Endogenous Peer Effects in School Participation*

Gustavo J. Bobonis and Frederico Finan

First version: September 2002

This version: February 2006

Abstract: The use of experimental designs has enabled researchers to identify social interactions or neighborhood effects on individual behavior. However, a remaining obstacle in the literature has been the inability to distinguish between peer effects that are determined by a person's reference group *behavior* (endogenous peer effects), and effects that are generated as a result of specific background characteristics of the groups themselves (contextual peer effects). This paper identifies and estimates endogenous peer effects on children's school participation decisions using evidence from the Progresa program. Under Progresa, payments were provided to poor mothers conditional upon school enrollment of their children. Because program eligibility was randomly assigned, we use this exogenous variation in school participation to identify peer effects on the school enrollment of ineligible children residing in the same communities. We find that peers have considerable influence on the enrollment decision of program-ineligible children, and these effects are nonlinear and concentrated among children from relatively poorer households. Our findings imply that educational policies aimed at encouraging enrollment can produce large social multiplier effects.

* A previous version of this paper was entitled "Do Transfers to the Poor Increase the Schooling of the Non-Poor: The Case of Mexico's Progresa Program". We are grateful to David Card, Ken Chay, Alain de Janvry, Asim Khwaja, David S. Lee, Craig McIntosh, Ted Miguel, Elisabeth Sadoulet, and T. Paul Schultz, whose suggestions greatly improved the paper. We also thank Michael Baker, Rob McMillan, Aloysius Siow, and seminar participants at UC Berkeley, Toronto, and NEUDC 2003 & 2005 for helpful comments. We thank Caridad Araujo, Paul Gertler, Sebastián Martínez, Iliana Yaschine, and the staff at Oportunidades for providing administrative data and for their general support throughout. Bobonis acknowledges financial support from the Institute of Business and Economics Research at UC Berkeley and NICHD Training Grant (T32 HD07275). Finan acknowledges financial support from the Social Science Research Council.

Contacts: G. Bobonis, Department of Economics, University of Toronto, Sidney Smith Hall, 100 Saint George Street Room 4057, Toronto, Ontario, M5S 3G3, Canada. Tel: 416-946-5299. E-mail: gustavo.bobonis@utoronto.ca
F. Finan, Department of Agricultural and Resource Economics, UC Berkeley, Giannini Hall 304, Berkeley, CA, 94720, USA. Tel: 510-643-5419. Fax: 510-643-8911. E-mail: finan@are.berkeley.edu.

1. Introduction

Recent empirical studies have made important contributions towards identifying the causal effect of social interactions or neighborhood effects on individual behavior. Sacerdote (2001), using the random assignment of roommates in a U.S. college, finds evidence of peer effects on the level of academic effort and membership in social organizations of college students. Kling, Ludwig, and Katz (2005) use experimental variation in assignment to different types of voucher relocation programs in five U.S. cities to identify long-term neighborhood effects on youth crime and find differential effects among females and males. On the other hand, Oreopoulos (2003) uses quasi-experimental variation in assignment to different types of public housing units in Toronto and finds no long-term neighborhood effects on individuals' labor market outcomes.[1]

While the use of experimental designs have enabled these studies to properly identify social interactions, a remaining obstacle in the literature has been the inability to distinguish between peer effects that are determined by a person's reference group *behavior* (endogenous peer effects), from those that are generated as a result of specific background characteristics of the groups themselves (contextual peer effects) (Manski 1993). The distinction between these two effects is crucial for policy, because endogenous peer effects imply potentially large social multiplier effects and efficiency gains through the feedback in the behavior of individuals within an existing social network (e.g., positive student behavior leads to more positive behavior in the network) (Hoxby 2000), as opposed to potential multipliers from contextual effects as a result of sorting into peer groups (Epple and Romano 1998; Bayer et al. 2004). This may be of particular importance for education policy. For instance, understanding the sources of peer effects, such as strategic complementarities in student effort (Kremer, Miguel and Thornton 2004) and 'learning' from others based on peers' cognitive abilities, may have important implications for the design of educational systems.

In this paper, we identify endogenous peer effects in children's school participation decisions using evidence from a human development program in rural Mexico. The Progresa program, initiated by the Mexican government in 1997, provides cash transfers to marginalized households in rural areas. The transfer is paid to mothers contingent on their children's primary and secondary school attendance and family visits to health services. Five hundred and six communities were selected to participate in an experimental evaluation of the program; the communities were randomly divided into two groups, with the treatment group being phased-in to the program in March-April 1998 and the control group in November-December 1999. Within these selected communities, a poverty indicator was constructed at baseline to classify eligible and ineligible households. While household eligibility was determined within

[1] Experimental and non-experimental studies in developing countries that examine strategic complementarity effects on student effort and pupil achievement include Kremer, Miguel, and Thornton (2004) and Ding and Lehrer (2005).

all (treatment and comparison group) communities, only households below a welfare threshold and within the treatment villages became program beneficiaries during the evaluation period.

Using experimental variation in the induced school participation of the subset of eligible children in these communities, we can identify the endogenous peer effects in school enrollment among children who were ineligible for the program within the program communities. Our main results suggest that children have an increased likelihood of attending secondary school of approximately 5 percentage points as a result of an increase of 10 percentage points in the network enrollment rate; however, these estimates are only marginally different from zero. Substantially larger effects of approximately 6.5 percentage points are found for children of relatively poorer households within the ineligibles group - a subgroup possibly more likely to be induced to continue their schooling, to the extent that these children are more likely to interact with treated children in these villages.

Moreover, our findings suggest that endogenous peer effects not only exist but operate in a nonlinear manner. Based upon semi-parametric estimates, an increase in the enrollment rates of a child's reference group has a substantially larger effect among children with peers who have low enrollment propensity. The magnitude of the effect has immediate policy implications, since it implies important efficiency gains from mixing students with high and low enrollment propensities, if enrollment propensities translate themselves into greater educational attainment or achievement. In summary, the study provides empirical evidence that behavioral (endogenous) peer effects are important determinants in children's school enrollment decisions. This implies that children may have a high degree of agency in their decision-making, even within poor families in developing countries.

A potential concern with our identification strategy is that the program may have affected ineligible children through other mechanisms. We follow two strategies to demonstrate that this is not the case. First, we exploit the richness of the data to test whether other potential externalities from program impacts or particular intricacies of the program had an effect on the behavior of ineligible households. We do not find any evidence of improvements in school quality, or that the program affected the consumption of ineligible households or children's health. Secondly, we condition on a large number of predetermined mean village-level contextual and environmental characteristics that may be correlated with the impacts of the intervention, and show that the effects are robust to these specifications. This sensitivity analysis confirms the validity of the identifying assumptions of the model.

We see our current work as complementing previous studies which have examined endogenous peer effects in school participation decisions (Cipollone and Rosolia 2003; Gaviria and Raphael 2001). Cipollone and Rosolia (2003) use exogenous variation in the school attainment of men as a result of an earthquake in Italy to identify the effect on the school attainment of women in these regions. Gaviria and Raphael (2001) use contextual interaction variables of parental involvement with children, parental

substance abuse behavior, and parental education levels as instrumental variables for behaviors of children's reference groups to identify endogenous peer effects. We assess the robustness of traditional approaches adopted to identify endogenous peer effects by comparing our IV estimates to estimates that rely on variation in contextual variables as IVs for the school enrollment of the reference group. This exercise provides suggestive evidence that reliance on non-experimental estimates in this case would lead to erred conclusions regarding the magnitude of these interactions.

The paper is structured as follows. Section 2 provides a brief discussion of the Progresa program and its evaluation component, as well as the data used in the analysis. In Section 3, we present an empirical model of social interaction effects and discuss its identification problems. We then describe our research design, and how it avoids these identification pitfalls. The main estimates are reported in Section 4, followed by sensitivity tests of the identifying assumption in Section 5. Section 6 provides evidence on the potential estimation biases from using non-experimental variation in contextual variables as IVs, and Section 7 concludes.

2. Progresa Program, Evaluation, and Data

2.1 Background on the Progresa Program Evaluation

In 1997, the Mexican government initiated a large-scale education, health, and nutrition program (the Progresa Program) aimed at improving human development among children in rural Mexico. The program targets the poor in marginal communities, where 40 percent of the children from poor households drop-out of school after the primary level. The program provides cash transfers to the mothers of over 2.6 million children conditional on school attendance, health checks and health clinics participation, at an annual cost of approximately one billion dollars, or 0.2 percent of Mexico's GDP in 2000. The education component of Progresa consists of providing subsidies, ranging from $70 to $255 pesos per month (depending on the child's gender and grade level), to children attending school in grades three to nine of primary and lower secondary school. Overall, the program transfers are important, representing 10 percent of the average expenditures of beneficiary families in the sample.

A distinguishing characteristic of Progresa is that it included a program evaluation component from its inception. Progresa was implemented following an experimental design in a subset of 506 communities located across seven states: Guerrero, Hidalgo, Michoacán, Puebla, Querétaro, San Luis Potosí, and Veracruz. Among these communities, 320 were randomly assigned into a treatment group, with the remaining 186 communities serving as a control group, thus providing an opportunity to apply experimental design methods to measure its impact on various outcomes. In addition, within these selected communities, a poverty indicator was constructed using the household income data collected from the baseline survey in 1997. A discriminant analysis was then separately applied in each of the seven regions in order to identify the household characteristics that best classified poor and non-poor

households. This resulting welfare index then determined eligibility into the program (see Skoufias et al. 2001) for a more detailed description of the targeting process). While household eligibility was determined within all (treatment and comparison group) communities, only households classified as eligible and within the treatment villages became program beneficiaries during the evaluation period. That the eligibility classification exists for both treatment and control communities and treatment was randomly assigned are critical design aspects for the identification of the endogenous peer effects, as will be discussed in Section 3.

An issue in the implementation of the program involved an increase (by the program administrators) in the number of eligible households during the first year of the program, after it was discovered that households with certain characteristics – namely, the elderly poor who no longer lived with their children – were excluded from the initial eligibility criteria. Because of this oversight, a new discriminant analysis was conducted, and households were again classified as either poor or non-poor households. Households that were originally classified as non-poor but included in this second set of eligible households - called the 'densificado' group – became program beneficiaries approximately 8 months after the start of the program (Skoufias, Davis, and de la Vega 1999). As a result of this change in program implementation, there are eligible households above and below the initial region-specific eligibility thresholds. For our analysis we classify these 'densificado' households as eligible, since these are eligible for treatment at some point during the evaluation period.

2.2 Data and Measurement

Since the baseline census in October 1997, extensive biannual interviews were conducted during October 1998, May/June 1999, and November 1999, on approximately 24,000 households of the 506 communities.[2] Each survey is a *community-wide census* containing detailed information on household demographics, individual socio-economic status, health, and school behavior, and household income, expenditures and consumption. More specifically, the surveys in October 1997, October 1998, May/June 1999, and November 1999 collected information on the school enrollment and grade completed of each child in the household between 6 and 16 years old. We thus have information on enrollment during three consecutive school years (1997-98, 1998-99, and 1999-2000). Since primary school enrollment is almost universal in rural Mexico, we restrict our interest to the enrollment decisions of children who have attained at least a primary education but have not completed secondary school at baseline. Secondary school enrollment is the most problematic decision for school attainment[3], and also the grade levels where Progresa has had its greatest impact among eligible households (Schultz 2004). In our sample, this concerns approximately 2,738 children who are eligible to enter any of three lower secondary school

[2] There was a round of data collection in March of 1998 just prior to the start of the intervention.
[3] In 1997, primary school enrollment was close to 96.5%, compared to 65% enrollment into secondary school.

grade levels. By selecting the sample based on grade completed at baseline rather than including children who start completing their primary schooling during the post-treatment evaluation period, we avoid issues of dynamic selection into secondary school (Cameron and Heckman 1998). Also, with village-level censuses, we can reliably construct village-level means of household and individual characteristics - including school behavior and contextual variables - that may affect a child's schooling behavior.

Table 1 presents the mean of various individual and household-level characteristics for both eligible and non-eligible children between treatment and control villages. The first row in the table demonstrates the hurdle that secondary school represents for children in rural Mexico, and highlights a clear objective of the program (Table 1, Panel A). In 1997, the enrollment rate of eligible children in secondary school is 66 percent, on average. Although enrollment rates are on average 4 percentage points higher among ineligible children, only 70 percent of these were enrolled in secondary school. As one would expect from the random assignment, the pre-program difference in enrollment rates between treatment and control villages among both eligible and ineligible households is small and statistically insignificant. In addition, the simple difference in 1998 and 1999 enrollment rates between treatment and control communities provides a straightforward measure of the program's impact on school participation. In both years, enrollment rates in treatment villages were roughly 6 percentage points higher than in control villages among the beneficiary households. Table 1 also shows our first indication of a possible spillover effect. Although the difference is statistically insignificant, secondary school enrollment in the treatment villages is approximately 6 and 4 percentage points higher than in control villages among children of ineligible families in 1998 and 1999, respectively. Given these low enrollment rates, it is perhaps not too surprising that the mean educational level of heads of households are also quite low, as heads of eligible and ineligible households have only completed 2.6 and 3.2 years of schooling, respectively (Panel B). These children also tend to come from large households, as the mean number of household members in these villages is 7.3 for eligible households and 6.8 for ineligible ones.

We also compare mean attributes at baseline (October 1997) across treatment and control villages to evaluate the randomization of our sample (Table 1, columns 2-4, 6-8). As one would hope from the random assignment, there are no statistically significant differences in the observed characteristics of these individuals in most dimensions.[4]

In addition to the village-census data, we use administrative data on the amount of Progresa transfers received by the households per survey-round. As expected, the administrative transfers data shows that eligible households in treatment villages received 170 pesos per month (on average) during the April 1998-December 1999 period (Table 1, Panel B). Control households start receiving program

[4] Behrman and Todd (1998) conduct an exhaustive analysis of the degree of success of the random assignment of villages in the Progresa Program, and conclude that the randomization was successful.

transfers by December 1999, therefore average transfers for this subgroup are non-zero. The difference in transfers between the two groups is large and substantial. More importantly, the administrative data shows no evidence of program leakage, i.e., households classified as ineligible receiving cash transfers. Although this does not prove that leakage was not an issue in the program's implementation, there is no evidence of it at the central level.

Finally, we also make use of administrative data on secondary schools in the evaluation regions (which contain information on number of pupils by grade, teachers, number of classrooms, and other infrastructure characteristics of the schools). Without information on which school each child attends, we match – using GPS data – children from the same village to the closest secondary school. This administrative data allow us to rule out alternative hypotheses and to test our identifying assumptions (see discussion in Section 5). Means of characteristics of schools attended by the children in the sample are reported in Table 2, and as expected there are no systematic differences between treatment and control villages.

Given our panel data structure, an important issue in the empirical analysis is the extent of sample attrition. If being out-of-sample is correlated with the likelihood of being in the program (treatment) group, then this could lead to bias in the coefficient estimates. Sample attrition rates through the two post-treatment survey rounds are approximately 20 percent for the sample of children in secondary school, both in eligible and ineligible households (Table A1, columns 1 and 4), and the likelihood of attrition is highly correlated with individuals' observable characteristics (columns 2 and 5). Fortunately, attrition rates are balanced across treatment groups and the observables correlates of attrition are not significantly different across program and comparison groups (columns 3 and 6). We use baseline individual, household, and community characteristics as controls for any potential attrition bias in all our estimations.

3. Econometric Strategy

In this section, we discuss the econometric models used to estimate endogenous peer effects and the assumptions needed for identification We propose two alternative identification strategies: i) a structural IV estimate of the endogenous peer effects, which controls for potential unobserved heterogeneity and contextual interaction effects; and (ii) a basic estimate which allows for estimation of both the direct effect of the program and the endogenous peer effect using information on the differences in mean changes in subgroup school enrollment between treatment and control groups.

3.1 Identification of Endogenous Peer Effects

The standard approach used to estimate endogenous peer effects assumes that individuals' school enrollment decisions follow a simple linear model:

$$y_{ic} = \alpha + \beta X_{ic} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + u_{ic} \qquad (1)$$

where y_{ic} is an indicator variable for the school enrollment behavior of child *i* in village *c*; X_{ic} are exogenous characteristics of the individual; $\overline{X}_c$ are the mean exogenous characteristics of the reference group; Z_c are characteristics of the environment (e.g., village) that may influence individuals' school enrollment decisions; and $\overline{y}_c$ is the enrollment rate of the reference group.[5] This linear model provides a formal expression to three hypotheses often advanced to explain the common observation that individuals belonging to the same group tend to behave similarly. The first, *correlated effects*, proposes that individuals in the same group tend to behave similarly because they have similar characteristics or face similar environments; these are represented in the model by the vector of parameters β and λ. The second, *contextual peer effects*, proposes that exogenous characteristics of the reference group (e.g., parent involvement in children's education in the village) influence individual behavior; the vector of parameters γ captures these contextual effects. Finally, the hypothesis of *endogenous peer effects* proposes that the behavior of the group influences individual behavior. The parameter θ in the model captures this effect. As Manski (1993) shows, this linear-in-means framework cannot separately identify the two types of social interaction effects, but does determine whether some social effect is present.[6]

Identification of parameter θ is possible however if the outcome variable of some randomly chosen members of the group is exogenously altered. Formally, we can assume that individuals' school enrollment decisions follow model (1) augmented for the existence of an exogenous treatment T_{ic} which equals unity for a subset of individuals in the reference group *c* and zero otherwise. The individual characteristics of this subgroup are denoted by superscript *E*:

$$y_{ic}^{E} = \alpha + \beta X_{ic}^{E} + \gamma \overline{X}_c + \lambda Z_c + \theta \overline{y}_c + \delta T_{ic}^{E} + u_{ic}^{E} \qquad (1')$$

In addition, there are individuals within the same reference group *c* (denoted with superscript *NE*) who do not receive treatment:

$$y_{ic}^{NE} = \alpha + \beta X_{ic}^{NE} + \gamma \overline{X}_c + \lambda Z_c + \theta \overline{y}_c + u_{ic}^{NE} \qquad (1'')$$

[5] Note that in this specification we are assuming that the reference group and the environment are one in the same. This clearly need not be the case.

[6] To see this, take the expectation of equation (1) conditional on X and Z, integrating over Z, and solving for $\overline{y}_c$ results in the mean equilibrium outcome in group *c*, which, substituted in equation (1) yields the reduced form for individual outcomes: $y_{ic} = \frac{\alpha}{1-\theta} + \beta X_{ic} + \frac{\gamma + \beta\theta}{1-\theta}\overline{X}_c + \frac{\lambda}{1-\theta} Z_c + u_{ic}$. Manski (1993) shows that, conditional on $\theta \neq 1$, this equation has a unique solution, parameters γ and θ are unidentified, but composite parameters $\frac{\alpha}{1-\theta}$, $\frac{\gamma+\beta\theta}{1-\theta}$, and $\frac{\lambda}{1-\theta}$ are identified. Although the identification of the composite parameters does not allow one to distinguish between endogenous and contextual social interaction effects, it permits one to determine whether some social effect is present.

Using equations (1') and (1"), and recalling that group averages are related to within-village treated (*E*) and untreated (*NE*) group averages by:

$$\bar{y}_c = m_c^E \bar{y}_c^E + (1 - m_c^E)\bar{y}_c^{NE} \qquad (2)$$

$$\bar{X}_c = m_c^E \bar{X}_c^E + (1 - m_c^E)\bar{X}_c^{NE}$$

where m_c^E is the share of treated individuals in the reference group *c*, we can show, based on Cipollone and Rosolia (2003), that the mean equilibrium outcome in the reference group is:

$$\bar{y}_c = \frac{\alpha}{1-\theta} + \frac{\beta+\gamma}{1-\theta}\bar{X}_c + \frac{\lambda}{1-\theta}Z_c + \frac{\delta}{1-\theta}m_c T_c \qquad (3)$$

Therefore, one could use coefficient estimates from equations (1") and (3) to identify the direct treatment and endogenous peer effects parameters.

The specifications that we adopt in this paper are based on equation (1'') and a slight variant of equation (3):

$$y_{ic}^{NE} = \alpha + \beta X_{ic}^{NE} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + u_{ic}^{NE} \qquad (1")$$

$$\bar{y}_c = \tilde{\alpha} + \tilde{\beta}\bar{X}_c + \tilde{\lambda}Z_c + \tilde{\delta}T_c + \bar{\varepsilon}_c \qquad (3')$$

where T_c is the Progresa treatment village indicator variable and composite coefficients $\tilde{\alpha} = \frac{\alpha}{1-\theta}$, $\tilde{\beta} = \frac{\beta+\gamma}{1-\theta}$, $\tilde{\lambda} = \frac{\lambda}{1-\theta}$, and $\tilde{\delta} = \frac{\delta}{1-\theta}$. Note that equation (3') uses T_c rather than the interaction term $m_c T_c$ as the instrumental variable. We allow for this discrepancy in the model because the share of treated individuals in the reference group, m_c, (in this case the share of Progresa-eligible children in the village) may not be exogenous if there is any sorting of individuals into and out of the village based on unobservable characteristics of the households or villages.

Under the conditions of (i) robust partial correlation between the instrumental variable and the endogenous regressor $\left(\tilde{\delta} \neq 0\right)$, and (ii) lack of correlation between the excluded IV and the disturbance term in equation (1") $\left(E\left[T_c u_{ic}^{NE}\right] = 0\right)$, IV estimation is a consistent estimator of parameter θ. Condition (i) can be tested in the data, and results will be discussed in Section 4. Condition (ii), the exclusion restriction, is not directly testable and is a maintained assumption of the model; the random assignment of the program across villages is *not* sufficient to ensure that this condition holds.

The IV exclusion restriction relies on the assumption that an increase in school participation among ineligible children in treatment villages is the effect of the exogenous increase in school participation among the eligible secondary-school children within the village. Since it is possible however, that the program affected ineligible children along other channels, we follow two strategies to

provide evidence that this is not the case. First, using rich micro data for both eligible and ineligible households, we directly test whether other potential externalities from program impacts or particular intricacies of the program had an effect on ineligible households. We do not find any evidence of changes in the consumption patterns or health status of ineligible households, or in measures of school quality. Secondly, we condition on a large number of predetermined mean village-level contextual ($\overline{X}_c$) and environmental (Z_c) characteristics that may be correlated with the impacts of the intervention, and show that the effects are robust to these specifications. We do not find any evidence of alternative mechanisms, and defer discussion of these results to Section 5.

Finally, note that we are also assuming that endogenous peer effects are at the village-level. Although we lack information on the specific individuals who belong to a child's reference group, we believe that the assumption of village-level effects may not be problematic for the following reasons. As is common in village economies in less-developed countries, there is substantial ethnographic evidence documenting social interactions at the village level in rural communities in Mexico (e.g., Foster 1967). Furthermore, rural villages in this sample are quite small, with 47 households per village and only 20 children of secondary-school age per village, on average. Thus in the context of Mexico, village peer effects may be a more credible assumption than studies that use city blocks (Case and Katz 1991), census tracks (Topa 2001; O'Reagan and Quigley 1996), or schools (Evans, Oates and Schwab 1992; Hoxby 2000; Gaviria and Raphael 2001).

3.2 Nonparametric Identification of the Endogenous Peer Effect

An alternative strategy to identify endogenous peer effects is to consider a model where contextual interaction effects and observed and unobserved child and environmental characteristics do not affect a child's enrollment decision i.e. $\beta = \gamma = \lambda = 0$ in equations (1') and (1'').[7] Then, we can rearrange the two structural equations (1') and (1'') to get reduced-form equations for school participation of eligible and ineligible children as functions of treatment assignment:

$$\bar{y}_c^{NE} = \left(1 + \frac{\theta}{1-\theta}\right)\alpha^{NE} + \frac{\theta\delta m_c^E}{1-\theta} T_c \qquad = \pi_0^{NE} + \pi_1^{NE} T_c \qquad (4')$$

$$\bar{y}_c^{E} = \left(1 + \frac{\theta}{1-\theta}\right)\alpha^{E} + \left(\frac{\theta\delta m_c^E}{1-\theta} + \delta\right) T_c \quad = \pi_0^{E} + \pi_1^{E} T_c \qquad (4'')$$

Therefore, from the means comparisons of both eligibles and ineligibles in the treatment and control groups, and the mean proportion of eligible children in the villages ($\overline{m}^E$), we can recover the values of the structural parameters δ (the direct effect of the program on eligible children) and θ:

[7] Note that because the program's random assignment we assume such a model without concerns for omitted variable bias.

$$\delta = \pi_1^E - \pi_1^{NE}$$

$$\theta = \frac{\pi_1^{NE}}{(\pi_1^E - \pi_1^{NE})\overline{m}^E + \pi_1^{NE}},$$

and construct estimates of the covariance matrix of the structural parameters using the delta method. This method has the advantage that it permits simple calculations of the direct effect of the scholarships on eligible children and of the endogenous peer effects in these communities, imposing relatively strong exogeneity and functional form (i.e., constant additive effects) assumptions of the treatment and peer effects.

4. Estimates of Spillovers and Endogenous Social Interaction Effects

4.1 Estimates of Reduced-Form Spillover Effects

In this section, we present evidence on the reduced-form spillover effects of the program on school enrollment. We start the discussion with a graphical analysis to shed light on the patterns in the data. Figure 1 presents a series of graphs, based on nonparametric estimates, depicting enrollment rates in secondary school by the welfare index used to classify eligible and ineligible households.[8] Enrollment rates do not differ at baseline among program and comparison villages (Figure 1, Panels A and B). However, for 1998 and 1999, enrollment rates in program villages among both eligible and ineligible children increase substantially relative to the comparison group (Panels C and D). Among the ineligible group, we observe a striking difference in enrollment rates between treatment and control villages among relatively poorer households. This enrollment difference remains until a household welfare index of approximately 900 units (the median welfare index of ineligible households), at which point the enrollment rates tend to converge. This figure suggests that any spillovers of the program may have been concentrated among ineligible households with welfare characteristics relatively similar to the eligible households but classified above the welfare qualification.

Parametric linear probability estimates of the reduced-form relationship between program and comparison villages enrollment rates mirror the results depicted in the Figure 1. As documented by Schultz (2004), children in eligible households increased their school enrollment by 7.6 percentage points relative to eligible children in control villages (Table 3, Panel A, regression 1). The point estimate with household and village-level controls implies an effect of 8.3 percentage points, or 14 percent. Overall, children from ineligible households residing in the Progresa villages increased their secondary school enrollment rate by 2.8 percentage points relative to ineligible households in control villages (Table 3, Panel B, regression 2); however, the point estimate is only measured at 89 percent confidence. There are

[8] The conditional means are estimated by taking the mean enrollment within a bandwidth of 0.5. The figure is robust to perturbations to the bandwidth size.

significant differential effects on school enrollment by household's welfare index level (regressions 3 and 4). Among ineligible households with a below-median welfare index, Progresa increased secondary school enrollment by 5.5 percentage points (statistically significant at 90 percent confidence), but had no effect for children among the upper welfare-index group (-0.9 percentage points and not statistically significant).[9] We will postpone detailed discussion of the likely explanations for these differential subgroup effects to Section 4.2. Interestingly, we do not find evidence of differential spillover effects among boys and girls once we include household and village-level controls (regressions 5 and 6).

4.2 Estimates of Direct and Endogenous Peer Effects

We continue the analysis by reporting estimates of endogenous peer effects (θ) from OLS and IV estimation of equations (1") and (3') (Table 4). The OLS estimate of the overall endogenous effect, which does not take into account the problems of self-selection into reference groups, the reflection problem, and unobserved heterogeneity in the population, implies a 0.67 percentage point (1.1 percent; significant at 99 percent confidence) increase in a child's probability of enrollment as a result of a 1 percentage point increase in the reference group's enrollment rate (results not shown). The estimated correlation using control group villages, which excludes any potential use of experimental variation, implies an effect of 0.72 (standard error 0.076; significant at 99 percent confidence) (Table 4, regression 1). In contrast, IV estimates of the overall endogenous effect imply an effect of 0.65 without any control-adjustment (regression 2), but an effect of 0.49 (8 percent) increase in the probability of enrollment in secondary school once household and village-level controls are included in the model (regression 3). However, the latter estimate is not statistically significant at conventional confidence levels (significant at 89 percent confidence); we cannot reject that the OLS and 2SLS estimates are significantly different from each other. Although IV estimates suggest that the extent of endogenous peer effects may be quite large in this population, the comparison of OLS and IV point estimates suggests that these are smaller than simple correlations would imply. Note that we control for individual and household-level controls, village-level predetermined contextual variables (i.e., the proportion of secondary school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportions of heads of households) to correct for potential contextual characteristics which may be correlated with the excluded instrument.

Substantially larger peer effects are found for relatively poor children within the ineligibles group; a subgroup possibly more likely to be induced to continue their schooling, or more likely to have peers in the eligible children group. The point estimate on the effect for children in the below-median welfare-index group is 0.671 (regression 5). The correlation of own and social network enrollment rate

[9] The difference in effects is statistically significant at 90 percent confidence.

for this subgroup in control villages implies an effect of 0.75 (regression 4). Again, the experimental evidence suggests that the OLS estimates are biased upwards, although we cannot reject that the coefficients are equal. Note that we cannot identify the effect on children with a high household welfare index, since the first-stage correlation is weak for this subgroup (Panel B, regression 6). The average enrollment rate effect is small and indistinguishable from zero in these villages. Therefore, no inferences can be made on the peer effects for children in the wealthier households; the point estimate for this high welfare index group is -5.112 (not statistically significant). This differential effect by household wealth index is also consistent with relatively poorer households being more credit constrained, and therefore, more likely to remove their children from secondary school (Jacoby and Skoufias 1997). However, these differential effects may also exist due to variation in the social networks of poorer and wealthier ineligible children within villages. If the relatively poor ineligible children are more likely to interact with eligible children in the village, then the induced school participation of eligible children should have an effect particularly on this subgroup of children. Next, we discuss some evidence consistent with this hypothesis.

Although we do not have information on the exact peer network of each student, we can construct – for each child in the village – a measure of the number of extended-family members (i.e., cousins) who live in different households and are at risk of being in secondary school; this measure serves as a proxy for a child's number of family-related peers in the village (a potential subset of a child's peer group).[10] Comparing relatively poor (below median welfare-index) and relatively well-off (above median welfare-index) ineligible children, we find that the number of eligible extended-family links at baseline is significantly greater for ineligible children in the first group (0.97 children) relative to the latter (0.65 children) *among children with some extended-family link in the village*. This implies a difference of approximately 0.31 children (standard error, 0.09, significant at 99 percent confidence; not reported in the tables), or 48 percent, in the number of eligible links.[11] Although we do not expect all interactions to occur in these villages solely at the extended-family level, this *suggestive* evidence is consistent with poorer ineligible children tending to interact more with eligible children, and therefore, that program-induced increases in school participation are more likely to affect this subset of ineligible children.

Weak instruments are not a main concern in the estimation. There is a robust partial correlation between the program village treatment indicator and the potentially endogenous regressor, the village-

[10] We construct identifiers for extended-families in the villages by grouping children according to unique identifiers of their parents' last names. In Latin America, each individual has two last names, the first being the father's first last name and the second the mother's first last name. Therefore, we can construct the households where individuals are related (within reasonable errors) by using unique numerical identifiers of each combination of last names.

[11] Assuming that other children who are not matched to an extended-family network actually have no extended-family eligible links, (therefore, we can impute a zero number of extended-family links for all these children), we can construct measures for all ineligible children in the village. We also find a greater number of links for children in the below-median welfare index group (0.58 children) relative to other ineligible children (0.41 children); a difference of 0.16 children (standard error 0.06, significant at 99 percent confidence).

level enrollment rate. The F-test statistics of the significance of the IV in the first stage equations excluding and including controls are 8.74 and 7.60 in the overall effect model (Panel B, regressions 2 and 3), and the F-statistic for the poorer ineligible group is 13.92 (the first-stage coefficient is significant at 99 percent confidence) (regressions 5).[12]

We continue the analysis by reporting estimates of the direct and endogenous peer effects on school enrollment calculated from estimates of the mean differences in school participation of eligible and ineligible children (Table 5). The estimate of δ implies a direct effect of the program's incentives of 4.3 percentage points among secondary-school children (6.5 percent; significant at 90 percent confidence), and the point estimate of the endogenous interaction effect (θ) is 0.37 (insignificantly different from zero at conventional confidence levels) (Table 5, Panel B). The latter estimate implies that a 10 percentage point increase in the reference group's school enrollment increases the own-enrollment probability by 3.7 percentage points. These estimates suggest that approximately one third of the increase in school enrollment as a result of the program are driven by the endogenous peer effects; however, we do not put much confidence in these estimates due to the imprecision of the interaction effect estimate.

In summary, this evidence is consistent with the hypothesis that changes in reference groups' school enrollment behavior affects children's own enrollment behavior, and that these effects differ depending on children and their family's inherent opportunity costs, as well as by the types of peers they interact with. As will be shown in Section 5, these results are very robust to specifications, alternative measures of peer behavior, and identifying assumptions.

4.3 Estimates of Nonlinearities in Endogenous Peer Effects

The baseline linear-in-means model, which is the most popular in practice, is one in which peer effects have distributional but no efficiency consequences. However, many questions involving the efficiency-improving potential of peer effects requires a model in which peer effects are non-linear in peers' outcomes (e.g., detracking in schools) (Hoxby 2000). Therefore, we specify a more flexible relationship between peer effects and school enrollment to explore any potential non-linearities. We estimate a model originally proposed by Robinson (1988) which allows peer effects to be estimated non-parametrically while still controlling for the determinants of school enrollment, including contextual characteristics. Econometrically, we estimate the following equation:

$$y_{ic}^{NE} = \alpha + \beta X_{ic}^{NE} + \gamma \overline{X}_c + \lambda Z_c + g(\overline{y}_c) + u_{ic}^{NE} \qquad (5)$$

[12] The LIML estimates of equations (1") and (3'), which are robust to the weak instruments problem (under certain conditions, see Hayashi, 2000) give endogenous interaction effects very similar to the IV/GMM results reported in the text. Results are available upon request.

which is similar to equation (1"), except that $g(\bar{y}_c)$ is some unknown functional form. Due to the potential endogeneity of peer effects, we follow Blundell and Duncan (1998) in proposing a control function approach and modify equation (5) to include:

$$y_{ic}^{NE} = \alpha + \beta X_{ic}^{NE} + \gamma \bar{X}_c + \lambda Z_c + g(\bar{y}_c) + \rho \hat{\varepsilon}_c + u_{ic}^{NE} \quad (6)$$

where $\hat{\varepsilon}_c$ are estimate residuals from equation (3'). The key assumption underlying the estimation of this model is that conditioning on $\hat{\varepsilon}_c$ corrects for the potential endogeneity of peer effects.

Taking expectation and subtracting, we can rewrite equation (6) as:

$$y_{ic}^{NE} - E(y_{ic}^{NE} \mid \bar{y}_c) = \beta(X_{ic}^{NE} - E(X_{ic}^{NE} \mid \bar{y}_c)) + \gamma(\bar{X}_c - E(\bar{X}_c \mid \bar{y}_c)) + \lambda(Z_c - E(Z_c \mid \bar{y}_c)) + \rho(\hat{\varepsilon}_c - E(\hat{\varepsilon}_c \mid \bar{y}_c)) + u_{ic}^{NE} \quad (7)$$

which implies the following estimator for $g(\bar{y}_c)$:

$$\hat{g}(\bar{y}_c) = E(y_{ic}^{NE} \mid \bar{y}_c) - \hat{\beta}E(X_{ic}^{NE} \mid \bar{y}_c) - \hat{\gamma}E(\bar{X}_c \mid \bar{y}_c) - \hat{\lambda}E(Z_c \mid \bar{y}_c)) - \hat{\rho}E(\hat{\varepsilon}_c \mid \bar{y}_c) \quad (8)$$

With nonparametric estimates of $E(r_{ic} \mid \bar{y}_c), \forall r_{ic} \in \{y_{ic}^{NE}, X_{ic}^{NE}, \bar{X}_c, Z_c, \hat{\varepsilon}_c\}$, we can estimate equation (7) by OLS to get estimates of $\{\hat{\beta}, \hat{\gamma}, \hat{\lambda}, \hat{\rho}\}$ and subsequently estimate $\hat{g}(\bar{y}_c)$. As Robinson (1988) points out, $\hat{g}(\bar{y}_c)$ converges at $\sqrt{nh}$ where *h* is the bandwidth size.

Figure 2 presents the estimated endogenous peer effects by the enrollment rate of the reference group along with the 95 percent pointwise confidence intervals.[13] The shape of the graph exhibits a negative and concave relationship between the magnitude of the peer effect and the average enrollment of the reference group. Moreover, we can reject the null hypothesis of constant marginal effects, based on the estimated pointwise confidence bands. At an average enrollment of 20 percent, 10 percentage points increase in the reference group's enrollment rate will increase individual enrollment by 7.6 percentage points. In comparison, at an average enrollment of 80 percent, a 10 percentage point increase in the reference group's enrollment rate will increase the individual's enrollment by 5.5 percentage points. That the effect of one's peers decreases as the average enrollment of the group suggest important efficiency gains in mixing student's with high and low enrollment propensities, if enrollment propensities translate themselves into greater educational attainment or achievement.

5. Sensitivity Analyses and Tests of Identifying Assumptions

[13] A locally-weighted regression, with a 0.9 bandwidth was used to estimate the conditional expectations in equation 8. These estimates are robust to slight perturbations of the bandwidth size. The 95 percent confidences intervals were constructed by bootstrapping the sample 100 times.

It has been well documented that the impact of Progresa was not restricted to schooling. That the program may have affected ineligible children in ways other than an increase in the enrollment rates of their reference groups remains a potential concern for our identification strategy. Such a situation would invalidate our exclusion restrict and we would be mistakenly attributing the effects of this other mechanism to peer effects. In this section, we present a series of robustness checks and tests of our underlying counterfactual assumption to show that we are in fact providing unbiased estimates of endogenous peer effects.

5.1 Tests of Alternative Mechanisms

In order for the treatment village indicator to serve as a valid instrument, the program cannot have indirectly affected other determinants of an ineligible child's enrollment decisions. This is a substantive assumption in the case of Progresa, where the program's multidimensionality affected the livelihoods of *beneficiary* households through a series of mechanisms. Apart from the increases in secondary school enrollment rates among eligible children (Schultz 2004), researchers have found significant increases in household consumption levels, food consumption, and food quality (Hoddinott and Skoufias 2004), improvements in health status, and increases in health care utilization (Gertler 2004; Gertler and Boyce 2001).[14] If any of these program impacts lead to externalities for ineligible households, in the form of, for example, inter-household resource transfers, correlated positive shocks to income, or positive health externalities, and these in turn lead to increases in school enrollment rates for ineligible children, then we would be attributing the positive externalities from these other mechanisms to the endogenous peer effects.

Moreover, changes in environmental or institutional factors affecting children's school enrollment decisions present other potential concerns. A set of particularly important changes affecting school enrollment decisions were school supply-side interventions which accompanied the implementation of the program. Although this was done to mitigate potential congestion effects due to the expected increase in schooling demand, the improvement in schooling facilities may have attracted children from ineligible households.

To verify whether any of these factors play a role in explaining the school enrollment spillover effect, we test for the existence of any post-treatment differences in household consumption and expenditures, health status of children, and certain school characteristics which may have been affected by the program (Table 6). We do not find any evidence that monthly household expenditures increased in the two post-treatment periods among ineligible households in program relative to comparison villages

[14] There is also evidence that the program improved women's status within the household (see Adato et al., 2000 and Bobonis 2004 for a discussion of women's empowerment effects). Evidence of program impacts on other outcomes, including children and adults' labor supply (Parker and Skoufias, 1999), migration patterns (Angelucci 2003), ability to mitigate shocks (de Janvry et al. 2004), and inter-household transfers (Attanasio and Rios-Rull 1999) suggest relatively small changes in these margins.

(point estimate reported in Table 6 is -12.93, and not statistically significant). Since expenditures do no take into account consumption from household production, we also estimate household consumption in the first post-treatment period, and, again, find no significant difference in total consumption among these households (point estimate is -50.56, not statistically significant).[15] Moreover, differential estimates by welfare-index subgroups also results in insignificant differences in expenditures and consumption (rows 1-2, columns 2-5). These expenditure and consumption patterns, as well as the evidence from the transfers data, provide evidence inconsistent with the possibility of inter-household income transfers from beneficiary to non-beneficiary households, correlated positive income shocks at the village-level, or evidence from program leakage (where some ineligible households may have been able to receive program transfers).

Households may be substituting expenditures in different areas as a result of the children's school enrollment. Consistent with the evidence on increased school participation, estimates suggest an increase in the share of the household budget spent on educational expenses (e.g., school supplies, school contributions). The point estimate implies an increase of 0.5 percentage points (9 percent, or approximately 5 pesos) on educational expenditures among all ineligible households and 0.4 percentage points (9 percent, approximately 4 pesos) among poorer ineligible households (row 3, columns 1 and 2). However, none of the estimates are significant at conventional confidence levels.

It is also possible that the liquidity injection from the program may have relaxed lending constraints of eligible households, enabling ineligible households to borrow when hit by negative idiosyncratic shocks, and making them less likely to remove their children from secondary school in the event of a shock (Jacoby and Skoufias 1997; Angelucci and De Giorgi 2005). We examine this potential alternative channel by showing evidence of the expenditure responses of ineligible households to natural shocks in both program and comparison villages in our subsample (Table 6, Panel B).[16] If the liquidity constraint hypothesis were correct, we would expect a relative positive effect on expenditures and school enrollment among households who suffer a shock in program villages. A potential concern to this test is that natural shock measures may not be very reliable: ineligible households seem to increase household expenditures in response to natural shocks, and we observe a similar pattern using household consumption data (Panel B, row 2). Given this caveat, we do not find evidence that ineligible households in program villages who suffer natural shocks have higher expenditure levels than those in comparison

[15] We use household expenditures and consumption as proxies for household income, since income is usually measured with substantial error in agricultural households, and these may better represent permanent incomes of households. Unfortunately, we only have home production data for the October 1998 survey round, and therefore, cannot estimate the consumption models in the second post-treatment round.

[16] We use household survey data to construct the shock measure, following Angelucci and De Giorgi (2005). The survey recorded whether the household has been hit by any of the following natural disasters in the six months preceding the interview: drought, flood, hail, fire, plague, earthquake, and hurricane. We create a variable which indicates whether the household has been hit by any natural disaster.

villages (Panel B, row 3). Furthermore, the school enrollment effect is lower among 'shock' than among 'no-shock' households: the estimated reduced form effects are -0.057 (standard error 0.032, significant at 10 percent confidence) and 0.083 (standard error 0.029, significant at 99 percent confidence), respectively (not reported in the tables).[17] We conclude that a potential relaxation of liquidity constraints did not affect the subsample of ineligible households with secondary-school children.

Regarding potential health externalities as a result of reduced contagion of communicable diseases (à la Miguel and Kremer 2004), or improved health status as a result of potential improvements in access to health facilities, we do not find evidence that any of these mechanisms took place, or at least that it led to significant improvements in the health status of secondary school-aged children. Unfortunately, the survey collected data from different questions across rounds regarding the self-reported health status of children. Therefore, we show evidence from the first post-treatment round (October 1998) on the number of days the child was ill in the past four weeks, and on answers to questions of difficulty with activities of daily living in the last survey round (November 1999).[18] There is no overall significant reduction or increase in the number of days ill reported among ineligible children in October 1998 (the point estimate is 0.10, not statistically significant; row 5, column 1). Differential effects by welfare subgroups suggest no difference in the morbidity of relatively poorer and wealthier households (Panel A, row 5, columns 2-3). Similar results are found using the ADL measures in November 1999 (rows 6-8). In summary, we find evidence inconsistent with any positive health effect hypotheses.

The evidence from administrative data on school characteristics is also inconsistent with supply-side interventions potentially affecting the school enrollment decisions of these children. The number of teachers in secondary schools did not significantly increase in program schools relative to comparison schools (Panel A, row 9, column 1). The differential effects across schools attended by different subgroups of children are not significantly different from zero either (row 9, columns 2-3). Interestingly, there is an (insignificant) increase in the mean pupil-teacher ratio in program schools (the point estimate is 1.19). Moreover, the secondary schools attended by the poorer ineligible children suffered a (marginally significant) increase in pupil teacher ratios of 1.78, as expected from the increased school enrollment among eligible and ineligible children from these villages. These estimated increases are within the expected range from the household-survey estimates of increases in school enrollment. A back-of-the-envelope calculation implies an expected increase of 1.42 in the pupil teacher ratio.[19] If any

[17] School enrollment effects are similar among relatively poor ineligible children: estimates are -0.029 (standard error 0.038) and 0.101 (standard error 0.036) for 'shock' and 'no-shock' household children, respectively.

[18] See Gertler and Boyce (2001) for a detailed discussion of this self-reported data in the Progresa evaluation surveys, and a thorough analysis of the heath impacts on eligible households.

[19] Approximately 76 percent of children in the villages were eligible, and approximately 59 percent of the ineligibles belonged to the below-median welfare-index group. In addition, there are approximately 20 children of secondary-school age per village.

negative congestion effect took place, we would expect a reduction in school enrollment among ineligible children (in equilibrium).

Secondly, both teachers and school directors, in separate focus groups, voiced that the improvements in educational outcomes resulted from improved student interest and attendance rather than improvements in school inputs. Skoufias and McClafferty (2001) reports from a series of interviews with school teachers that "[t]he general perception was that [the] supply-side [intervention] was not sufficient to deal with the increase demand, although better attendance and attitudes towards schooling made teaching easier and more rewarding." Unfortunately, we do not have quantitative data to test whether there were systematic improvements in teacher motivation, a factor that may have affected school enrollment decisions of all children.

A final concern may be that instability in the implementation of the program during its first year. Since some ineligible households were phased into the program during the first and second years (the 'densificado' households) this could have lead to uncertainty about the potential future eligibility of other non-beneficiary households. In addition, a large proportion of eligible households (27% of the total eligible population and mostly 'densificado' households) never received program payments during the evaluation period.[20] To the extent that this mismanagement led to uncertainty and changes in expectations about future eligibility, ineligible households could have increased their children's school participation in order to maximize their opportunity of becoming beneficiaries (although it is theoretically possible that they would have reduced their children's school participation as well).

Although expectations of program eligibility are unfortunately unobserved, rending this hypothesis untestable, we do provide some indirect evidence to address this issue. If the extent of uncertainty surrounding the implementation of the program was more prevalent in villages where the incorporation of 'densificado' households was higher, we should expect higher increases in the school participation of ineligible children in these specific villages. However, when we estimate a schooling decision reduced-form model with an interaction term of the Progresa treatment indicator and the proportion of 'densificado' households in the village, we find that the interaction term is not significantly different from zero (not reported in the tables).[21]

5.2 Robustness Checks to Contextual Interaction Effects

In addition to these reduced-form tests, we report estimates of the endogenous peer effect conditioning on a series of expenditure-related village contextual controls (in addition to the pre-

Using the estimate of program impacts among eligible children of 8.3 percentage points increase in the secondary school enrollment rate, and the 5.5 percentage point increase among the below-median welfare-index group of ineligibles, we can estimate the mean increase in the number of pupils as 20*[(0.76)*0.083 + (0.14)*0.055] = 1.42 pupils.

[20] Previous researchers of the program suspect that these households were never formally incorporated into the program (Hoddinott and Skoufias, 2004).

[21] Estimates available from the authors upon request.

determined contextual controls): mean village-level household expenditures, mean educational, food, boys and girls' clothing, alcohol & tobacco expenditure shares, and an indicator variable for whether the village suffered a rainfall shock (i.e., flood) in the past six months.[22] Panel A in Table 7 reports estimates of θ from a series of regressions which gradually condition on village-level predetermined and expenditure-related contextual variables, and also compares these to OLS estimates of θ. Conditioning on these sets of contextual variables reduces the point estimate of the overall effect slightly, from 0.54 to 0.49, and the latter becomes insignificantly different from zero (Table 7, Panel A, column 1). However, the point estimates for the below-median welfare-index group does not vary significantly with the inclusion of additional controls (Panel A, column 2). Also note that the F-statistics of the first-stage regression coefficients (reported in brackets) do not vary substantially once we condition on potential exogenous interaction factors. The exercise suggests that, especially among the specific subgroup with significant endogenous effects, the estimates are robust to these potential contextual effects.

Finally, we use an alternative measure of the reference group school enrollment behavior to assess whether the results are robust to the functional form assumptions used. We estimate both OLS and 2SLS models using as the endogenous measure the number of children in the village enrolled in secondary school, conditioning on the total number of children in the village, and use the treatment village indicator as an IV for the former in the 2SLS specifications (Table 7, Panel B). Consistent with the previously reported estimates, the 2SLS results imply slightly larger effects. The main estimate implies that an increase by one child in attending secondary school in the reference group increases a child's probability of enrollment by 1.6 percentage points (2.6 percent, significant at 95 percent confidence). The estimates for relatively poor children and for 10-13 year old children imply an increase in probability of 1.9 percentage points (2.4 percent; significant at 95 percent confidence).

6. Comparison with Contextual Variables IV Estimates

Previous observational studies that estimate endogenous peer effects use contextual variables of the reference group as instrumental variables for the reference group's behavior (Case and Katz 1991; Evans, Oates, and Schwab 1995; Gaviria and Raphael 2001). Estimates using this identification strategy are consistent only if the mechanism through which these particular contextual variables of the reference group affect own behavior is through affecting the reference group's behavior. However, if this IV exclusion restriction is not satisfied, the non-experimental estimates would be inconsistent.

In this section, we compare our IV estimates using experimental variation in the school enrollment of the reference group to traditional peer effects estimates, which use reference groups'

[22] We do not control for mean village-level measures of health status in the population of interest, because the survey questions assessing health status vary across survey rounds. However, as seen in Section 5.1, there is no evidence of positive health externalities among this group of children (see discussion above).

contextual variables as IVs for the school enrollment of the reference group (Table 8). We restrict our sample to children in the control villages, and use the mean educational achievement of heads of households in the village and the proportion of female household heads in the village as instrumental variables for the reference group enrollment rate, since we do not have detailed information on parent's attitudes towards schooling and other factor that may influence school drop-out. We report estimation results from regressions analogous to those presented in Table 4. Note that the Hansen J specification tests do not reject the validity of these alternative instruments.

The non-experimental overall effect estimate is 0.577 (significant at 99 percent confidence) (Table 8, regression 1). The point estimate is slightly larger in magnitude than the experimental point estimate, but the difference is not statistically significant. However, the subgroup results by household welfare would imply that effects are *not* concentrated among the subgroups identified by the experimental estimates. These would imply that the effects are concentrated among wealthier households in the villages; the non-experimental estimate for children in the first welfare group is 0.449 (not statistically significant; regression 2), whereas the estimate for children in the upper welfare index group is 0.743 (significant at 95 percent confidence; regression 3). These results suggest that endogenous interaction effects are convex on welfare among ineligible children in the villages. However, note that we cannot reject that the effects by subgroup are significantly different from each other. This exercise therefore provides suggestive evidence that reliance on non-experimental estimates in this case would lead to different conclusions regarding the magnitude of these interactions.

An alternative hypothesis that may explain the contextual instrumental variables estimates is that, even if the instrument satisfies the exclusion restriction, this variation may be inducing school participation among a different subset of children in the village, not just the Progresa-eligible children. Therefore, if children have different subsets of peer groups, even within these small villages, there may exist an induced school participation effect among different subgroups of ineligible households. Since we do not have information on the exact peer groups of each individual, we cannot reject this alternative hypothesis. This exercise provides evidence of Manski (1993)'s point that researchers should optimally have perfect measures of peer groups, since – even in this small village context – children may be selecting into smaller peer group networks. This exercise therefore provides suggestive evidence that reliance on non-experimental estimates in this case would lead to different conclusions regarding the magnitude of these interactions.

7. Conclusions

In 1997, the Mexican government introduced a randomly phased-in human development program designed to increase human capital among the rural poor. This study uses experimental variation in the school enrollment rates among program eligible households to estimate how endogenous social

interactions affect the school decisions of children ineligible to receive these program benefits. Our findings suggest that the enrollment behavior of one's peers has an important role on a child's decision to enroll into school. A 10 percentage point increase in the enrollment rate of a child's reference group, increases his likelihood of attending secondary school approximately 5 percentage points. These endogenous peer effects are more pronounced among children of relatively poorer households within the ineligibles group and among children who's reference groups have a low average enrollment rate, which provides evidence that peer effects operate ina nonlinear fashion. Furthermore, we are able to reject other potential contextual interaction effects hypotheses using rich micro data on household consumption and expenditures, health of individual members, and administrative data on program transfers and school characteristics. This sensitivity analyses confirms the validity of the identifying assumptions of the empirical social interactions model.

These results have important implications for education policy in rural areas in Mexico. Behavioral peer effects seem to have substantial effects in these communities; policies which would promote positive sorting of students by ability may have negative social multiplier effects by excluding interactions of different types of children, as found for Chile in Hsieh and Urquiola (2004). These interaction effects may have important implications for inhibiting or promoting child labor in these poor areas, and future work will be assessing this aspect of the social interaction effects.

Theoretical models within economics incorporate endogenous peer effects as a result of identity formation behavior (Akerlof and Kranton 2002), conformity behavior (Bernheim 1994), informational externalities (Bikhchandani, Hirschleifer, & Welch 1992), and preferences depending on group behavior (Becker 1996), among others. Future research should work in trying to empirically differentiate the specific mechanisms for which we observe these 'reduced-form' endogenous interactions. Current work attempting to distinguish these effects, such as Akerlof and Kranton (2002), Miguel and Kremer (2003), Munshi and Myaux (2002), and Hoxby and Weingarth (2005) could serve researchers as guides for these types of studies.

7. References

Adato, M., B. de la Brière, D. Mindek, & A. Quisumbing (2000). "Final Report: The Impact of PROGRESA on Women's Status and Intrahousehold Relations". July. International Food Policy Research Institute, Washington, D.C.

Akerlof, George and Rachel Kranton (2002). "Identity and Schooling: Some Lessons for the Economics of Education", *Journal of Economic Literature*, 40(4), 1167-1201.

Angelucci, Manuela (2004). "Aid and Migration: An Analysis of the Impact of Progresa on the timing and size of labour migrations", IZA Working Paper DP1187.

Angelucci, Manuela and Giacomo De Giorgi (2005). "Indirect Effects of an Aid Program: the Case of Progresa on Consumption", unpublished manuscript, University of Arizona.

Attanasio, Orazio and José Víctor Ríos-Rull (2000). "Consumption Smoothing in Island Economies: Can Public Insurance Reduce Welfare?", *European Economic Review*, 44(7), 1225-1258.

Bayer, Patrick, Fernando Vendramel Ferreira, and Robert McMillan (2004). "Tiebout Sorting, Social Multipliers and the Demand for School Quality", unpublished manuscript, University of Toronto.

Becker, Gary (1996). Accounting for Tastes. Cambridge, MA: Harvard University Press.

Behrman, Jere R., Piyali Sengupta, and Petra E. Todd (2001). "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment". *Economic Development and Cultural Change* (forthcoming).

Behrman, Jere R. and Petra E. Todd (1999). "Randomness in the Experimental Samples of PROGRESA (Education, Health, and Nutrition Program)". International Food Policy Research Institute.

Bernheim, B. (1994). "A Theory of Conformity", *Journal of Political Economy*, 102, 841-877.

Bikhchandani, S., D. Hirschleifer, & I. Welch (1992). "A Theory of Fads, Fashion, Custom, Cultural Changes as Informational Cascades", *Journal of Political Economy*, 100, 992-1026.

Blundell, Richard and Alan Duncan (1998). "Kernel Regression in Empirical Microeconomics", *Journal of Human Resources*, 33(1), pp. 62-87.

Bobonis, Gustavo J. (2004). "Income Transfers, Divorce, and Intra-Household Resource Allocation: Evidence from Rural Mexico". unpublished manuscript, University of Toronto.

Cameron, Stephen and James Heckman (1998). "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, 106(2), 262-333.

Case, Anne C. and Lawrence F. Katz (1991). "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youths." National Bureau of Economic Research Working Paper No. 3705.

Cipollone, Piero and Alfonso Rosolia (2003). "Endogenous social interactions in schooling: evidence from an earthquake", unpublished manuscript, Bank of Italy Research Department.

Ding, Weili and Steven F. Lehrer (2005). "Do Peers Affect Student Achievement in China's Secondary Schools?", unpublished manuscript, Queen's University.

Epple, Dennis and Richard F. Romano (1998). "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects", *American Economic Review*, 88(1), 33-62.

Gaviria, Alejandro and Steven Raphael (2001). "School-based Peer Effects and Juvenile Behavior", *Review of Economics and Statistics*, 83(2) 257-268.

Gertler, Paul (2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment", *American Economic Review*, 94(2), 336-341.

Gertler, Paul and Simone Boyce (2001). "An Experiment in Incentive-Based Welfare: The Impact of Progresa on Health in Mexico". unpublished manuscript, University of California-Berkeley.

Hayashi, Fumio (2000). Econometrics. Princeton, NY: Princeton University Press.

Hoddinott, John and Emmanuel Skoufias (2004). "The Impact of PROGRESA on Food Consumption", *Economic Development and Cultural Change,* 53(1), 37-61.

Hoxby, Caroline M. (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation", National Bureau of Economic Research Working Paper No. 7867.

Hoxby, Caroline M. and Gretchen Weingarth (2005). "Taking Race Out of the Question: School Reassignment and the Structure of Peer Effects", unpublished manuscript, Harvard University.

Hsieh, Chang-Tai and Miguel Urquiola (2004). "When Schools Compete, how do they compete?", unpublished manuscript, Columbia University.

Jacoby, Hanan G. and Emmanuel Skoufias (1997). "Risk, Financial Markets, and Human Capital in a Developing Country", *Review of Economic Studies*, 64(3), pp. 311-335.

Kling, Jeffrey, Jens Ludwig, and Lawrence F. Katz (2005). "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment", *Quarterly Journal of Economics,* 120(1), 87-130.

Kremer, Michael, Edward Miguel, and Rebecca Thornton (2005). "Incentives to Learn", unpublished manuscript, UC Berkeley.

Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60(3). 531-542.

Miguel, Edward and Michael Kremer (2003). "Networks, Social Learning, and Technology Adoption: The Case of Deworming Drugs in Kenya", unpublished manuscript, University of California-Berkeley.

Miguel, Edward and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities", *Econometrica*, 72(1), 159-217.

Munshi, Kaivan and Jacques Myaux (2002). "Development as a Process of Social Change: An Application to the Fertility Transition", unpublished manuscript, Brown University.

O'Reagan, K. and J. Quigley (1996). "Spatial Effects upon Employment Outcomes: The Case of New Jersey Teenagers", *New England Economic Review*: Federal Reserve Bank of Boston. 41-57.

Parker, Susan and Emmanuel Skoufias (2000). "Final Report: The Impact of PROGRESA on Work, Leisure, and Time Allocation". October. International Food Policy Research Institute, Washington, D.C.

Robinson, P.M. (1988). "Root-N-consistent semiparametric regression", *Econometrica*, 56(4), pp. 931-954.

Sacerdote, Bruce (2001). "Peer Effects with Random Assignment: Results for Dartmouth Roommates" *Quarterly Journal of Economics*. 116(2). 681-704.

Schultz, T. Paul (2004). "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program", *Journal of Development Economics*, 74, 199-250.

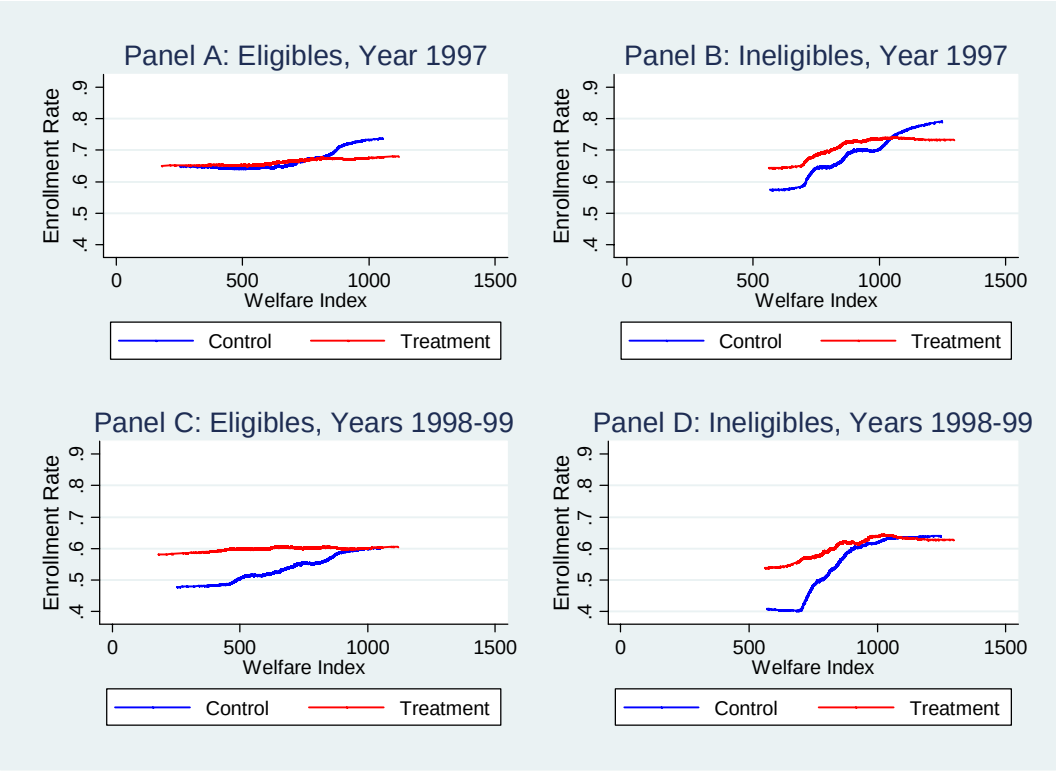
Skoufias, Emmanuel, Benjamin Davis, and Sergio de la Vega (2001). "Targeting the Poor in Mexico: An Evaluation of the Selection of Households into PROGRESA." *World Development*. 29(10). 1769-1784.

Skoufias, Emmanuel, Benjamin Davis, and Sergio de la Vega (1999). "An Addendum to the Final Report: An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico dated June 4, 1999", International Food Policy Research Institute, Washington, D.C.

Skoufias, Emmanuel and Bonnie McClafferty (2001). "Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI", FCND Discussion Paper 118.

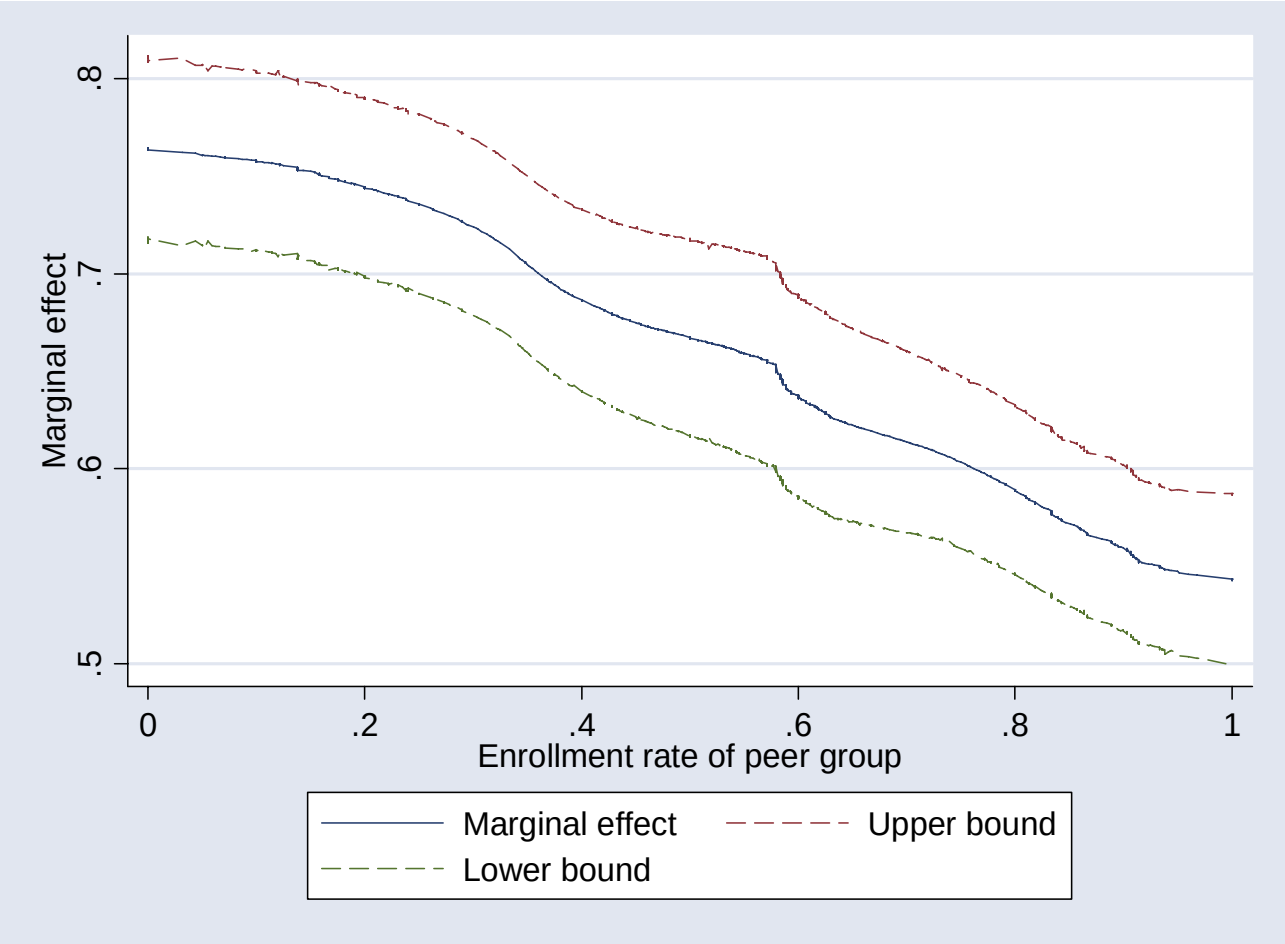
Topa, Giorgio. (2001). "Social Interactions, Local Spillovers and Unemployment", *The Review of Economic Studies.* 68(2) 261-296.

Figure 1: Nonparametric Estimates of Enrollment Rates by Household Eligibility Index, Years 1997-1999



Notes to Figure 1: Locally weighted smoothing of the proportion of individuals enrolled in secondary school by the welfare index of program eligibility; bandwidth = 0.8. The numbers of ineligible and eligible children are 2,738 and 11,147, respectively. Vertical lines drawn at welfare index levels 550 and 822.

Figure 2: Semi-parametric estimates of endogenous peer effects
by the enrollment rate of the peer group



Notes to Figure 2: Estimates of marginal effects of village enrollment rate on enrollment propensity of individual from semi-parametric Fan locally-weighted regressions; bandwidth = 0.9. The 95 percent pointwise confidence intervals were constructed by bootstrapping the sample 100 times.

Table 1: Individual and Household Characteristics across Program and Comparison Villages

	----------Ineligible Households----------				----------Eligible Households----------			
	Mean [Std.Dev.]	Program	Comparison	Difference	Mean [Std.Dev.]	Program	Comparison	Difference
Panel A: Child Characteristics								
School Enrollment in 1997	0.699	0.712	0.680	0.032	0.663	0.664	0.662	0.002
	[0.459]			(0.029)	[0.473]			(0.020)
School Enrollment in 1998	0.655	0.679	0.618	0.061*	0.635	0.661	0.592	0.069***
	[0.475]			(0.033)	[0.481]			(0.024)
School Enrollment in 1999	0.515	0.532	0.489	0.042	0.516	0.540	0.479	0.061***
	[0.500]			(0.034)	[0.500]			(0.023)
Child's age in 1997	13.43	13.41	13.46	-0.05	13.36	13.36	13.35	0.02
	[1.72]			(0.07)	[1.67]			(0.04)
Grade completed in 1997	6.25	6.27	6.23	0.05	6.03	6.03	6.04	-0.01
	[1.01]			(0.05)	[0.93]			(0.03)
Gender (Boy)	0.495	0.497	0.494	0.003	0.504	0.511	0.492	0.019*
	[0.500]			(0.020)	[0.500]			(0.010)
Indigenous	0.115	0.129	0.093	0.036	0.306	0.305	0.308	-0.003
	[0.319]			(0.040)	[0.461]			(0.052)
Panel B: Household Characteristics								
Head of Household's Schooling	3.19	3.25	3.10	0.15	2.57	2.58	2.57	0.01
	[2.97]			(0.20)	[2.39]			(0.11)
Head of Household's Gender (Male)	0.926	0.932	0.918	0.014	0.921	0.921	0.922	-0.001
	[0.261]			(0.013)	[0.269]			(0.007)
Head of Household's Age	48.78	48.82	48.73	0.08	45.88	45.62	46.30	-0.68**
	[10.65]			(0.62)	[10.84]			(0.33)
Household size	6.85	6.78	6.97	-0.19	7.34	7.33	7.38	-0.05
	[2.32]			(0.17)	[2.36]			(0.09)
Total Household-level Progresa Transfers (Post-treatment)	-	-	-	-	111.48	170.27	14.93	155.34***
					[131.44]			(5.84)

Notes to Table 1: Standard deviations of variables reported in brackets. Differences estimated in OLS regression models. Robust standard errors in parentheses; disturbances are allowed to be correlated within village; significantly different from zero at (*) 90%, (**) 95%, and (***) 99% confidence. The numbers of ineligible and eligible children are 2,738 and 11,147, respectively.

Table 2: School Characteristics across Program and Comparison Villages

	All Villages		
	Program	Comparison	Difference
Tele-secondary school	0.85	0.88	-0.03
			(0.03)
General sec. school	0.05	0.04	0.01
			(0.02)
Technical sec. school	0.09	0.07	0.02
			(0.02)
Rural	0.93	0.95	-0.02
			(0.02)
Semi-urban	0.06	0.04	0.02
			(0.02)
Classrooms in Grade 7	1.14	1.15	-0.01
			(0.08)
Classrooms in Grade 8	1.05	1.02	0.03
			(0.08)
Classrooms in Grade 9	0.98	0.94	0.04
			(0.08)
Number of teachers	3.10	3.06	0.04
			(0.37)
Pupil teacher ratio	22.14	21.59	0.55
			(0.92)

Notes to Table 2: Differences estimated in OLS regression models. Robust standard errors in parentheses; disturbances are allowed to be correlated within village; significantly different from zero at (*) 90%, (**) 95%, and (***) 99% confidence. The number of secondary schools is 506.

Table 3: Estimates of Reduced-Form Spillover Effects among Ineligible Children

Panel A: No Controls

	Dependent variable: School enrollment indicator					
Sample	All Eligible Children	All Ineligible Children	Welfare < Median	Welfare > Median	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Treatment indicator	0.076***	0.050	0.085**	-0.043	0.041	0.059*
	(0.023)	(0.031)	(0.035)	(0.036)	(0.038)	(0.036)
Individual & household controls	No	No	No	No	No	No
Village contextual controls	No	No	No	No	No	No
State indicators	No	No	No	No	No	No
Mean of dependent variable	0.579	0.587	0.539	0.641	0.601	0.574
Observations	13124	4211	2757	1454	2119	2092
R-squared	0.006	0.003	0.007	0.000	0.002	0.004

Panel B: Including Controls

	Dependent variable: School enrollment indicator					
Treatment indicator	0.083***	0.028	0.055*	-0.009	0.033	0.027
	(0.018)	(0.025)	(0.029)	(0.044)	(0.030)	(0.031)
Individual & household controls	Yes	Yes	Yes	Yes	Yes	Yes
Village contextual controls	Yes	Yes	Yes	Yes	Yes	Yes
State indicators	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	0.579	0.587	0.539	0.641	0.601	0.574
Observations	13124	4211	2136	1454	2119	2092
R-squared	0.091	0.12	0.12	0.13	0.13	0.12

Notes to Table 3: Coefficient estimates from OLS regressions are reported. Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages; significantly different from zero at (*) 90%, (**) 95%, (***) 99% confidence. Individual and HH-level controls are the child's gender, indigenous status, the household's welfare index, education, age, and gender of the head of household , family size, and distance to secondary school. Village contextual controls are the proportion of secondary-school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportion of heads of households.

Table 4: OLS and IV Estimates of Endogenous Peer Effects among Ineligible Children

Panel A: IV Estimates

	Dependent variable: School enrollment indicator					
Sample	All Children in Control Group	All Children	All Children	Welfare<Median Control Group	Welfare < Median	Welfare > Median
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IV	OLS	IV	IV
Social Network Enrollment Rate	0.716***	0.649***	0.492	0.750***	0.671***	-5.112
	(0.076)	(0.239)	(0.310)	(0.092)	(0.246)	(18.120)
Individual & household controls	Yes	No	Yes	Yes	Yes	Yes
Village contextual controls	Yes	No	Yes	Yes	Yes	Yes
State indicators	Yes	No	Yes	Yes	Yes	Yes
First-stage F-statistic	-	8.74	7.60	-	13.92	0.94
[p-value]		[0.0033]	[0.0061]		[0.0002]	[0.3320]
Mean of dependent variable	0.557	0.587	0.587	0.507	0.559	0.642
Observations	1678	4211	4211	1075	2757	1454
R-squared	0.19	0.13	0.17	0.20	0.17	0.04

Panel B: First Stage Regressions

	Dependent variable: Social Network Enrollment Rate				
Treatment indicator	0.077***	0.057***		0.082***	0.025
	(0.026)	(0.021)		(0.022)	(0.026)
Individual & household controls	No	Yes		Yes	Yes
Village contextual controls	No	Yes		Yes	Yes
State indicators	No	Yes		Yes	Yes
Observations	4211	4211		2757	1454

Notes to Table 4: Coefficient estimates from OLS and IV regressions are reported. Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages; significantly different from zero at (*) 90%, (**) 95%, (***) 99% confidence. Individual and HH-level controls are the child's gender, indigenous status, the household's welfare index, education, age, and gender of the head of household, family size, and distance to secondary school. Village contextual controls are the proportion of secondary-school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportion of heads of households.

Table 5: Estimates of Direct Program and Endogenous Peer Effects

Panel A: Mean Differences in School Enrollment Status		(1)	**Panel B:** Structural Parameters Estimates		(2)
Eligible children	π_1^E	0.064***	Direct Program Effect	δ	0.043*
		(0.013)			(0.02)
Ineligible children	π_1^{NE}	0.021		θ	0.37
		(0.020)	Endogenous Peer Effect		(0.39)
Mean proportion of eligible children in village	$\overline{m}^E$	0.846			

Notes to Table 5: Mean differences in school enrollment status estimates from difference-in-differences OLS regressions of school enrollment status; disturbance terms are allowed to be correlated within villages but not across villages. Structural parameter estimates computed from OLS coefficient estimates of reduced-form differences in Panel A; delta method estimates of standard errors.

Table 6: Tests of Alternative Mechanisms for Spillover Effect

Panel A: Alternative Channels, Reduced Form

	Coefficient Estimate on Treatment Village Indicator (s.e.)			
Sample	All children	Welfare < Median	Welfare > Median	Mean of dep. variable
	(1)	(2)	(3)	(4)
Dependent variables	OLS	OLS	OLS	
Total HH Expenditures	-12.93	19.00	-44.56	1029.0
	(38.22)	(44.38)	(56.62)	
Total HH Consumption, October 1998	-50.56	-7.76	-98.56	1201.8
	(58.10)	(67.10)	(85.36)	
Schooling Expenditure Share	0.005	0.004	0.007	0.052
	(0.005)	(0.005)	(0.008)	
Food Expenditure Share	0.010	0.013	0.003	0.654
	(0.009)	(0.010)	(0.011)	
Days ill, October 1998	0.10	-0.01	0.25	0.357
	(0.11)	(0.14)	(0.19)	
Days of Difficulty with Daily Activities due to Illness, Nov. 1999	-0.047	-0.127	0.025	0.178
	(0.061)	(0.093)	(0.073)	
Days of No Daily Activities due to Illness, November 1999	0.029	-0.029	0.100	0.093
	(0.046)	(0.071)	(0.080)	
Days in Bed due to Illness, November 1999	0.030	-0.031	0.106	0.046
	(0.044)	(0.067)	(0.079)	
Number of teachers	0.05	0.06	-0.01	3.45
	(0.26)	(0.29)	(0.34)	
Pupil teacher ratio	1.19	1.78*	0.12	21.96
	(1.01)	(1.04)	(1.32)	

Panel B: Treatment and Shocks Interactions in HH Expenditures Regressions

	Dependent variable: Total HH Expenditures			
Treatment indicator	2.07	23.55	-15.19	1029.0
	(44.18)	(45.00)	(81.82)	
Shock indicator	36.31	80.26	-51.26	
	(44.81)	(56.05)	(68.52)	
Treatment * Shock	-37.79	-14.44	-68.15	
	(60.76)	(72.65)	(99.02)	

Notes to Table 6: Each coefficient in Panel A is from a separate regression. Columns in Panel B report coefficient estimates from one regression. Coefficient estimates from OLS regressions are reported. Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages; significantly different from zero at (*) 90%, (**) 95%, and (***) 99% confidence levels. Sample sizes of each regression in brackets. Individual and HH-level controls are the child's gender, indigenous status, the household's welfare index, education, age, and gender of the head of household, family size, and distance to secondary school. Village contextual controls are the proportion of secondary-school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportion of heads of households.

Table 7: Robustness Checks of Endogenous Peer Effects

Specification (dependent variable is the school enrollment indicator)	Sample	Coefficient Estimate on Social Network Enrollment Measure (s.e.) All children (1)	Welfare < Median (2)
Panel A: Village-level Enrollment Rate			
OLS [a]		0.668***	0.660***
		(0.042)	(0.053)
IV, no contextual controls [a]		0.541**	0.652***
		(0.264)	(0.235)
		[8.91]	[14.37]
IV, predetermined contextual controls [a,b]		0.492	0.671***
		(0.310)	(0.246)
		[7.60]	[13.92]
IV, predetermined & exp. related contextual controls [a,b,c]		0.486	0.657***
		(0.357)	(0.282)
		[6.70]	[12.53]
IV, predetermined contextual & school characteristics controls [d]		0.495	0.636**
		(0.305)	(0.263)
		[7.83]	[13.31]
IV, predetermined, exp. related contextual controls, & school chars. [a,b,c,d]		0.486	0.630**
		(0.367)	(0.302)
		[6.70]	[11.60]
Panel B: Total Enrollment (conditional on number of children in the village)			
OLS [a]		0.016***	0.017***
		(0.001)	(0.002)
IV/2SLS, no contextual controls [a]		0.014**	0.019**
		(0.007)	(0.007)
		[9.04]	[16.33]
Mean of dependent variable		0.587	0.559
Observations		4211	2757

Notes to Table 7: Each coefficient estimate is from a separate regression. Coefficient estimates from OLS and 2SLS regressions are reported. Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages; significantly different from zero at (*) 90%, (**) 95%, and (***) 99% confidence. First stage F-statistics of significance of partial correlation between IV (treatment indicator) and social network measure are reported in brackets.

(a) Individual and HH-level controls are the child's gender, indigenous status, the household's welfare index, education, age, and gender of the head of household, family size, and distance to secondary school. These are included in all specifications.

(b) Village predetermined contextual controls are the proportion of secondary-school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportion of heads of households.

(c) Expenditure related contextual characteristics are mean village-level HH expenditures, mean educational, food, boys' clothing, girls' clothing, alcohol & tobacco expenditure shares, and an indicator variable for whether the village suffered a flood shock.

(d) School characteristics are indicator variables for general, technical, secondary schools (relative to 'tele-secundaria' schools), urban and semi-urban school indicators (relative to rural schools), school-level pupil/teacher ratio, and the number of home teachers, teaching assistants, PE teachers, and art teachers in school.

Table 8: Estimates of Endogenous Peer Effects using Contextual Variables as IV

	Dependent variable: School enrollment indicator		
	All children	Welfare < Median	Welfare > Median
	(1)	(2)	(3)
	IV	IV	IV
Social Network Enrollment Rate	0.577***	0.449	0.743**
	(0.221)	(0.408)	(0.364)
Individual & household controls	Yes	Yes	Yes
Village contextual controls	Yes	Yes	Yes
State indicators	Yes	Yes	Yes
Observations	1678	1075	603
Hansen J-statistic	0.165	0.00	0.69
P-value	[0.68]	[0.99]	[0.40]
First Stage F-statistic [p-value]	10.00 [0.001]	3.41 [0.036]	8.25 [0.001]
Df	(2,136)	(2,122)	(2,76)
Mean of dep. Variable	0.557	0.507	0.647
IV Estimate using experimental variation (same as Table 4)	0.492	0.671***	-5.112
	(0.310)	(0.246)	(18.120)

Notes to Table 8: Coefficient estimates from IV/GMM regressions are reported. Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages; significantly different from zero at (*) 90%, (**) 95%, (***) 99% confidence. Instrumental variables are mean educational achievement of heads of households in the village, and proportion female household heads in the village. Individual and HH-level controls are the child's gender, indigenous status, the household's welfare index, education, age, and gender of the head of household, family size, and distance to secondary school. Village contextual controls are the proportion of secondary-school-age girls and the proportion of indigenous children in the village, mean village-level family size and educational level, age, and gender proportion of heads of households.

Table A1: Relationship between attrition and characteristics of children at baseline

	Dependent variable: Attrition indicator							
	Ineligible Children				Eligible Children			
	Treatment	Correlates	Main effect of correlates	Interaction of correlates with treatment	Treatment	Correlates	Main effect of correlates	Interaction of correlates with treatment
	(1)	(2)	(3)		(4)	(5)	(6)	
	OLS	OLS	OLS		OLS	OLS	OLS	
Treatment village	-0.006	0.008	0.102		0.000	0.005	0.016	
	(0.005)	(0.010)	(0.122)		(0.003)	(0.006)	(0.075)	
Treatment * Year 1998	0.008		-0.005		0.002		0.004	
	(0.020)		(0.021)		(0.011)		(0.010)	
Treatment * Year 1999	0.034		0.017		0.022		0.029**	
	(0.025)		(0.026)		(0.014)		(0.014)	
Year 1998	0.208***	0.150***	0.154***		0.196***	0.140***	0.138***	
	(0.016)	(0.010)	(0.016)		(0.009)	(0.005)	(0.008)	
Year 1999	0.201***	0.095***	0.087***		0.199***	0.096***	0.078***	
	(0.018)	(0.013)	(0.018)		(0.011)	(0.007)	(0.010)	
Child's age		0.062***	0.058***	0.005		0.059***	0.061***	-0.005
		(0.003)	(0.004)	(0.005)		(0.002)	(0.003)	(0.003)
Grade completed in 1997		-0.029***	-0.027***	-0.004		-0.019***	-0.026***	0.012**
		(0.005)	(0.008)	(0.010)		(0.003)	(0.004)	(0.005)
Gender (boy)		-0.013*	-0.022*	0.016		-0.021***	-0.017***	-0.008
		(0.008)	(0.012)	(0.015)		(0.004)	(0.006)	(0.008)
Indigenous		0.034	0.019	0.029		0.003	0.011	-0.014
		(0.021)	(0.031)	(0.041)		(0.007)	(0.012)	(0.015)
Family size		-0.004**	-0.001	-0.005		-0.002**	-0.002	0.000
		(0.002)	(0.003)	(0.004)		(0.001)	(0.002)	(0.002)
HOH education		0.002	0.003	-0.002		0.000	-0.001	0.001
		(0.002)	(0.003)	(0.004)		(0.001)	(0.002)	(0.002)
HOH gender (male)		-0.010	-0.012	0.006		-0.029***	-0.027*	-0.004
		(0.017)	(0.023)	(0.034)		(0.010)	(0.016)	(0.020)
HOH age		0.001*	0.001	0.000		0.000	0.000	-0.001
		(0.000)	(0.001)	(0.001)		0.000	0.000	0.000
Dist. to sec. school		0.003	0.004	0.000		0.003*	0.003	0.000

		(0.003)	(0.004)	(0.005)		(0.002)	(0.003)	(0.003)
Dist. to urban center		0.000	0.001**	-0.001*		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)		0.000	0.000	0.000
General sec. school		-0.053	-0.012	-0.056		0.016	0.002	0.012
		(0.034)	(0.072)	(0.083)		(0.019)	(0.028)	(0.038)
Technical sec. school		-0.067**	-0.073	0.026		0.001	-0.014	0.030
		(0.027)	(0.051)	(0.066)		(0.015)	(0.025)	(0.033)
Urban school		0.011	0.149	-0.156		-0.044**	0.078	-0.145**
		(0.029)	(0.174)	(0.176)		(0.022)	(0.059)	(0.064)
Semi-urban school		-0.003	-0.033	0.043		0.005	0.014	-0.009
		(0.023)	(0.025)	(0.041)		(0.013)	(0.018)	(0.026)
Num. of home teachers		0.001	0.009	-0.013		0.007	-0.003	0.013
		(0.012)	(0.029)	(0.032)		(0.006)	(0.014)	(0.015)
PE teachers		0.029	0.023	-0.001		0.003	-0.025	0.042
		(0.038)	(0.069)	(0.086)		(0.019)	(0.025)	(0.035)
Art teachers		-0.064*	-0.077	0.043		-0.007	-0.017	0.014
		(0.034)	(0.054)	(0.067)		(0.018)	(0.021)	(0.035)
Teaching teachers		0.021	-0.008	0.039		-0.001	0.004	-0.006
		(0.013)	(0.021)	(0.026)		(0.007)	(0.008)	(0.013)
Num. of teachers		0.007	0.003	0.008		-0.007	0.002	-0.012
		(0.013)	(0.031)	(0.034)		(0.006)	(0.014)	(0.016)
Pupil teacher ratio		0.000	0.000	-0.001		0.000	0.000	-0.001
		(0.001)	(0.001)	(0.001)		0.000	(0.001)	(0.001)
Constant	0.011**	-0.681***	-0.755***		0.006***	-0.596***	-0.605***	
	(0.005)	(0.056)	(0.104)		(0.002)	(0.031)	(0.066)	
State indicators	Yes	Yes	Yes		Yes	Yes	Yes	
State * Treatment indicators	No	No	Yes		No	No	Yes	
Observations	8214	8184	8184		33441	33351	33351	
R-squared	0.08	0.16	0.17		0.08	0.15	0.15	
Interactions F-statistic		-	1.29			-	1.36	
P-value			[0.149]				[0.105]	

Notes to Table A1: Robust standard errors in parentheses; disturbance terms are allowed to be correlated within villages, but not across villages. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. F-statistic of joint significance of interaction terms (F(28,379) reported at bottom of the table.