>
<tr>
<td>12-15</td>
<td>13.5</td>
<td>8.9</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>(400,38%)</td>
<td>(400,10%)</td>
<td>(400,0)</td>
</tr>
</tbody>
</table>
References


27
Histogram of Min Monthly Laborer Wage

Histogram of Total Family Income

Figure 1