Within-Family Program Effect Estimators: The Impact of *Oportunidades* on Schooling in Mexico[1]

Susan W. Parker, Petra E. Todd, Kenneth I. Wolpin

August 7, 2006

[1]This research was supported by a grant from the Mellon Foundation/Population Studies Center (PSC)/University of Pennsylvania to Parker, Todd and Wolpin. Parker is a Profesora/Investigadora, División de Economía, CIDE, Mexico City. Todd is an Associate Professor of Economics and a PSC Research Associate at the University of Pennsylvania. Wolpin is a Professor of Economics and a PSC Research Assocaite at the University of Pennsylvania. We are grateful to *Oportunidades* for permission to use these data and to Iliana Yachine for her help in interpreting the data.

Abstract

This paper explores the use of dynamic panel data models and sibling-based estimation procedures for evaluating the effects of educational interventions aimed at children or youth. It specifies statistical models of the schooling decision that are consistent with dynamic behavioral models and that allow for unobservable heterogeneity. The frameworks presented provide a way of directly assessing short-term program impacts and of simulating longer term impacts that may extend beyond the program exposure times observed in the data. This paper also proposes a new estimation approach for the model, which combines matching, differencing and instrumental variables in a way that minimizes the need for parametric modeling assumptions. The approach uses retrospective data on older siblings' schooling histories to identify how an educational intervention affects younger siblings. We apply the proposed method as well as other methods to study the effects of the Mexican *Oportunidades* conditional cash transfer program on schooling outcomes of children and youth living in urban areas. The analysis samples come from the 2002-2004 ENCELURB evaluation datasets. The empirical results show significant short-term impacts of the program on school enrollment. Simulations based on the dynamic model indicate that long-term exposure to the program increases educational attainment by about half a year.

1 Introduction

This paper has three goals. The first is to explore the use of a dynamic panel data schooling model as a means for evaluating the short- and long-term effects of educational interventions. We contrast the dynamic modeling approach with standard potential outcomes frameworks commonly adopted in the evaluation literature. The second goal is to study the application of sibling-difference estimators for evaluation of social programs. Although widely used in other contexts, kinship-based estimators are rarely used for program evaluation.[1] The method we use identifies program effects by exploiting within-family variation in the timing of the program with respect to children's ages. In particular, retrospective data on older siblings' schooling histories are used to assess how an educational intervention affects the school attendance and schooling attainment of younger siblings. We consider both standard sibling-difference estimators as well as a new semiparametric approach. The third goal of the paper is to implement the estimators for the purpose of evaluating the effects of the *Oportunidades* conditional cash transfer program on the schooling outcomes of children and youth in urban Mexico.

The statistical models of the schooling decision problem that we specify are motivated by and are consistent with dynamic behavioral models, although we do not structurally estimate a behavioral model. We follow a previous literature (e.g. Rosenzweig (1986), Rosenzweig and Wolpin (1986, 1988, 1995)) in illustrating how different assumptions about the underlying behavioral model justify different types of estimators. For example, a dynamic model of parents making decisions about their children's schooling that allows for intrafamily investment behavior and unobserved heterogeneity leads to a dynamic panel data formulation of the schooling model with endogenous regressors. We consider alternative modeling frameworks as well as estimation strategies that are closely tied to the information structure of the dynamic behavioral models.

[1] Exceptions are Rosenzweig and Wolpin (1986), which used child fixed-effect estimators to investigate the effects of a family planning intervention on child health, and Currie and Thomas (1995), which used siblings to study the impact of Head Start.

One attractive feature of the methods studied in this paper is that some of them can be implemented even in the presence of severe data limitations, such as missing baseline data or only having data on treated households. Another advantage is that the identifying assumptions that justify their application may be more plausible than the kinds of assumptions typically invoked by other evaluation estimators. For example, matching-on-observables estimators usually compare outcomes of individuals who receive an intervention to outcomes of observably similar individuals from other families who do not receive the intervention. Our estimators restrict comparisons to be among children within the same family and of the same gender and age, arguably controlling for many unobservable determinants of outcomes that are not controlled by standard procedures. A potential concern with the sibling comparison approach is that siblings of the same age will necessarily be observed in different time periods, assuming they are not twins, but we discuss ways of nonparametrically accounting for time effects.

The estimators studied in this paper use a combination of exact matching, differencing, and instrumental variables in estimating program effects. They can be viewed as an application of the Ahn and Powell (1993) pairwise differencing estimator, developed for the sample selection model, to a panel data, treatment effects setting. Specifically, we consider panel data models of the form:

$$y_{ijt} = x_{ijt}\beta + D_{ijt}\alpha + \varphi(f_i, z_{ijt}) + \varepsilon_{ijt},$$

where i indexes the child, j the family and t the time period. In the above equation, f_i denotes an unobserved family fixed effect, D_{ijt} is an indicator for program participation and x_{ijt} and z_{ijt} represent state variables (implied by a dynamic model) that may be time varying. The coefficient α represents the one-period treatment effect, which is one of the main parameters of interest. The model can also be used to simulate the effects of multiple years of exposure to the treatment, as described in section two of the paper.

Our estimator for β and α treats $\varphi(f_i, z_{ijt})$ as a nuisance function to be eliminated by differencing among siblings with the same value of z_{ijt} (e.g. age, gender). Under the assumption that the outcome equation is additively separable in $\varphi(f_i, z_{ijt})$, the procedure is nonparametric with regard to the influence of f_i and z_{ijt} on y_{ijt}. We consider two behavioral models, one in which the x_{ijt} and z_{ijt} are exogenous (conditional on f_i) and the other in which a subset of the x_{ijt} are endogenous.

We use the sibling-difference estimators to study the effects of the Mexican *Oportunidades* conditional cash transfer program on schooling outcomes of children and youth living in urban areas. The program provides cash grants to poor families if they send their age-eligible children to school and regularly visit health clinics for check-ups. In recent years, conditional cash transfer programs have become the anti-poverty program of choice for many governments in Latin America and South America. Such programs now exist in Argentina, Brazil, Chile, Colombia, Costa Rica, Equador, Honduras, Nicaragua, and Uruguay, and Peru.[2] The success of the Mexican *Oportunidades* program (formerly called PROGRESA), as measured by the experimental evaluation carried out in the first two years of its implementation (1998-1999), is often cited as justification for developing new school subsidy programs designed to promote investment in human capital. (Becker, 1999, Krueger, 2002) Impact evaluations based on the rural evaluation datasets demonstrated significant impacts of the program on reducing child labor, improving health outcomes, and increasing schooling enrollment and attainment.[3] Partly in response to these encouraging impact results, the Mexican government expanded the program at a rapid rate into semiurban and urban areas. Today, the program covers five million families–about one quarter of all families in Mexico.

The implementation of the *Oportunidades* program in urban areas deviates in several important ways from the earlier-established rural program. One is the procedure by which families become beneficiaries. In the rural program, a census of the targeted localities was conducted and all families that met the eligibility criteria for the program were informed of their eligibility status.

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

[3] For evaluations of educational impacts in rural areas, see Schultz (2000,2004), Behrman, Sengupta and Todd (2005), Parker and Skoufias (2000), Buddelmeyer and Skoufias (2003), and Todd and Wolpin (2004).

Almost all eligible families decided to participate in the Program. For cost reasons, this type of census was not deemed feasible in urban areas, and an alternative system of advertised sign-up modules was adopted. Households that deemed themselves potentially eligible had to apply for the program during a time window of opportunity when the module was open. Because of this new mechanism for incorporation, about 40% of households who were eligible did not solicit incorporation.[4] Another difference between the current *Oportunidades* program and the initial rural program is that subsidies are now provided for grades 10 to 12 (high school), whereas initially the program gave subsidies only until grade 9 (secondary level). We therefore would expect to see impacts of the current *Oportunidades* program on older children as well as potentially larger impacts on younger children than in the original rural program. Program impacts in urban areas might also be expected to differ from those in rural areas, because access to schooling and health facilities differs and work opportunities differ.

In this paper, we study the short-term and longer-run effects of the urban *Oportunidades* program, focusing on the so-called intent-to-treat parameter, where treatment is defined as being offered the program. We analyze observational data on girls and boys age 6-17 that were collected as part of the urban evaluation. We make use of data on three groups: (a) households who are eligible for the program and who live in intervention areas, (b) households who just missed the cut-off for eligibility for the program and who live in intervention areas, and (c) households who live in areas where the program was not available but who otherwise satisfy program eligibility criteria. The data were gathered in three rounds. One round is baseline data, gathered in the fall of 2002, just prior to the start of the program.[5] The other two rounds are post-program data collected in the fall of 2003 and 2004 after participating households had experienced one and two

[4] In about 30% of cases, the households report not being aware of the program.

[5] At the time of the application of the 2002 survey, many households had already applied and been notified of their beneficiary status, although they had yet to receive any payments. It is likely, therefore, that fall 2002 school enrollments were affected by the expectation of receiving the program. For this reason, we consider the 2001-2002 school year as being pre-program and the 2002-2003 school year as post-program.

years in the program. Using the dynamic panel data within family evaluation estimators, we find that the program has significant effects on school attendance for older girls and boys, in the range of 9.0-14.4 percentage points, depending on gender and on the source of comparison group data used. When we use the estimated model to simulate long-term program effects for a typical family, we find that a child participating in the program from age 6-17 would complete about 0.5 additional years of schooling.

The paper develops as follows. Section two describes the theoretical schooling models that motivate our estimation approaches and defines parameters of interest. Section three presents the estimators. Section four describes the data used to implement them and gives summary statistics. Section five presents the main empirical findings concerning short and long run program effects. Section six concludes.

2 The Sequential Schooling Model and Parameters of Interest

The key parameter of interest in this paper is the so-called *intent-to-treat* parameter (ITT), where treatment is defined as receiving the offer of the program. This parameter combines the effect of treatment for those who took up the program as well as for those who did not take up the program.[6] In arriving at our statistical model, we first consider the decision problem in light of a dynamic behavioral model of parents making decisions about their children's schooling.

2.1 Two Alternative Models of Schooling Decisions

In specifying our model for schooling outcomes, we are guided by earlier work in Todd and Wolpin (2002) and Attanasio and Meghir (2002), which develops dynamic structural models of decision making about child time use in the context of rural Mexico.[7]

[6] If a fraction of the population offered treatment does not take up the program, then the intent-to-treat parameter reflects the lack of impact on that segment of the population.

[7] Also relevant is the earlier economic literature on sequential modeling of the schooling decision. See, for example, Cameron and Heckman (1998), Keane and Wolpin (1997) and the references cited therein.

The Basic Model The first schooling model we consider is one in which parents are assumed to make decisions about each child's school attendance taking into account only that child's characteristics, but not taking into account other children in the family.[8] Schooling decisions are also assumed to depend on family-specific unobservables (time invariant) that may reflect permanent heterogeneity in preferences for schooling, differences in average ability levels of children across families, or differences in expectations about the future wage returns to schooling. Within such a model, school decisions depend on the state variables at the time of making the decision, which we specify as the child's age and highest grade completed, whether the child attended school last year, the education levels of the parents, and the unobserved family component. We refer to this model as the "basic" model.

The Intrafamily Model In addition to the basic model specification, we also consider a more general model in which parents consider all their children's schooling outcomes when making decisions about any particular child. For example, if an older child dropped out of school at a young age, say because of a negative health shock, then parents may place more emphasis on a younger child finishing school. In a model with intrafamily investment behavior, the set of state variables needs to be augmented to include variables that describe the characteristics of other children in the family, such as their ages and schooling attainment.

In the empirical work, we implement both the "basic model" and the "intrafamily model."

2.2 Using a Sequential Model to Identify One-Year Program Effects and to Simulate Multi-Year Effects

To fix ideas, we start with the basic model where school attendance decisions depend only on the single child's characteristics and not on those of siblings. Let Y_{ijt} denote the school attendance decision for child/youth i from family j at time t. Let x_{ijt} denote a vector of state variables at

[8] The basic model described here could be viewed as an approximation to the decision rule in the model that is structurally estimated in Attanasio and Meghir (2002).

time t determining choices. For now, assume that x_{ijt} may contain time varying variables, but that how they evolve over time is unaffected by treatment (e.g. age of the child). The lag variable Y_{ijt-1} denotes whether a child attended last year. Let $D_{ijt} \in \{0,1\}$ indicate whether the individual is treated, where we assume that $D_{ijt} \geq D_{ijt'}$ for $t > t'$, so that once a person is in the treated state they remain in that state. For simplicity, we also assume here that treatment effects α are constant.

The sequential schooling model can be written as

$$Y_{ijt} = x_{ijt-1}\beta + Y_{ijt-1}\psi + D_{ijt}\alpha + f_i + \varepsilon_{ijt}. \tag{*}$$

where α is the one-year program impact, f_i is an unobserved family fixed effect that may be correlated with D_{ijt} and ε_{ijt} is a random shock, which is assumed to be serially uncorrelated for a given child, uncorrelated across children and uncorrelated with the x_{ijt}, Y_{ijt-1}, and D_{ijt}.[9]

To relate this framework to a more conventional potential outcomes framework, substitute repeatedly for the lagged outcome variables Y_{ijt-1} to get

$$Y_{ijt} = \sum_{k=0}^{t} x_{ijt-1-k}\beta\psi^k + \alpha\sum_{k=0}^{t} D_{ijt-k}\psi^k + f_i\sum_{k=0}^{t}\psi^k + \psi^{t+1}Y_0 + \sum_{k=0}^{t}\varepsilon_{ijk}\psi^k.$$

This equation writes the outcome in period t as a function of past x values, the number of years that the individual received treatment, a fixed effect, the initial condition (Y_0), and all the past shocks. In this equation, the term $\alpha\sum_{k=0}^{t} D_{ijt-k}\psi^k$ gives the effect of cumulative exposure to treatment on current period outcomes.

Let Y_{ijt}^1 denote the potential outcome if an individual participates in the program for time periods beginning with period t_0 up until t. Let Y_{ijt}^0 denote the potential outcome in the untreated

[9] Recall that D_{ijt} denotes being offered the program, because our focus is on estimating an intent-to-treat (ITT) parameter. For this reason, it is reasonable to assume that there is no correlation between which sibling is offered the program (D_{ijt}) and ε shocks.

state. In terms of the parameters of the sequential model, the potential outcomes are:

$$
\begin{aligned}
Y_{ijt}^1 &= \sum_{k=0}^{t} x_{ijt-1-k}\beta\psi^k + \sum_{k=t_0}^{t} D_{ijt-k}\alpha\psi^k + f_i\sum_{k=0}^{t}\psi^k + \psi^{t+1}Y_0 + \sum_{k=0}^{t}\varepsilon_{ijk}\psi^k \\
Y_{ijt}^0 &= \sum_{k=0}^{t} x_{ijt-1-k}\beta\psi^k + f_i\sum_{k=0}^{t}\psi^k + \psi^{t+1}Y_0 + \sum_{k=0}^{t}\varepsilon_{ijk}\psi^k.
\end{aligned}
$$

Now suppose that the x_{ijt} variables are also potentially affected by the treatment. In that case, to derive the cumulative program effect, we need an auxiliary model for how the program affects x_{ijt}. For example, if x_{ijt} were highest grade completed, it would evolve as a function of previous highest grade completed and the current attendance decision:

$$x_{ijt-1} = x_{ijt-2} + Y_{ijt-1}.$$

Substituting repeatedly for x_{ijt} gives:

$$x_{ijt-1} = \sum_{k=0}^{t-1} Y_{ijk}.$$

Substituting this equation into (*) gives:

$$Y_{ijt} = (\sum_{k=0}^{t-1} Y_{ijk})\beta + Y_{ijt-1}\psi + D_{ijt}\alpha + f_i + \varepsilon_{ijt},$$

which can be written as

$$Y_{ijt} = \sum_{k=0}^{t-1} b_k Y_{ijk} + D_{ijt}\alpha + f_i + \varepsilon_{ijt},$$

where $b_k = \beta$ for $k = 1..t-2$ and $b_{t-1} = \beta + \psi$. Obtaining the expression of the cumulative program effect in terms of the parameters of the sequential model requires substituting in for lagged outcomes until the model can be written as a function of initial conditions (Y_0, x_0), the history of program participation, the fixed effect and all past shocks. Substituting repeatedly for

Y_{ijk}, we get:

$$Y_{ijt} = [(1+b_{t-1})(b_0 + \sum_{k=1}^{t-2} b_k b_0)]Y_0 \qquad (**)$$
$$+(1+\sum_{k=1}^{t-1} \Pi_{s=k}^{t-1} b_s) f_i + (\alpha D_{ijt} + \alpha \sum_{k=1}^{t-1} \Pi_{s=k}^{t-1} b_s D_{ijs})$$
$$+(\varepsilon_{ijt} + \sum_{k=1}^{t-1} \Pi_{s=k}^{t-1} b_s \varepsilon_{ijs})$$

The coefficients associated with the indicators for treated at time t give the cumulative treatment effect that is now inclusive of any effect coming through changing the distribution of x_{ijt}. The potential outcome without treatment is the Y_{ijt} obtaining by setting $D_{ijt} = 0$ and D_{ijt-k} to zero, for all k. The potential outcome with treatment is the one obtained by setting $D_{ijt'} = 1$ for $t' = 0$ to t, $= 0$ for $t' < t_0$.

The cumulative program effect can be obtained from estimating (**). Once the parameters are estimated, (**) can be used to determine the effects of longer terms of program exposure that may extend beyond those observed in the data. An alternative approach is to estimate the dynamic version of the model (*) using the methods described in section three and then to use it to simulate the effects of longer-term exposure. This requires simulating how Y_{ijt} evolves over time, where the values of x_{ijt-1} are updated in each period. An advantage of estimating the parameters using the dynamic model instead of the cumulative model is that the data requirements are less stringent. The cumulative model requires the entire history of lagged x's and D's, whereas the dynamic equation requires one period lagged regressors plus some additional lags to use as instruments (using the methods described in section three). We implement the dynamic model approach (*) in the empirical work.

3 Estimation

We next describe the sibling-difference estimator used to estimate the sequential schooling model. We take as a point of departure the basic model without intrafamily investment behaviors, where the regressors (x_{ijt}, z_{ijt}) are assumed to be exogenous conditional on family type. To simplify the presentation, we also assume a common effect model in which the treatment effect is assumed to be the same for everyone, denoted by α. After presenting the key ideas in the context of the simpler model, we consider estimation in the more general case with lagged sibling outcomes, and in the case with heterogeneous program effects.

3.1 The Within Family Estimator

Let Y_{1ijt} and Y_{0ijt} denote the potential outcomes for child/youth i from family j at time t in the treated and untreated states. For example, Y_{1ijt} and Y_{0ijt} could denote potential school attendance with or without the offer of the program at time t. Using the same notation as in section two, x_{ijt} denotes a vector of state variables at time t (individual, family or locality level characteristics) determining choices and D_{ijt} denotes whether the individual is treated.[10] The potential outcomes model can be written as:

$$\begin{aligned} Y_{1ijt} &= Y_{0ijt} + \alpha \\ Y_{0ijt} &= x_{ijt}\beta + \varphi(z_{ijt}, f_i) + \varepsilon_{ijt}, \end{aligned}$$

where f_i is a fixed effect and the index ijt denotes individual i in family j at time t. ε_{ijt} is a random shock that is assumed to be serially uncorrelated for a given child and uncorrelated across children.[11]

[10] It is assumed that $D_{ijt} \geq D_{ijt'}$ for $t > t'$, so that once a person is in the treated state they remain in that state.

[11] It is possible to estimate a model in which the effect of x_{ijt} is also modeled nonparametrically, for example, using the estimation methods described in Porter (1996). We do not pursue this more general form of the model here, in part because there is little practical experience with such estimators and the standard error formulae are nonstandard.

Assumption #1: $E(\varepsilon_{ijt}, \varepsilon_{kst'}) = 0$ for all i, j, s, t and t' , whenever $t \neq t'$ or $i \neq k$.

The observed outcome is

$$\begin{aligned} Y_{ijt} &= D_{ijt} Y_{1ijt} + (1 - D_{ijt}) Y_{0ijt} \\ &= x_{ijt}\beta + D_{ijt}\alpha + \varphi(z_{ijt}, f_i) + \varepsilon_{ijt}. \end{aligned}$$

The statistical problem addressed in much of the evaluation literature is that D_{ijt} is potentially correlated with the unobservables (here, f_i and ε_{ijt}.). Such correlations might results from self-selection into treatment on the basis of unobserved characteristics that are correlated with outcomes. In our case, however, D_{ijt} represents the offer of the program. The program eligibility criteria do not depend on children's schooling outcomes, so the offer of the program D_{ijt} should not be correlated with ε_{ijt}. The program eligibility criteria do, however, depend on the number of children in the family and it is conceivable that shocks to fertility might be correlated with school attendance shocks ε_{ijt}. Thus, consistency of the sibling difference estimator requires that any correlation between the unobservables of the outcome equation and D_{ijt} comes through the fixed effect f_i (See Assumption #2 below).

Let C denote the set of all *child* $\times$ *family* $\times$ *time period* observations in the data, where a single observation is denoted by the 3-tuple (i, j, t), where $j = 1..J$, $t = 1..T$ and $i = 1..N$. Let $F(j)$ denote the subset of C consisting of observations belonging to family j. Also, $S(i)$ denotes the subset of C consisting of observations on siblings of child i. That is,

$$S(i) = \{(k, j, t) \in C \text{ such that } i \in F(j), k \in F(j), i \neq k\}.$$

Let $R(\tilde{z})$ denote the subset of C for which $z_{ijt} = \tilde{z}$:

$$R(\tilde{z}) = \{(i, j, t) \in C \text{ such that } z_{ijt} = \tilde{z}\ \}.$$

One of the elements of z_{ijt} is assumed to be the age of the child. For example, $R(\tilde{z} = 9)$ denotes the subset of C pertaining to nine-year-old children.

The set $A(i, z_{ijt}) = R(z_{ijt}) \cap S(i)$ is the set of observations on siblings of child i who match i in terms of their characteristics z_{ijt}. Let $S^*(i, z_{ijt}) \in A(i, z_{ijt})$ denote the youngest older sibling of child i within set A.[12] $S^*(i, \tilde{z})$ is either a singleton or the empty set.

Using this notation, the sibling difference estimator of (β, α) solves

$$\min_{\{\beta,\alpha\}} \sum_{j=1}^{J} \sum_{\substack{\{(i,j,t)\in \\ F(j)\}}} \sum_{\substack{\{(k,j,t')\in \\ S^*(i,z_{ijt})\}}} \{y_{ijt} - y_{kjt'} - (x_{ijt} - x_{kjt'})\beta + (D_{ijt} - D_{kjt'})\alpha\}^2.$$

The estimation procedure takes pairwise differences between siblings who match in terms of z_{ijt}. The effect of the nuisance function $\varphi(z_{ijt}, f_i)$ on the outcome variable is eliminated by the differencing. The functional form of $\varphi(z_{ijt}, f_i)$ is left unspecified, so the procedure is nonparametric with regard to $\varphi(z_{ijt}, f_i)$.

Consistency requires that the error terms be orthogonal to the regressors:

Assumption #2: $E(\varepsilon_{ijt} - \varepsilon_{kjt'} | x_{ijt} - x_{kjt'}, D_{ijt} - D_{kjt'}, z_{ijt} = z_{kjt'}) = 0.$

For example, the fact that one boy in a family was offered the program and his brother was not is assumed to be arbitrary and uncorrelated with differences in the unobservables.[13] This arbitrary variation in the timing of treatment with respect to children's ages provides the source of variation that we use to identify program effects.

We impose assumption #2 instead of a stronger strict exogeneity assumption to accommodate the possibility of lagged child outcomes (such as lagged attendance and lagged highest grade completed) as regressors. As discussed in section 2.1, these variables would be in the state space at the time schooling decisions are made. Assumption #2 is also sufficient for consistency of a standard sibling fixed-effect estimator that is implemented by deviating the outcome variable and

[12] Birth order can be determined from information on age (contained in z_{ijt}) at time period t.

[13] This assumption is less problematic in the context of estimating an intent-to-treat (ITT) parameter than it would be if the focus were on a treatment on the treated (TT) parameter. Although the decision to take up the program may depend on child-specific unobservables, the offer of the program did not depend on parental decisions or on child outcomes.

the regressors from their means, where the means are taken over siblings of the same age.[14] In the empirical work, we apply both sibling-difference and fixed-effect estimators.

In our application, $\varphi(z_{ijt}, f_i)$ represents the effects of age, gender and the family fixed effect. Unlike standard sibling fixed effect estimators, the semiparametric estimator permits interactions of arbitrary form between the family fixed effect f_i and the elements of z_{ijt}. For example, it allows for permanent family heterogeneity in preferences that may differ by the age or gender of the child. The only required assumption is that:

Assumption #3: The outcome equation is additively separable in $\varphi(z_{ijt}, f_i)$.

The procedure previously described sequentially matches children with their next older (matched on z_{ijt}) sibling. For example, suppose that z_{ijt} consists of age and gender and that there are three girls within a family observed at a given age. The procedure takes differences between the youngest child and the middle child and between the middle child and the oldest child. There is no comparison between the youngest and the oldest, because that comparison would be a function of the other two comparisons.

In addition to assumptions 1-3, we also require that the treatment status varies for children within the same family with the same values of z_{ijt}–a plausible assumption given the way in which the program was introduced at a single time period when children in the family were of different ages. Holding age constant, younger siblings will have been exposed to the program and older siblings not exposed. An implicit assumption of this approach is that the introduction of the

[14] Let $a(t)$ denote the age of an individual. Write the outcome equation after deviating from means as:

$$\begin{aligned} y_{ija(t)} - y_{\cdot ja} &= \beta(x_{ija-1(t-1)} - x_{\cdot ja-1}) + \\ &\quad (D_{ija(t)} - D_{\cdot ja})\alpha + (\varepsilon_{ija(t)} - \varepsilon_{\cdot ja}), \end{aligned}$$

where $y_{\cdot ja}, D_{\cdot ja}$, $x_{\cdot ja-1}$, and $\varepsilon_{\cdot ja}$ denote means that are taken over siblings of the same age (observed at different time periods). The assumption that the error terms are serially uncorrelated and uncorrelated across siblings implies that there is no correlation between the regressors $(x_{ija-1(t-1)} - x_{\cdot ja-1})$ and the error term, even if x includes own lagged dependent variables. Thus, the sibling fixed effect estimator is unbiased for the basic model specificaiton. However, the same estimator would be biased in an intrafamily model (described in section 3.2) for two reasons. First, older siblings shocks' would potentially affect decisions about younger siblings outcomes, leading to a correlation between $\varepsilon_{\cdot ja}$ and $x_{\cdot ja-1}$. Second, lagged sibling outcomes also enter directly as regressors.

program was unanticipated, so that families did not take into account future program availability when making preprogram schooling decisions.

It is useful to compare the proposed estimator to a standard sibling-difference estimator. Define $S^{**}(i,t)$ as the youngest older sibling of child i, observed at some time $t' \neq t$. There is some ambiguity as to how to implement the standard sibling-difference estimator. Holding time constant in taking the differences would mean that $D_{ijt} = D_{kjt}$ for all sibling pairs. Therefore, differences must be taken across different time periods, which leads to the question of which time periods to use. Under the assumption that some time period $t' \neq t$. is used and under the functional form assumption $\varphi(z_{ijt}, f_i) = z_{ijt}\gamma + f_i$, the standard sibling difference estimator can be written as:

$$\min_{\{\beta,\alpha\}} \sum_{j=1}^{J} \sum_{\substack{\{(i,j,t)\in \\ F(j)\}}} \sum_{\substack{\{(i,k,t')\in \\ S^{**}(i,t)\}}} \{y_{ijt} - y_{kjt'} - (x_{ijt} - x_{kjt'})\beta + (D_{ijt} - D_{kjt'})\alpha + (z_{ijt} - z_{kjt'})\gamma\}^2.$$

The difference between our approach and a standard approach is that our method eliminates the effect of the z_{ijt} along with the fixed effects. The main advantage of doing so is that the estimator is not subject to potential misspecification of the functional form of $\varphi(z_{ijt}, f_i)$.

As previously noted, children within the same family of the same age will almost always be observed in different time periods, raising concerns about how to account for time effects on outcomes. For example, there might be a general upward time trend in school enrollment that is unrelated to the introduction of the program. When additional data are available on households who did not receive the intervention, they can be used to nonparametrically identify year effects on the outcome variables. That is, when the comparison group data are used in estimation, the vector x_{ijt} can include unrestricted year indicators. In the case where only treatment group data are available, separating time effects from treatment effects in post-program years requires parametric restrictions on the functional form of the time effect. For example, one could assume a linear time trend and that deviations from the linear trend in post-program years represent treatment effects.

3.2 Allowing for Intra-family Investment Decisions

The previous discussion assumed that the regressors x_{ijt} did not include sibling lagged outcomes. As noted in section 2.1, if families make intrafamily investment decisions, then the decision model for one child would include the lagged schooling outcomes for other children in the family as part of the state space.

To allow for this type of decision-making, assume now that the regressors include own child lagged schooling outcomes as well as sibling's outcomes (or summary measures of lagged sibling outcomes, as in our empirical application). We denote the regressor vector by $\mathbf{y}_{jt-1}$. Taking differences across younger and older siblings, matched on z_{ijt} (e.g. age,sex), who are observed s time periods apart gives

$$y_{ijt} - y_{kjt-s} = (\mathbf{y}_{jt-1} - \mathbf{y}_{jt-s-1})\beta + (D_{ijt} - D_{kjt-s})\alpha + \{\varepsilon_{ijt} - \varepsilon_{kjt-s}\}.$$

Allowing for the dependence on sibling's lagged outcomes introduces an endogeneity problem between ε_{kjt-s} and $\mathbf{y}_{jt-1}$, because lagged sibling shocks were known at the time of making decisions about the components of $\mathbf{y}_{jt-1}$ (y_{ijt-1} and y_{kjt-1}). For example, a bad shock experienced by sibling in the past (such as a random illness that leads to missed school), might affect parents' decisions about sending their children to school in period t.

We address the endogeneity problem using instruments for $\mathbf{y}_{jt-1} - \mathbf{y}_{jt-s-1}$. Given the structure of the model, $\mathbf{y}_{jt-s-1}$ can serve as an instrument for the difference $(\mathbf{y}_{jt-1} - \mathbf{y}_{jt-s-1})$, because the shocks $\{\varepsilon_{ijt} - \varepsilon_{kjt-s}\}$ were unknown at the time of making decisions about $\mathbf{y}_{jt-s-1}$. Also, if the model includes other time varying endogenous variables (such as fertility), then characteristics of the family that are fixed across time (for example, years of education of the mother and father) are potential instruments.

3.3 Estimation when Program Effects are Heterogeneous

The above discussion assumed that program effects were homogenous. We now consider the case where program effects can vary across individuals. For example, families who are eligible for the program but are unaware of their eligibility (or of the program) might respond very differently to the program offer than families who choose to take up the proram. Quite possibly, they would have no response to being eligible for treatment. For this and other reasons, it is important to allow for heterogenous responses to treatment.

To simplify the presentation, consider a version of the previous models with no regressors other than z_{ijt} and the treatment indicator D_{ijt},

$$y_{ijt} = \beta_0 + D_{ijt}\alpha_i + \varphi(f_i, z_{ijt}) + \varepsilon_{ijt},$$

where α_i is the treatment effect for individual i (now a random effect). Let B be defined as the subset of observations of C for which matched siblings can be found, that is, $B \subset C$, where

$$B = \{(i, j, t) \text{ such that } S^*(i, z_{ijt}) \text{ is non-empty}\}.$$

Define $\bar{\alpha} = E(\alpha | D_{ijt} = 1, i \in B)$ as the average program effect individuals in B who received the offer of the program. The outcome equation can be written as:

$$\begin{aligned} y_{ijt} &= \beta_0 + D_{ijt}\bar{\alpha} + \varphi(f_i, z_{ijt}) + \{\varepsilon_{ijt} + D_{ijt}(\alpha_i - \bar{\alpha})\} \\ &= \beta_0 + D_{ijt}\bar{\alpha} + \varphi(f_i, z_{ijt}) + u_{ijt}. \end{aligned}$$

When program effects are heterogeneous, the following additional assumption on the random component of the program effect is required to justify application of our estimators:

Assumption 5: $E(u_{ijt} | D_{ijt}, f_i, z_{ijt}) = 0$

This assumption implies that, within families and matched siblings, treatment status is independent of the idiosyncratic component of the treatment effect. While these kinds of independence assumptions can be problematic in situations where treatment is self-selected (see discussion in Heckman, Lalonde and Smith (1999)), the assumption seems relatively benign in this context, where treatment is the offer of the program and the program was unanticipated. The timing of the offer and, therefore, the realization of which child in the family was offered the program at which age did not depend on child-specific responses to the treatment.[15]

Under assumption #5, consider estimation of the parameter $\bar{\alpha}$ as a minimizer of

$$\sum_{i=1}^{T}\sum_{t=1}^{T}[y_{ijt} - D_{ijt}\bar{\alpha} - y_{kjt} + D_{kjt}\bar{\alpha}]^2 1(i \in B, k = S^*(i, z_{ijt})),$$

where the indicator function $1(i \in B, k = S^*(i, z_{ijt}))$ makes it clear that the effective estimation sample consists only of children for whom matches can be found.

We can divide the summation into two components, one component including all observations for which $D_{ijt} = D_{kjt}$ and the other including all those for which $D_{ijt} = 1$ and $D_{kjt} = 0$. (We do not have the case of $D_{kjt} = 1$ and $D_{ijt} = 0$, because the differences are always taken between younger and older siblings.)

$$\sum_{i=1}^{T}\sum_{t=1}^{T}[y_{ijt} - D_{ijt}\bar{\alpha} - y_{kjt} + D_{kjt}\bar{\alpha}]^2 1(i \in B, k = S^*(i, z_{ijt}), D_{ijt} = D_{kjt})$$

$$\sum_{i=1}^{T}\sum_{t=1}^{T}[y_{ijt} - D_{ijt}\bar{\alpha} - y_{kjt} + D_{kjt}\bar{\alpha}]^2 1(i \in B, k = S^*(i, z_{ijt}), D_{ijt} = 1, D_{kjt} = 0),$$

The first expression does not depend on $\bar{\alpha}$ and therefore does not contribute to its estimation. The minimizer of $\bar{\alpha}$ is solely determined from the second term, where the matched sibling pairs differ in terms of their treatment status. Solving for $\hat{\alpha}$, we get

$$\hat{\alpha} = \frac{\sum_{i=1}^{T}\sum_{t=1}^{T}[y_{ijt} - y_{kjt}]1(i \in B, k = S^*(i, z_{ijt}), D_{ijt} = 1, D_{ikt} = 0)}{\sum_{i=1}^{T}\sum_{t=1}^{T} 1(i \in B, k = S^*(i, z_{ijt}), D_{ijt} = 1, D_{ikt} = 0)},$$

which shows that the program effect is estimated by the mean difference in the outcomes of treated children and of those of their matched pairs, for the subset of B for which siblings differ in treatment status.

[15] Similar independence assumptions are commonly invoked in application of matching on observables approaches.

In this example, the sibling pairs for which $D_{kjt} = D_{ijt}$ do not contribute to the estimation of $\bar{\alpha}$. However, in a model with other x_{ijt} regressors and associated coefficients β, the observations would contribute not only to the estimation of β, but also to the estimation of the program effect through the covariance of β and $\bar{\alpha}$.[16]

A limitation of the semiparametric sibling difference approach applied in a situation with heterogenous program effects is that the estimator only identifies the average program effect for the subsample of matched siblings that differ in treatment status. This limitation is an instance of the support problem that also plagues matching procedures (see Heckman, Ichimura and Todd, 1997). The conventional sibling fixed-effect estimator also suffers from the problem; families have to have at least two children to be included in the estimation, so the estimation is valid only for families with two or more children (See Rosenzweig and Wolpin, 1995).

The semiparametric sibling-difference estimator requires that families have at least two children, that children be of the same gender and that we observe them (either contemporaneously or retrospectively) at the same age. Conditioning the analysis on having at least two children at the time of the survey is potentially problematic if fertility is a choice that depends on school attendance shocks. Our instrumental variables estimation procedure accounts for the endogeneity of fertility, but only within the subsample of mothers that have at least two children. Our application is to Mexico, where the birth rate is still relatively high and the large majority of women (about 85%) have two or more children. Some women have not completed their child-bearing at the time of the survey, though. About one-third of the observations are excluded by the restriction that families have at least two children at the time of the survey.

[16] Whether it increases the efficiency of the estimator of the program effect depends on the particular covariance structure.

4 Description of the Oportunidades Program and Datasets

The *Oportunidades* program has two main subsidy components: a schooling subsidy and a food subsidy.[17] To receive these subsidies, households have to satisfy certain coresponsibilities. To receive the food benefit, households have to attend clinics for regular check-ups and informational health talks. To receive the school subsidy, children have to attend school in one of the subsidy-eligible grade levels for at least 85% of days.[18] Because of these coresponsibilities, one would expect to find some program impacts on school enrollment and grade accumulation. However, the magnitude of these impacts is not forecastable, because households can participate to different extents in the program. For example, households can choose to participate only in the health component, so it is possible that the program has no impact on schooling for some participating households. Second, participating households can choose to send only a subset of their children to school for the required time. For example, rather than send two children to school fifty percent of the time, households might now send their older child to school 85% of the time (to receive the subsidy) and keep the other child at home. Thus, even though school enrollment and attendance are coresponsibilities of participating in the program, children within the household may be affected in different ways by the program. One goal of this paper is to assess the magnitude of the *Oportunidades* impacts on a variety of school outcomes. Some of these outcomes, such as school enrollment and educational attainment are closely tied to coresponsibilies of the program, while other outcomes, such as work and fertility, are not. Table 1 shows the subsidy amounts and how they increase with grade level to offset the higher opportunity costs of working for older children. The slightly higher subsidies for

[17] In 2003, an additional component called Jovenes con Oportundades was introduced, which gives a financial incentive for older youth to complete highschool. Through the program, youth earn points that they can use after high school graduation towards certain purposes, such as to pay college expenses, to start a business, to enter a housing lottery, or to put in a savings account with restrictions on withdrawal. In 2003, this new component of the *Oportunidades* program was still in its infancy and was relatively unknown, so we do not expect the youth in our evaluation samples to have been affected by it.

[18] They cannot receive a subsidy more than twice for the same grade. If a child fails a grade more than once, the benefits for that child are discontinued.

girls reflect one of the emphases of the *Oportunidades* program, which is to increase the schooling enrollment and schooling attainment for girls who traditionally have lower enrollment in secondary and higher grade levels.

The *Oportunidades* program is targeted at families who meet the official nation-wide criterion of poverty, which is measured by a marginality index summarizing characteristics of the household. The index depended on criteria that are intended to capture poverty.[19] Households with index scores above a cutoff level are eligible to participate.

The data used in this paper come from the ENCELURB (Urban Evaluation Survey) surveys, which were designed for the purpose of evaluating the impacts of *Oportunidades* in urban areas. The sampling frame was as follows. First, a screener survey was applied in 149 randomly selected blocks from a sample of localities slated for incorporation into the program in 2002 as well as a matched comparison group in 288 blocks, slated for incorporation at a later date.[20] The survey elicited information on economic characteristics needed to calculate the poverty index used to determine program eligibility. The survey also included, for households in areas where the program had been introduced, questions about whether the household was a program beneficiary and whether the respondent knew about the existence of the program. Using the information from the screener, the poverty index was calculated for each of the households and they were classified into three categories: poor (eligible), near-poor (ineligible, but close to the eligibility cut-off), and non-poor. A stratified random sample based on these classifications and on the self-reported beneficiary status was used to select the households into the baseline Urban Household Socio-economic Characteristics

[19] Specifically, the index depends on a crowding index (number of people in the household / number of rooms), whether the head of household is female, whether the household has access to medical service, the total number of children in the household less than eleven years old, years of education of the household head, age of the household head, whether the household has a bath and whether the bath has running water, whether the floor of the house has a dirt floor, whether the household has a gas heating system, a refrigerator, a washing machine, whether assets include a vehicle, whether the home is in a rural area and region of residence dummies for the 19 census regions.

[20] The blocks were chosen by a stratified sampling scheme which placed a higher weight on blocks for which more than 50% of the population is poor, according to the last available census information.

Questionnaire, or "ENCELURB."[21]

Thus far, three ENCELURB surveys have been carried out annually between 2002 and 2004. Retrospective information on education was used to construct a complete educational history for each child/youth in the household. In particular, the baseline 2002 survey asks all individuals between the ages of 6 and 20 about repetition history (grades failed, number of times and age when failed) as well as age at entry to primary school. The follow up rounds in 2003 and 2004 ask about completion over the previous year. Using this information as well as current educational attainment, the history of enrollment and highest grade completed was constructed from the time the children were age 6 until the current age.

Our estimation sample is restricted to households in which mothers are under 40 years of age at the time of the 2002 survey. This restriction was imposed to increase the likelihood that all children ever born to the household were in the household at the time of the survey.[22] There were 6777 such households, consisting of 2615 eligible households, 986 near-poor households as 603 non-poor households living in treatment areas, as well as 1733 eligible households and 373 near-poor households in non-intervention areas.

Figures 1a-1b graph the enrollment rates before and after the program for the treatment group (eligibles in intervention areas) as well as for the two comparison groups. Figure 1a shows enrollment rates in 2001, the year just prior to the initiation of the program. In general, preprogram enrollment rates are lower in the treatment group than in either of the comparison groups, with the difference being largest for older children. Enrollment rates at age 6 are about 0.9, which reflects a

[21] All households who were beneficiaries were selected with probability one to be interviewed. For those reporting that they were not beneficiaries, a random sample was selected for three classifications: poor, near poor and non-poor. In the case of the matched comparison group, all poor households were selected and a random sample of near-poor were selected. Non-poor households in comparison areas were not interviewed. In our analysis, we use weights to reweight the observations back to random sample proportions.

[22] A subsample of women were asked about the number of children ever born to them, so for these women we can assess to what extent children are still living in their household. The average difference between the number of children ever born and the number of children living in the household is small (less than 0.5) for women age 35-39, but increases to 1.2 for women 40-44. For this reason, we restrict the sample to women under age 40.

tendency for delays in the age of starting school. There is a sharp drop in enrollment between ages 12 and 15, which for most children coincides with the time of finishing primary school (grades 1-6). Figure 1b shows the school enrollment in the 2002 post-program year. Much of the preprogram gap between the treatment and comparison groups has been eliminated by 2002, which is suggestive of positive impacts of the program. The closing of the gap occurs across all ages. There is an especially substantial increase in the average enrollment rates of children age 13 and older.

Table 2 presents means and descriptive statistics for our sample in the year 2001, which was the year just prior to the introduction of the program. The three columns show means and standard deviations for the three subsamples corresponding to the eligible intervention group (the treated) and the two comparison groups. The average age of program-eligible children is 10.19, which is similar to the average for the comparison groups. School attendance rates and the highest grade completed are somewhat lower in the eligible intervention group than in the comparison groups. Mother's and father's schooling are also slightly lower. These patterns are consistent with eligible families being poorer on average than ineligible families and with the program having first been introduced into the poorest localities. On average, program eligible households have slightly more than 2 children.

5 Empirical Results

This section presents estimates of program impacts based on the dynamic panel data models of schooling that were described in section two. We first discuss estimates of short-term (one-year) program impacts and then turn to results based on simulating longer-term effects.

As noted in section two, treatment is defined as being offered the program, so that children from program-eligible families constitute the treatment group. Although the estimation procedure could be implemented only using data on treated families, a major advantage of having comparison group data is that they allow nonparametric identification of post-program year effects on the outcome variable. We base our analysis on two alternative comparison groups that are available in the ENCELURB data:

(i) families living in intervention areas who were ineligible for the program but whose index scores fell close to the eligibility cut-off (the near-poor), and

(ii) families living in nonintervention areas who meet the criteria for eligibility but who do not receive the program because the program was not yet available in their area.

There are some relative advantages and disadvantages to using either group. For example, with group (i), we assume that there are no spillover effects or leakage effects of the program on ineligible families, so that their outcomes represent the untreated state.[23] Such an assumption is plausible for families in group (ii), who lived in an area where the program was not available and are unlikely to be affected by the program. On the other hand, our estimation procedure also requires that year effects be the same for the treatment and comparison group, after eliminating an additive locality level effect. Because group (i) lives in the same localities as the treated group, the assumption that year effects are the same for them is reasonable.[24] A possible concern with using group (ii) might be that year effects differ by localities, which is not permitted by the estimation procedure.[25] Because each comparison group has relative strengths and weaknesses, we report estimates for both of the groups.

We estimate the schooling model that was introduced in section two:

$$y_{ijt} = x_{ijt}\beta + D_{ijt}\alpha + \varphi(z_{ijt}, f_i) + \varepsilon_{ijt}$$

using the methods described in section three. For comparison purposes, we also estimate versions of the model using standard cross-section and difference-in-difference estimators. Following the

[23] About 10% of the ineligible families who were close to the cut-off for eligibility report receiving some program benefits. If the program was extended to some of the ineligible families, then our estimated program effects likely understate the impact of the program.

[24] Group (ii) families live in similar broad geographical census regions, but not in the same localities.

[25] The model assumes that locality effects and year effects are additively separable and does not include a locality-year interaction.

discussion in section 2.1, we implement two versions of the model. The "basic" model assumes that parents do not take into account the existence of siblings when they make decisions about any particular child. The "intrafamily" model assumes that parents may take into account the outcomes of all their children in making schooling decisions.

The conditioning variables in the analysis are chosen to capture the state space at the time that schooling decisions were being made. The regressors in the basic model include the child's highest grade completed, indicators for highest grade completed equal to six or nine (which indicate that the child completed primary or secondary school), lagged schooling attendance, the child's gender, mother's and father's education levels and a constant term. The intrafamily model augments this set to include the child's birth order, the number of children in the family, the fraction of children age six or older who are boys, the average age of children in the family, and the average highest grade completed of the children. The models also include year indicator variables.

5.1 Short-term (one-year) program impacts

Table 3a shows the estimated program impact on school attendance under alternative estimators for the basic model and the "near-poor" comparison group. The column labeled "Cross-Section" refers to a conventional cross-section model (without fixed effects) that is implemented using only post-program data. The column labeled "Diff-Diff" is a conventional difference-in-difference estimator that uses all the available years of data (pre-program, baseline and post-program). The "Before-after" specification only uses data on the treated group. As noted in section two, when only treatment group data are used, then post-program year effects are not separately identified from program effects without parametric assumptions. In implementing the before-after estimator, we assume that the year effect follows a linear time trend. The other model specifications include unrestricted year effects. The column labeled "Sibling Fixed Effect" implements a conventional sibling fixed effect estimator, implemented by subtracting family-specific means. All specifications include unrestricted age indicator variables.

The columns "Sibling Diff, Matched on Age" and "Sibling Diff, Matched on Age, Sex" show the results of applying our semiparametric sibling difference estimator, where z_{ijt} includes age alone, or age and sex. These estimators are nonparametric with respect to how age and gender affect the outcomes. They also permit interactions between the fixed effect, age and gender. Because these estimators restrict matched siblings to be of the same age and sex, they effectively exclude a fraction of the sample in estimation. To assess how this exclusion affects the empirical estimates, we also present as a benchmark in the last column of the table the standard sibling fixed effect estimator, implemented using the matched data subsample.

As seen in Table 3a, most of the estimated coefficients have the expected sign and are statistically significant. In terms of magnitude, the most important determinants of current attendance are the highest grade completed and lagged attendance. Having attended in the previous period makes attendance in the current period much more likely, which is not surprising if children attend school for consecutive periods and then drop out.[26] Having just finished primary or secondary school makes attendance less likely. Children whose parents have more education are also more likely to attend school. The estimated coefficients on father's and mother's education are both statistically significantly different from zero and are almost identical.

Results based on a standard difference-in-difference estimator indicate that the program increases attendance by 3.5 percentage points. The estimate based on the "before-after" estimator would suggest that the program had no effect on school attendance, although this estimate is far outside the range of the others shown in the table. Using the sibling fixed effect estimator, we obtain an estimate of 2.5 percentage points. The sibling difference estimators generally yield larger estimates, ranging from 4.1-5.1 percentage points. The results shown in the last column, which implements the sibling fixed effect estimator on the matched sample, suggests that the higher impact

[26] Many factors could explain this type of schooling pattern, such as a Ben-Porath (1969) type of investment model, fixed costs of reentering after dropping out, or an opportunity cost of schooling that is rising with age.

estimate obtained using the semiparametric estimators may be partly due to the restriction of the sample to matched sibling pairs.

Table 3b shows the same set of results, obtained for the alternative comparison group of eligible families living in non-intervention areas. The impact estimates are also positive, but they tend to be smaller in magnitude than the ones in Table 3a. The estimates derived from the sibling-based estimators range from 1.4-1.7.[27]

Table 4a and 4b present analogous results for the intrafamily specification, which includes as additional conditioning variables summary measures of family composition and lagged sibling outcomes. As noted in section three, estimating the intrafamily model requires application of instrumental variables. The following variables are treated as endogenous: the difference between siblings in the highest grade completed, the difference in the average highest grade completed, the difference in the siblings' lagged school attendance, the difference in the indicators of highest grade completed equal to six or nine years, the difference in the number of children present in the household, and the difference in the average age of children in the household. The set of instruments include the mother's and father's highest grade completed, an indicator for whether the siblings are twins, differences in the year indicator variables, the difference in the fraction of boys in the household greater than six years old, and the age indicator variables.[28] The instrument set also includes the following variables that are measured at the time of the oldest sibling observation: the number of children, attendance, indicators for highest grade completed equal to 6 or 9, fraction of boys greater than or equal to six years old, average highest grade completed of all children, average age of siblings, year indicators, and the treatment indicator. The use of these instruments is justified by the structure of the dynamic model, as described in section three. The R-squares from

[27] The before-after estimate is not shown in this table, because it does not depend on the comparison group and would be the same as in Table 3a.

[28] The estimation procedure matches children on age. Therefore, the age effect is eliminated by the differencing and age can be used as an instrument. In specifications that also match on gender, we use gender as an additional instrument.

the first stage regressions range from 0.2 to 0.5. Interestingly, application of two stage least squares in this context is not subject to the common problem of weak instruments, because the error term in the first stage equation is uncorrelated with the error term in the second stage equation. [29]

As seen in Tables 4a and 4b, the additional conditioning variables are statistically significant at conventional levels, which we interpret as evidence that families do take into account siblings' characteristics when making schooling decisions. In the specifications that do not allow for fixed effects, a higher average highest grade completed of siblings (holding constant the average age) increases the probability of a child attending school. However, controlling for fixed effects, we see that a higher average highest grade completed of siblings lowers the probability that a child attends school. Such a pattern might be expected if parents consider schooling across children to be substitutes. For example, if an older sibling has a lot of schooling, then it might be less important that a younger child attends school.

Comparing Tables 4a with Tables 3a, we see that the program impact estimates based on the standard cross-section and difference-in-differences specifications are relatively invariant to changing the set of included regressors and instrumenting for them. The estimates based on the cross-section and difference-in-differences specifications also tend to be much smaller than those obtained using the sibling-based procedures. The before-after estimator again yields an estimate outside the range of the others and would indicate no effect of the program. The estimates based on the sibling difference iv procedures range from 4.6-6.0, which is similar to the estimates obtained using the basic model specification. The statistical significance of the additional regressors indicate strong support for the intrafamily rather than the basic model, but the short-term program impact estimates are not very sensitive to which model is used. Below, we demonstrate that the longer-term impact estimates are sensitive to the choice of model.

Table 4b presents results based on the alternative comparison group of eligibles leaving in

[29] See Hahn and Hausman (2002) for the formula for the bias of the two-stage-least-squares estimator and how it depends on the covariance of the error terms.

nonintervention areas. The program impact estimates are generally smaller in magnitude than those observed in Table 4a. The largest impact estimate of 4.2 percentage points is obtained using the semiparametric difference estimator that is nonparametric with regard to child age and gender.

Table 5 presents program impact estimates for different age and gender groups, obtained using the sibling difference iv estimator applied to the intrafamily model, where the children are matched on age in implementing the estimator (i.e. z_{ijt} includes age). The second column presents estimates for the near-poor comparison group and the third column for the eligible, non-intervention area comparison group. The disaggregated estimates reveal a pattern of significant variation in impacts by age. For boys and girls age 6-11, the impact estimates are not statistically significantly different from zero. However, there are substantial impact estimates for the age 12-17 group. For older boys, the impact estimates are 12.0 percentage points using the near-poor comparison group and 9.0 percentage points using the group of eligibles in nonintervention areas. We also find large impacts for older girls, on the order of 12.6-14.4 percentage points.

The lack of discernible impacts for young children is perhaps not surprising, given the very high attendance rates for children at these ages even in the absence of the program, which places an upper bound on the conceivable effects of the program. For example, if attendance is already 96% without the program, then program could at most increase attendance by 4%. As previously noted, attendance rates in the absence of the program fall dramatically after completion of primary school and again after completion of secondary school. Thus, there is a larger scope for the program to have an effect on the attendance of older children, which is what we observe. The higher school subsidies for higher grades may also in part explain why we observe larger impacts on older children, although older children also face higher opportunity costs of schooling (e.g. higher wage offers, which are being captured in the estimation by the age-specific family effect).

Table 6 presents analogous results when the outcome variable is grades completed rather than school attendance. Children who were not attending are deemed to not have completed a grade.

The pattern of larger impacts for the older age group is generally upheld. However, we also observe an effect of the program on grades completed for younger (age 6-11) boys and girls when the near-poor comparison group is used.

5.2 Simulated longer-term program impacts

We next use the estimated model to simulate the effects of longer-term participation in the program. Specifically, we simulate children's school-going from age 6-17 first without the program and then assuming the program was offered in every year. For the intrafamily model, we assume a family with two children (boys), who are spaced two years apart. We simulate both children's school-going, updating the state variables in each time period as required by the model to reflect changes in lagged own schooling outcomes and/or lagged siblings' schooling outcomes. We performed this simulation using both the basic and intrafamily models and using the sibling difference iv matched on age estimator (estimated coefficients shown in Table 3a and Table 4a).[30]

Table 7 shows how the program affects the highest grade completed over the longer-term. The basic model indicates that if the program were available from ages 6-17, then the highest grade completed would increase by 0.575 years.[31] The intrafamily model indicates a different impact of 0.426 years, which is 25% smaller. Thus, the small differences in short-term impact estimates that were observed between the basic model and the intrafamily model are magnified in the longer-term predictions.

[30] Our simulation assumes that the family would be eligible throughout for the program and does not attempt to model the dynamics of program eligibility. The simulation also assumes that fertility is exogenous and is unaffected by the the program. Incorporating indirect effects of the program on schooling outcomes that changes in fertility would require specifying an auxilliary model of how fertility responds to the program.

[31] The estimated long-term impacts obtained under the intrafamily model are similar in magitude to those obtained (using alternative methods of doing the long-term prediction) by Schultz (2004), Behrman, Sengupta and Todd (2005), and Todd and Wolpin (2002) for the rural evaluation sample.

6 Conclusions

This paper explores the use of dynamic schooling models and sibling-based estimators for evaluating both the short-term and longer-term effects of program interventions. Our approach uses retrospective data on older siblings' schooling histories to study how an intervention affects younger siblings. The statistical models are motivated by dynamic behavioral models in which parents make decisions about whether to send children to school. We considered two behavioral models, one in which parents make schooling decisions about individual children without regard to their siblings (the basic model) and another in which their schooling decisions could also depend on siblings and lagged sibling outcomes (the intrafamily model). The empirical results strongly support the use of the richer intrafamily model.

This paper also proposed an estimation method that combines matching, differencing and instrumental variables in estimating program effects. It can be viewed as an application of Ahn and Powell's (1993) pairwise differencing estimator for the selection model to a panel data, treatment effects setting. The estimator does not impose any parametric structure on how age and gender affect outcomes. It also allows for arbitrary interactions between age, gender and the fixed effect. For example, it accommodates unobserved local labor market effects that may differ by age and gender and affect opportunity costs of school attendance. Additive year effects can also be handled nonparametrically, when there is an alternative comparison group data on families not exposed to the program.

The identifying assumptions that justify the application of sibling difference estimators may be more plausible than those invoked by other evaluation estimators, such as cross-sectional matching estimators, which are widely used in evaluating effects of programs. Matching estimators typically pair treated individuals to observably similar untreated individuals from other families. A concern with the application of these estimators is that there may be unobservable determinants of outcomes that differ across the treatment group and the matched comparison group, which could bias the estimated program impacts. The sibling difference approach also uses matching when it restricts comparisons to be among children of the same gender and age. However, it imposes the additional restriction that children come from the same family, which controls for family-specific unobservable determinants of outcomes.

We apply the sibling difference estimators as well as other conventional evaluation estimators to study the effects of the *Oportunidades* conditional cash transfer program on schooling outcomes of children and youth living in urban areas of Mexico. The empirical results based on the proposed estimator as well as those based on some common alternative estimators show significant impacts of the program on school attendance. The before-after estimator found no effect of the program, but, as noted, this estimator requires strong parametric assumptions on time effects.

Our preferred estimates, based on the sibling difference iv estimator applied to the intrafamily model, reveal substantial heterogeneity in impacts by age and gender of the child. For younger children age 6-11, the program appears to have no discernible effect on school attendance. The lack of impact is partly due to high attendance levels for children in this age range even in the absence of the program. For older girls and boys, we observe substantial program effects ranging from 9.0-14.4 percentage points, depending on gender and on the source of comparison group data used. We also used the estimated dynamic model to simulate long-term program effects for a typical family. Our simulations indicate that a child participating in the program from age 6-17 would complete about 0.5 additional years of schooling.

The kind of retrospective data required to apply the sibling difference methods studied in this paper are rarely collected in evaluation studies, but they could easily be collected. Finally, while we have applied the method in the context of studying a conditional cash transfer program, there are other potential applications. For example, the estimators may be useful for evaluating effects of child health interventions, where retrospective data on illnesses might be collected. They could also be used to study effects of other educational interventions, such as school voucher programs, high school graduation bonus programs, or college scholarship programs.

References

[1] Ahn, H. and J. L. Powell (1993): "Semiparametric Estimation of Censored Sample Selection Models with A Nonparametric Selection Mechanism." *Journal of Econometrics*, 58, No. 1-2, 3-29.

[2] Altonji, J. G. and Dunn, T. A. (1996). 'Using siblings to estimate the effect of school quality on wages.' *Review of Economics and Statistics*, vol. 78, no. 4, (November) pp. 665-71.

[3] Becker, Gary S. 1999. "Bribe" Third World Parents to Keep their Kids in School", *Business Week*, November 22nd.

[4] Becker, G. and Tomes, N. (1976). 'Child endowments and the quantity and quality of children,' *Journal of Political Economy*, vol. 84, no. 4, part 2, pp. S143-162.

[5] Behrman, Jere R., Piyali Sengupta and Petra Todd, 2005, "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment," *Economic Development and Cultural Change,* 54(1), 237-275.

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

[7] Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), part 1, 352-65.

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

[9] Cameron, Steven and James J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," in *Journal of Political Economy*, Vol. 106, no. 2, p.262-333.

[10] Coady, D. and S. Parker (2003): "Combining Self-Selection with Proxy Means Targeting in Mexico," Mimeo.

[11] Currie, Janet and Ducan Thomas (1995): "Does Head Start Make a Difference?", *American Economic Review*, 85(3):341-64.

[12] Hahn, Jinyong and Jerry Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," Econometrica, Vol. 70, 163-190.

[13] Heckman, James, Hidehiko Ichimura and Petra Todd, 1997, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, Vol. 64(4), 605-654.

[14] Heckman, James, Robert Lalonde and Jeffrey Smith (1999): "The Economics and Econometrics of Active Labor Market Programs" in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 3A* (Amsterdam: North-Holland), 1865-2097.

[15] Keane, Michael and Kenneth I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, Vol. 105, No. 3, 473-522.

[16] Krueger, Alan. 2002. "Economic Scene: A Model for Evaluating the Use of Development Dollars, South of the Border." *New York Times*. May 2nd.

[17] Parker, Susan W., Jere R. Behrman and Petra E. Todd, 2004, "Medium-Term Effects on Education, Work, Marriage and Migration in Rural Areas," Philadelphia: University of Pennsylvania, Technical Document Number 1 on the Evaluation of Oportunidades 2004, mimeo.

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

[19] Parker, Susan W. "Evaluation del impacto de Oportunidades sobre la inscripcion, reprobacion y abandono," mimeo, 2004.

[20] Porter, Jack R. (1996): "Essays in Econometrics," MIT Ph.D. Dissertation..

[21] Rosenzweig, Mark R. (1986): "Bith Spacing and Sibling Inequality: Asymmetric Information within the Family," *International Economic Review*, 27, 55-76.

[22] Rosenzweig, Mark R. and Wolpin, Kenneth I. (1988): "Heterogeneity, Intrafamily Distribution and Child Health," in *Journal of Human Resources,* Volume XXIII, 4, p. 437-461.

[23] Rosenzweig, Mark R. and Wolpin, Kenneth I. (1986): "Evaluating the Effects of Optimally Distributed Public Programs: Child Health and Family Planning Interventions," in *American Economic Review*, Volume 76, No. 3, p. 470-482.

[24] Rosenzweig, Mark R. and Wolpin, Kenneth I. (1995): "Sisters, Siblings and Mothers: The Effect of Teen-age Childbearing on Birth Outcomes in a Dynamic Family Context," in *Econometrica*, Vol. 63, No. 2, 303-326.

[25] Schultz, T. Paul, 2000, "Impact of PROGRESA on school attendance rates in the sampled population," February. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C.

[26] Schultz, T. Paul, 2004, "School subsidies for the poor: Evaluating a Mexican strategy for reducing poverty," *Journal of Development Economics*, vol. 74, 199-250

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

[28] Todd, Petra and Kenneth Wolpin, 2003, " Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," manuscript.

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

[30] Todd, Petra E. and Kenneth I. Wolpin (2002): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," forthcoming in *American Economic Review*.

[31] Wooldridge, Jeffrey M. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press: Cambridge, MA.

Table 1
Transfer Levels by Grade and Gender
(pesos per month in second semester of 2002)

	Boys	*Girls*
Primary School		
Grade 3	100	100
Grade 4	115	115
Grade 5	150	150
Grade 6	200	200
Middle School		
Grade 7	290	310
Grade 8	310	340
Grade 9	325	375
High School		
Grade 10	490	565
Grade 11	525	600
Grade 12	555	635

Note: Education transfers are conditional on 85% school attendance. Households also receive a monthly “food transfer” of 150 pesos, conditional on regular attendance at health centers. $1US=10 Mexican pesos.

Table 2
Baseline characteristics (year 2002) of the treatment and comparison groups
Children age 6-17

	Eligible Intervention		Eligible Comparison		Near-poor intervention	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age	10.19	3.01	10.07	2.95	10.63	3.19
Attend	0.83	0.38	0.88	0.33	0.86	0.34
Highest grade completed	3.42	2.63	3.73	2.69	4.14	2.86
Mother's schooling	4.75	3.02	5.48	3.03	6.00	3.08
Mother's age	32.16	4.16	32.21	4.10	32.78	4.15
Father's schooling	5.13	3.26	5.84	3.17	6.51	3.33
Father's age	35.80	6.71	35.76	6.69	36.81	7.09
Average number of children in the household	2.29	1.22	2.29	1.24	2.08	1.06
Number of Households	2615		1733		986	

Table 3a
Estimated Program Impacts on Attendance
Basic Model, Robust Standard Errors in Parentheses
Groups: Eligible and Ineligible Households (Near Poor) Living in Intervention Areas

Variable	Estimator						
	Cross-Section Post Program	Diff-Diff	Before-After	Sibling Fixed Effect	Sib. Diff. Matched on Age	Sib. Diff. Matched Age, Sex	Sibling FE - Matched Sample
Treated*	0.010	0.035	-0.014	0.025	0.041	0.051	0.052
	(0.005)	(0.006)	(0.008)	(0.006)	(0.004)	(0.011)	(0.011)
Highest grade completed	0.033	0.028	0.028	-0.003	0.013	0.013	0.016
	(0.0022)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)
Lagged Attendance	0.406	0.431	0.441	0.372	0.368	0.356	0.379
	(0.012)	(0.007)	(0.008)	(0.004)	(0.008)	(0.010)	(0.011)
Highest Grade Completed. = 6	-0.063	-0.096	-0.097	-0.108	-0.109	-0.106	-0.105
	(0.009)	(0.008)	(0.009)	(0.005)	(0.009)	(0.011)	(0.011)
Highest Grade Completed = 9	-0.068	-0.129	-0.128	-0.125	-0.177	-0.212	-0.219
	(0.017)	(0.015)	(0.018)	(0.009)	(0.021)	(0.028)	(0.028)
Male	-0.003	-0.004	-0.003	-0.005	-0.004	…	…
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)		
Mother's highest grade completed	0.002	0.004	0.004	…	…	…	…
	(0.0008)	(0.0005)	(0.0006)				
Father's highest grade completed	0.002	0.004	0.005	…	…	…	…
	(0.0007)	(0.0005)	(0.0005)				
Constant	0.179	0.627	-58.21		-0.013	-0.018	-0.018
	(0.015)	(0.080)	(5.8)		(0.004)	(0.004)	(0.005)
Eligible+	…	-0.025	…	…	…	..	…
		(0.004)					
Linear Time Trend	…	…	0.029	…	…	…	…
			(0.003)				
Includes Age Indicators	Yes	Yes	Yes	Yes	No	No	Yes
Number of Observations	15401	41935	30410	57092	28495	18019	18019
R-square	0.50	0.38	0.38	0.36	0.17	0.16	0.16

* Treated means that the household satisfied the eligibility criteria to participate in the program. All specifications also included unrestricted age effects. All but the before-after specification include unrestricted year effects. The before-after specification includes instead a linear time trend.
+The coefficient on Eligible gives the preprogram difference in mean enrollment between the eligible intervention and ineligible intervention groups.

Table 3b
Estimated Program Impacts on Attendance
Basic Model, Robust Standard Errors in Parentheses
Groups: Eligibles Living in Intervention and Nonintervention Areas

Variable	Estimator: Cross-Section Post Program	Diff-Diff	Sibling Fixed Effect	Sib. Diff. Matched on Age	Sib. Diff. Matched Age, Sex	Sibling FE on Matched Sample
Treated*	0.008	0.017	0.017	0.014	0.014	0.015
	(0.003)	(0.005)	(0.004)	(0.007)	(0.008)	(0.008)
Highest grade completed	0.027	0.028	-0.002	0.009	0.012	0.016
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Lagged Attendance	0.413	0.425	0.323	0.355	0.348	0.380
	(0.010)	(0.006)	(0.004)	(0.007)	(0.009)	(0.010)
Highest Grade Completed. = 6	-0.079	-0.104	-0.108	-0.101	-0.083	-0.084
	(0.007)	(0.007)	(0.004)	(0.008)	(0.010)	(0.009)
Highest Grade Completed = 9	-0.058	-0.105	-0.104	-0.109	-0.126	-0.133
	(0.015)	(0.013)	(0.008)	(0.019)	(0.025)	(0.025)
Male	0.00001	-0.003	-0.007	-0.003	…	…
	(0.003)	(0.003)	(0.003)	(0.003)		
Mother's highest grade completed	0.002	0.004	…	…	…	…
	(0.0006)	(0.0005)				
Father's highest grade completed	0.003	0.005	…	…	…	…
	(0.0005)	(0.0004)				
Intervention Group+	…	-0.007	…	…	…	…
		(0.004)				
Constant	0.218	0.616	0.642	-0.017	-0.018	-0.018
	(0.013)	(0.073)	(0.035)	(0.003)	(0.004)	(0.004)
Includes Age Dummies	Yes	Yes	Yes	No	No	Yes
Number of Observations	20378	53007	69261	35241	22450	22450
R-square	0.49	0.38	0.36	0.16	0.16	0.15

* Treated means that the household lived in an intervention area and satisfied the eligibility criteria to participate in the program. All specifications also included unrestricted age effects. All but the before-after specification include unrestricted year effects. The before-after specification includes instead a linear time trend.

+ The coefficient on Intervention Group gives the preprogram difference in mean enrollment between the eligible intervention and nonintervention groups.

Table 4a
Estimated Program Impacts on Attendance
Intrafamily Model, Robust Standard Errors in Parentheses
Groups: Eligible and Ineligible Households (Near Poor) Living in Intervention Areas

Variable	Estimator				
	Cross-Section Post Program	Diff-Diff	Before-After	Sib. Diff. IV Matched on Age	Sib. Diff. IV Matched Age, Sex
Treated*	0.016 (0.005)	0.037 (0.006)	-0.011 (0.008)	0.046 (0.012)	0.060 (0.015)
Highest grade completed	0.026 (0.003)	0.014 (0.002)	0.013 (0.003)	0.008 (0.004)	0.005 (0.005)
Lagged Attendance	0.403 (0.012)	0.427 (0.007)	0.436 (0.008)	0.454 (0.019)	0.434 (0.023)
Highest Grade Completed. = 6	-0.066 (0.009)	-0.100 (0.008)	-0.100 (0.010)	-0.113 (0.164)	-0.117 (0.019)
Highest Grade Completed = 9	-0.064 (0.017)	-0.125 (0.015)	-0.125 (0.018)	-0.224 (0.037)	-0.227 (0.051)
Birth order	0.006 (0.004)	-0.001 (0.003)	-0.002 (0.003)	…	…
Number of Children	-0.007 (0.003)	-0.004 (0.002)	-0.003 (0.002)	-0.070 (0.042)	-0.050 (0.047)
Fr. of boys >= 6	-0.014 (0.008)	-0.006 (0.006)	-0.009 (0.007)	-0.002 (0.016)	-0.009 (0.028)
Male	0.002 (0.006)	-0.001 (0.005)	0.001 (0.005)	-0.003 (0.004)	…
Mother's highest grade completed	0.001 (0.001)	0.003 (0.0005)	0.004 (0.0006)	…	…
Father's highest grade completed	0.001 (0.001)	0.003 (0.0005)	0.004 (0.0005)	…	…
Average highest grade completed of all children	0.010 (0.003)	0.023 (0.003)	0.030 (0.003)	-0.053 (0.020)	-0.050 (0.022)
Average age of all children	-0.009 (0.002)	-0.016 (0.002)	-0.017 (0.002)	0.067 (0.022)	0.079 (0.024)
Eligible+	…	-0.022 (0.004)	…	…	…
Constant	0.248 (0.021)	0.737 (0.081)	-52.54 (6.15)	-0.019 (0.005)	-0.024 (0.007)
Linear Time Trend	…	..	0.027 (0.003)	…	…
Includes Age Dummies	Yes	Yes	Yes	No	No
Number of Observations	15083	41617	30163	21117	13366
R-square	0.50	0.38	0.38	0.16	0.15

* Treated means that the household satisfied the eligibility criteria to participate in the program. All specifications also included unrestricted age effects. All but the before-after specification include unrestricted year effects. The before-after specification includes instead a linear time trend.
+The coefficient on Eligible gives the preprogram difference in mean enrollment between the eligible and ineligible groups living in intervention areas.

Table 4b
Estimated Program Impacts on Attendance
Intrafamily Model, Robust Standard Errors in Parentheses
Groups: Eligibles Living in Intervention and Nonintervention Areas

Variable	Estimator CS Post Program	Diff-Diff	Sib. Diff. IV Matched on Age	Sib. Diff. IV Matched Age, Sex
Treated*	0.009	0.018	0.012	0.042
	(0.004)	(0.005)	(0.008)	(0.014)
Highest grade completed	0.022	0.014	0.005	0.004
	(0.003)	(0.002)	(0.003)	(0.005)
Lagged Attendance	0.409	0.418	0.431	0.442
	(0.010)	(0.006)	(0.017)	(023)
Highest Grade Completed. = 6	-0.082	-0.107	-0.094	-0.117
	(0.008)	(0.007)	(0.013)	(0.018)
Highest Grade Completed = 9	-0.058	-0.102	-0.173	-0.228
	(0.015)	(0.014)	(0.033)	(0.051)
Birth order	0.001	-0.0004	...	...
	(0.003)	(0.002)		
Number of Children	-0.0001	-0.002	-0.066	0.048
	(0.002)	(0.001)	(0.036)	(0.047)
Fr. of boys >= 6	-0.006	0.005	-0.005	-0.007
	(0.007)	(0.005)	(0.013)	(0.024)
Male	0.002	-0.006	-0.002	…
	(0.005)	(0.004)	(0.003)	
Mother’s highest grade completed	0.003	0.003	…	…
	(0.0006)	(0.0005)		
Father’s highest grade completed	0.002	0.004	...	…
	(0.0006)	(0.0004)		
Average highest grade completed of all children	0.009	0.023	0.112	-0.030
	(0.003)	(0.002)	(0.020)	(0.021)
Average age of all children	-0.007	-0.015	0.112	0.013
	(0.002)	(0.002)	(0.020)	(0.024)
Constant	0.261	0.719	-0.018	0.002
	(0.019)	(0.074)	(0.005)	(0.006)
Intervention Group+	…	-0.007	...	…
		(0.004)		
Includes Age Dummies	Yes	Yes	No	No
Number of Observations	20028	52605	26986	11183
R-square	0.49	0.38	0.15	0.22

* Treated means that the household lived in an intervention area and satisfied the eligibility criteria to participate in the program. All specifications also included unrestricted age effects. All but the before-after specification include unrestricted year effects. The before-after specification includes instead a linear time trend.
+The coefficient on Intervention Group gives the preprogram difference in mean enrollment between the eligible intervention and nonintervention groups.

Table 5
Estimated Program Impacts on Attendance by Demographic Group
Intrafamily Model, Sibling Diff IV Estimator Matched on Age and Gender

Demographic Group	Groups: Eligible and Ineligible Households (Near Poor) Living in Intervention Areas	Groups: Eligibles Living in Intervention and Nonintervention Areas
Boys		
Age 6-17	0.074	0.017
	(0.021)	(0.013)
Age 6-11	0.036	-0.005
	(0.024)	(0.014)
Age 12-17	0.120	0.090
	(0.057)	(0.038)
Girls		
Age 6-17	0.033	0.004
	(0.021)	(0.013)
Age 6-11	0.032	-0.017
	(0.025)	(0.014)
Age 12-17	0.126	0.144
	(0.042)	(0.045)

Table 6
Estimated Program Impacts on Completed Grades by Demographic Group and by Comparison Group, Sibling Diff IV
Intrafamily Model

Demographic Group	Groups: Eligible and Ineligible Households (Near Poor) Living in Intervention Areas	Groups: Eligibles Living in Intervention and Nonintervention Areas
Boys		
Age 6-17	0.189	0.0002
	(0.054)	(0.029)
Age 6-11	0.203	-0.048
	(0.077)	(0.037)
Age 12-17	0.154	0.104
	(0.106)	(0.054)
Girls		
Age 6-17	-0.037	-0.021
	(0.036)	(0.034)
Age 6-11	0.109	-0.050
	(0.055)	(0.034)
Age 12-17	0.078	0.097
	(0.059)	(0.068)

Table 7
Simulated Program Effect on Attendance and on
Highest Grade Completed by Age after Longer Terms of Exposure

Age	Basic Model[a]		Intrafamily Model [a,b] Two Child Family	
	Attendance (Percent)	Highest Grade Completed	Attendance (Percent)	Highest Grade Completed
6	3.81	0.038	4.50	0.045
7	3.16	0.070	3.48	0.080
8	3.35	0.103	3.22	0.112
9	3.36	0.137	2.88	0.141
10	3.15	0.168	2.69	0.168
11	3.21	0.201	2.44	0.192
12	4.37	0.244	3.13	0.224
13	5.65	0.301	3.60	0.260
14	6.98	0.370	4.28	0.302
15	5.89	0.429	3.01	0.333
16	7.12	0.500	4.29	0.375
17	7.50	0.575	5.04	0.426

a Regressions based on the same observation set using sibling difference IV estimator, matched on age.
b Assumes two male children who are two years apart in age.

Figure 1a: Pre-program (2001) Enrollment by Age

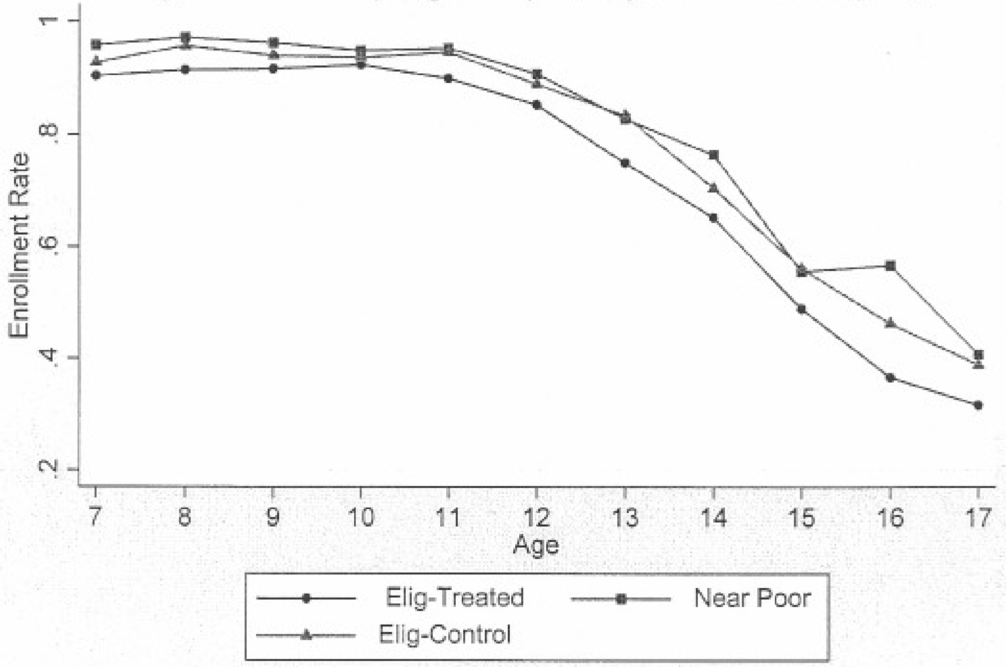

Figure 1b: Post-program (2002) Enrollment by Age