CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Serveur académique lausannois

Social Interactions and Schooling Decisions

Rafael Lalive*
University of Lausanne and CEPR

Alejandra Cattaneo
Swiss Coordination Center for Research in Education

September 14, 2007

Abstract

The aim of this paper is to study whether schooling choices of a child are affected by the schooling choices of other children. Identification is based on a randomized targeted intervention that grants a cash subsidy conditional on school attendance to a sub-group of *eligible* children within small rural villages in Mexico (PROGRESA). This policy change spills over to ineligible children if social interactions are relevant. Results indicate that the *eligible* children tend to attend school more frequently, and the *ineligible* children acquire more schooling when the subsidy is introduced in their local village. Moreover, the overall effect of PROGRESA on *eligible* children is the sum of a direct effect due to cash transfers and an indirect effect due to changes in peer group schooling. Interestingly, the social interactions effect is almost as important as the direct effect.

JEL-Classification: C93, I21, I28

Keywords: social interactions, social multiplier, schooling, field experiment, PROGRESA.

*We would like to thank Michael Bognanno, Armin Falk, Bernhard Ganglmair, Jason Fletcher, Michael Kosfeld, Jeffrey Kling, Alan Krueger, Chuck Manski, Ulrich Müller, Tanya Rosenblat, Florian Pelgrin, Jesse Rothstein, Alois Stutzer, Till von Wachter, Georg Weizsäcker, Rainer Winkelmann, Josef Zweimüller, seminar participants at Columbia, Innsbruck, IZA, Ringberg, Princeton, Uppsala, and Zürich, and an anonymous referee whose comments have substantially improve on an earlier version of this paper. We would like to thank Iliana Yaschine Arroyo at the Evaluation Center of the Oportunidades Program for access to the data, and Beatrix Brgger for excellent research assistance. Manuela Angelucci and Giacomo De Giorgi provided generous access to the information on consumption in PROGRESA. Rafael Lalive acknowledges financial support by IEW.
Rafael Lalive is also affiliated with IEW (University of Zurich), CESifo, IZA, and IFAU. Address: Rafael Lalive, Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne-Dorigny, Rafael.Lalive@unil.ch. Alejandra Cattaneo, SKBF, Entfelderstrasse 61, CH-5000 Aarau, maria.cattaneo@skbf-csre.ch

1 Introduction

> "Curiously enough, when one of the students would get sick, one or two of his closest friends would also not show up to school." – School principal in a rural school in Mexico (Adato *et al.*, 2000, p. 54)

All social sciences agree that education is very important in every person's life. Understanding when and why children leave school for work is thus crucial. Economists have been quite successful in developing a framework that explains schooling decisions (Becker, 1964). Yet, while economists have typically focused on the *individual* costs and benefits of acquiring further schooling (Card, 1999), sociologists have long pointed out that schooling decisions could be affected by the social environment of a child (Coleman, 1961).

This paper aims to provide evidence of the importance of the social determinants of schooling among children enrolled in the upper grades of primary school in rural Mexico. A comprehensive understanding of the factors shaping education investment decisions is particularly needed in this setting. Children are often absent from school and many leave school when they complete primary school at the age of 12 years – notably two years before they reach the minimum age required for employment. To address these schooling problems, the Mexican ministry of education set up an innovative program – the PROGRESA program – designed to encourage school attendance among poor families living in small villages in rural Mexico. The program consists of a cash grant paid to the mother for each child in grade 3-6 of primary school or grade 1-3 of secondary school who attends school regularly, i.e. more than 85 % of all school days during every 2 month period of the school year. Clearly, this program directly increases the incentive to attend school among eligible children, i.e. children living in households that are classified as poor. Indeed, the existing evidence indicates that this program is highly successful in increasing primary and secondary school attendance rates by around 6-10 percentage points.[1]

The basic idea of this paper is to assess the response of the ineligible children – living in a household that has not been classified as poor – to introducing the program in their peer group. Clearly, ineligible children's families do not receive additional income when they send their child to school. However, as the initial quote suggests, ineligible children might want to spend more time in the classroom because their peers – children in the same grade living in the same village – attend school more frequently. The spillover effect of the program on the

[1] See Behrman *et al.* (2001), Buddelmeyer and Skoufias (2003), and Attanasio *et al.* (2003) for evaluations of this program.

ineligible children can be used to identify the role of social interactions in schooling decisions (Moffitt, 2001).

Identifying social interactions is difficult because individuals tend to select into similar groups – an "omitted variable" problem – and because each member of the peer group is affecting every other member – the "reflection problem" (Manski, 1993).[2] PROGRESA is ideally suited to address these two important identification problems. First, the program was randomly implemented. Randomization balances all determinants of school attendance directly addressing the "omitted variable" problem. The "reflection problem" can be addressed because PROGRESA grants are only paid to a sub-group of *eligible* children within small villages. Thus, PROGRESA increases peer group school attendance while leaving unaffected the *ineligible* child's monetary incentive to attend school.[3] This means that the response among ineligible children provides information on how strongly the peer group affects the individual and not vice versa.

Our empirical results document a sizeable effect of schooling subsidies on school attendance among *eligible* children. With regard to social interactions, we find, first, that there is a positive and statistically significant average spillover effect of the program on *ineligible* children. We also find that the ineligible child's response is larger the higher is the eligible fraction of children among their classroom peers. Second, the social interaction effect is sizeable. When we combine the response of the ineligible student with information on the effect of PROGRESA on peer group schooling, we find that the ineligible students' schooling decisions are strongly, and statistically significantly, affected by their peer's decision. Third, we find that the direct effect of the cash subsidy on school attendance is about as large as the social spillover effect among children from poor households. Thus, the total effect of PROGRESA on eligible children can be explained via both the direct effect of the subsidy on the individual and the indirect effect of the subsidy on the social environment of poor children.

Our findings are important for at least three reasons. First, these findings suggest that the social environment of an individual is relevant in affecting one of the most important investment decisions. This implies that economists should pay attention to social interactions in analyzing schooling choices both theoretically and empirically. Second, endogenous social interactions in schooling decisions entail amplification of the effects of schooling interventions

[2]Manski (1993) called this the "reflection problem" because solving it is analogous to determining whether the mirror image is causing me to act like I do or vice versa. Also see the more recent discussions in Manski (1995) and Manski (2000) on the identification problems with endogenous social interactions.

[3]The ineligible children might be affected indirectly by the program if the poor share their grants with the non-poor. However, there is no evidence of direct sharing of the benefits. Furthermore, the poor appear to have used PROGRESA transfers to cover schooling costs (school material and transportation) or purchase children's clothes. Only about 20 out of the 506 villages have local markets where children clothes are sold. This means that the additional income of the poor is unlikely to have altered income among the non-poor.

because endogenous social interactions give rise to a social multiplier (Glaeser *et al.*, 2003). Third, evidence on the quantitative importance of social interactions is essential in thinking about targeting of this program. PROGRESA was set up to target poor households. If social interactions are important, the average effect of the program on the poor is larger if many other poor households within the same village are eligible for the program.

The existing literature on social interactions in schooling can be grouped in two strands. The first strand of the literature that addresses the "omitted variable problem".[4] Hoxby (2000) identifies peer effects from gender and race variation. Sacerdote (2001) studies peer effects among college freshman at Dartmouth college who are assigned to dorms at random. Hanushek *et al.* (2001) study how peer ability affects student achievement. Betts and Zau (2004) use administrative panel data to study peer groups and academic achievement. However, while these studies address the "omitted variable bias", panel data does not allow addressing the "reflection problem". Case and Katz (1991) use instrumental variables to study neighborhood effects in the Boston area. Duflo and Saez (2003) study the role of information and social interactions in retirement plan decisions in a field experiment. Angrist and Lang (2004) measure peer effects in academic achievement using quasi-experimental features in the placement of disadvantaged children in a Boston desegregation program. Miguel *et al.* (2004) study a merit-based incentive program in Kenya that generates strong effects among eligible girls but also spillover effects on ineligible boys. Cipollone and Rosolia (2007) study social interactions in high school using the effects of an earthquake induced exemption from compulsory military service for young men on the schooling choices of young women in Southern Italy.[5]

The paper that is closest to ours is Bobonis and Finan (2005) who study social interactions in schooling decisions using the PROGRESA as an instrument for peer schooling. This paper goes beyond the existing literature in at least three important respects. First, Bobonis and Finan (2005) use the PROGRESA status of a village to identify social interactions. This analysis can not address village specific shocks to schooling. This paper proposes an identification strategy that uses information on a child's eligible classroom peers in PROGRESA villages as an instrument for peer group schooling. Importantly, this instrument varies within villages allowing us to identify peer effects *within* rather than *across* villages. Our results indicate that the spillovers on ineligible children indeed arise within villages rather than across villages. Second, the overall treatment effect of a targeted intervention on eligible individuals

[4]See Glaeser *et al.* (1996) on social interaction in crime, and Lalive (2003), Topa (2001), and Topa and Conley (2002) for social interaction in unemployment.

[5]A related issue arises in studying the impacts of de-worming treatment in developing countries. Miguel and Kremer (2004) find strong epidemiological spillovers generated by de-worming treatment in rural Kenya.

consists of a direct effect due to changes in economic incentives and of an indirect effect due to endogenous social interactions. From the policy point of view, decomposing the overall effect into the direct effect and indirect effect is important because it allows assessing whether a program's overall impact is sensitive to the way it is targeted. To our knowledge, this is the first paper that discusses how to separate the direct effect of the program from the indirect effect of the program on eligible children. Our findings indicate that indirect effects are very important in the PROGRESA context in the sense that it is almost as important as the direct effect due to economic incentives. Third, this is the first paper that empirically documents how endogenous social interactions amplify the effects of individual level interventions. We find that PROGRESA accounts for twice as much of the variance in grade level differences in school attendance than it does at the individual level.[6]

The remainder of this paper is organized as follows. Section 2 presents background information on Mexico and PROGRESA. Section 3 discusses the data and presents descriptive evidence. Section 4 discusses the identification strategy. Section 5 presents the main results, and Section 6 concludes.

2 Background

This section discusses the education system in Mexico, and it provides important background information on PROGRESA. The section then discusses why social interactions might arise in the PROGRESA setting and it addresses possible identification concerns.

2.1 Primary and Secondary Education in Mexico

According to the general education act from 1993 the educational system in Mexico comprises 3 levels: basic, which is sub-classified in pre-primary, primary and lower secondary; upper secondary and tertiary. However the only two types that are obligatory for all Mexican citizens according to the Mexican political constitution are primary and lower secondary school. The Mexican school system is quite centralized. Schools have limited decision-making autonomy, with only 22 % of all decisions taken at the school level, 45 % of all decisions taken at the state level and 30 % taken at the central level (OECD, 2004).

Whereas Mexico has made substantial progress in terms of average educational attainment, there is still a strong discrepancy in terms of education attainment between rural and central areas (Hanson, 2002). This is because many rural villages have a local primary school but

[6] Glaeser *et al.* (2003) estimate the social multiplier for exogenous background characteristics of Dartmouth college roommates and for the returns to schooling. These estimates of the social multiplier do not allow discussing amplification of actual policy interventions.

the secondary schools are only present in larger villages or small cities. Thus, distance to secondary school is an important factor explaining low educational attainment. Distance to school is perceived to be particularly problematic and dangerous for girls (Adato *et al.*, 2000, p. 73). Also, poor families simply can not afford to send all children to primary and secondary school. Moreover, children do not want to continue with school due to laziness, boredom with school and preference for work, girls would rather be with their boyfriend than in school, teachers treat children badly, children want their own income rather than study (Adato *et al.*, 2000, p. 72). After school, most children in rural areas are expected to perform a variety of household chores such as taking care of animals, help out in the kitchen, gather firewood, help out in building a fence, etc. These tasks are demanding and require balancing the schoolwork and the housework schedules (Adato *et al.*, 2000, p. 66).

2.2 PROGRESA

In order to encourage enrollment and permanence in school of children and teenagers under 18 years old who attend grades between third of primary and third of secondary school the Mexican government created PROGRESA (Programa de Educacion, Salud y Alimentacion)[7] which is a program aimed at increasing the opportunities and complementing the income of Mexican families living in conditions of extreme poverty. It has three components: education, health and nutrition. PROGRESA's health and nutrition components primarily target children aged 0-60 months (Gertler, 2004). Other family members visit clinics once a year for checkups and receive information concerning health prevention and nutrition at monthly meetings (*pláticas*).

PROGRESA's educational component consists of cash transfers which are provided to poor families every two months during the school year (August to June) conditional on sending their children to school. The cash grants have two particularities: the sums granted increase as children reach higher grades and – in the secondary school – the sums awarded to girls are slightly higher than those for boys to compensate for the slightly higher proportion of girls dropping out of school (Table A1).[8] The nominal values of the cash transfers are adjusted every 6 months to take into account changes in cost of living. The cash subsidy is handed out to the mother because of the belief that the mother is usually better administrating the household resources and because women are disproportionately vulnerable to poverty. The grants are awarded only after confirming that the child has been present on more than 85% of all school days in successive bi-monthly periods during the school year. If they fail to fulfill

[7]The Program was re-named Oportunidades under the Fox administration.

[8]Beneficiary families also receive funding to pay for school material of 135 pesos for children in primary school and 170 pesos for children in secondary school.

this requirement they loose the grant, at first temporarily and then permanently. Attendance is monitored by school teachers. There are only very few reports of parents trying to influence teachers to misreport attendance (Adato *et al.*, 2000). The PROGRESA transfers go directly from the Federal Budget to beneficiary households.

The most important advantage of this program from the perspective of this paper is that PROGRESA is a partial-population intervention that was phased-in at random.[9] Specifically, the program was implemented in three steps. In the first step, the Mexican government selected an initial set of 506 rural villages which are characterized by a high degree of "marginality", but with access to education and health facilities, on the basis of the 1990 and 1995 census.[10] These 506 villages are located in seven states (*Guerrero, Hidalgo, Michoacán, Puebla, Querétaro, San Luis Potosí and Veracruz*) which are loosely clustered around Mexico City.

In the second step, PROGRESA determined the poverty status of each of about 24,000 households living in these 506 villages based on survey information collected in October 1997. Basically, the poverty status of the household was determined using information on educational attainment of the household head and her or his partner, dwelling characteristics, and other information that predicts the per capita household income. The poverty status of a household was determined by condensing this information into a so-called "poverty index". A household was classified as poor if its poverty index exceeded a state specific poverty threshold. On average, this procedure led to 52 % of all households being classified as poor but there is also substantial variation with respect to the percentage of households that are eligible within a village.

In a third step, PROGRESA determined at random a set of 320 villages where the program was implemented as of August 1998. The remaining 186 villages were excluded from the list of PROGRESA villages until the end of the 1999 / 2000 school year. This randomized phasing-in of the program allows evaluating the impacts of PROGRESA in a randomized design.[11]

In the second year of the program, (August 1999 to June 2000), PROGRESA added a further 26 % of all household to the list of beneficiaries due to complaints that the initial procedure discriminated against households whose children had already left home. However, by the year 2000, PROGRESA staff found that none of the newly admitted households had collected any cash benefits. Apparently, none of the newly admitted households had been notified of their eligibility for the program (Buddelmeyer and Skoufias, 2003). Because an

[9] See Skoufias (2001) for an in-depth discussion of the implementation of the program.

[10] The marginality index compresses information on literacy, share of dwellings without water, drainage or electricity, average number of occupants in one room, share of dwellings with dirt floor, and share of population working in primary sector into one variable by means of a principal components analysis.

[11] Randomized evaluation proved crucial in securing a loan from the World Bank to roll out the program in all of Mexico in 2000.

in-depth understanding of how the program was implemented is crucial but missing, we limit all analyses to the first year of the program.

2.3 Social Interactions

The basic idea of this paper is that children from non-poor households also might decide to stay in school longer when PROGRESA induces children from poor households within their village to stay in school longer. The main idea is that children like to spend time with their friends as the statement by the school director cited in the introduction clearly shows. Thus, when prior to PROGRESA a child from a poor household reported sick and some of his or her friends from non-poor households curiously also missed school, the non-poor children now have no reason to miss school when all children from poor households attend school regularly.

Clearly, such social interactions are expected to be strong among children who have reached the same grade level and who are living in the same village. Most of these children are classmates because villages are so small that most of them only have one local school. The relevant peer group therefore consists of all children living in the same village who have reached the same grade level. Is PROGRESA useful in identifying social interactions between poor and non-poor families? If children from poor households only interact with other children from poor households, there could be important social spillover effects that can not be detected with the PROGRESA experiment. However, any social interactions that we find can be thought to represent a lower bound on within poor / within non-poor social interactions. Moreover, there is strong evidence that the inhabitants of the 506 rural and extremely poor villages in Mexico felt that the selection of poor families was quite arbitrary.[12] This suggests that social relationships exists also between the poor and the non-poor. Moreover, the fact that the program was implemented did not seem to change existing social relationships.[13]

2.4 Confounding Effects

It is essential to discuss four alternative hypotheses that motivate a spillover effect on the children of ineligible households. It might, first, be possible that poor households share the education grants with non-poor households. In the official evaluation Adato *et al.* (2000) do not find sharing of benefits to any substantial extent. This is probably due to the fact that the transfers are not perceived as salient. According to in-depth focus group research, these

[12]"Among beneficiaries, non-beneficiaries and *promotoras*, there was a strong view expressed that 'everyone is poor' – a sense of common identity in poverty." (Adato, 2000, p. vi)

[13]"Many comments were made suggesting that beneficiaries and non-beneficiaries continue to get along with each other fine and 'the same' as before." (Adato, 2000, p. vi)

benefits are helpful in financing a child's education but they do not pay for much else (Adato *et al.*, 2000).

Second, even if the poor do not share the grants directly with the non-poor, they might spend the additional income in shops owned by the non-poor thus also directly affecting the incomes of the non-poor. The existing evidence shows that the transfers are mainly used to finance the children's education and clothes for children (Bobonis, 2004). Detailed village level data suggests that the grants are spent outside the village rather than in the local village. Only 20 out of the 506 villages have a local supermarket or street market. Nevertheless, it is possible to investigate whether there are indirect effects of PROGRESA on the non-poor households as we discuss in section 4. The basic idea is that non-social spillover effects are likely to be similar for children living in the same village. Our main identification strategy uses information on the eligible fraction of each child's peers which varies within villages. As we explain in section 4 we can thus construct an identification strategy that is not affected by household specific spillover effects. [14]

Third, it might be that the non-poor misunderstood the working of the program and believed that they are eligible as well. This is unlikely to be the case. Between October 1997 and the start of the program in August 1998, PROGRESA held public meetings in which the eligibility status of each household was clearly communicated. Moreover, the education subsidy was administered using two forms. The E1 form recorded background information on all children from eligible households. These E1 forms were the basis of the E2 form, a list of eligible children, which was sent out to each school which is attended by eligible children. The E1 forms were not distributed to non-beneficiary households. Nevertheless, we investigate the sensitivity of our main results to possible anticipation effects by focusing on those families that are least likely to have been re-admitted to the program during the second year.

Fourth, ineligible children may also have attended school more regularly due to increases in the quality of teaching, or the quality of schools. However, in their in-depth evaluation of PROGRESA, Adato *et al.* (2000) do not find any change in terms of the overall quality of education in PROGRESA villages compared to control communities. In PROGRESA villages, 54.9 % of all school directors state that the overall conditions of the school has improved since PROGRESA started. In control villages, the corresponding figure is 9 percentage points *higher* (63.9 %) motivating a concern with underestimation of the spillover effect.

[14] An effect on ineligible children's school attendance could also be due to a health spillover when the incidence of communicable diseases is reduced among beneficiary children (Miguel and Kremer, 2004). However, Gertler (2000) does not find an effect of PROGRESA on health among beneficiary children aged 6-17 years. Moreover, positive health spillover effects for ineligible parents again are household specific. These can be addressed using the same identification strategy discussed in section 4.

3 Data and Descriptive Evidence

This section provides background information on the data set that we use in the empirical analysis. The section then provides evidence on random assignment and on the effects of PROGRESA on school attendance.

3.1 Data

The official PROGRESA evaluation database contains annual survey information on school attendance, socio-economic characteristics, and localities between October 1997 and November 1999.[15] The empirical analysis primarily uses information on two waves, October 1997 and October 1998. The first wave provides information on school attendance and socio-economic background before the program was implemented. The second wave is useful in assessing the effect of the program because the program was implemented in August 1998.

We concentrate on children living with their mother who have completed grades 3 to 6 of primary school in October 1997.[16] This sample consists of children from poor households whose family directly became eligible when PROGRESA was introduced, and children from non-poor households whose family did not become eligible in August 1998 but who are expected to know eligible children in their classroom and village. Moreover, this sample covers children who have not yet finished primary school (children having completed grades 3 and 4 of primary school) as well as children making the transition from primary school to secondary school or drop out of school (children having completed grades 5 or 6 of primary school). Thus, the sample allows discussing whether social interactions are relevant in attendance and drop-out decisions. Second, we concentrate on children with non-missing information on school attendance in 1997 and 1998.[17] We end up with a sample of 15,653 children of which 9,690 live in "treated" villages – where the program was implemented in August 1998 – and 5,963 live in "control" villages that were denied access to the program in August 1998. This sample comprises both

[15] To our knowledge, the administrative data generated in paying out the subsidy is not available for researchers. PROGRESA also collected information during the spring. The number of completed interviews in the spring is, however, 20 % lower than in the fall. This likely reflects the seasonal pattern of field work in the villages. Since non-response may be non-random, we focus on the interviews conducted in the fall rather than in spring.

[16] The sample selection can not be based on the grade attained in October 1998 because grade attainment in 1998 is already affected by PROGRESA and, therefore, endogenous. Grade attained in 1997 is also determined by a number of factors we do not observe. However, randomized implementation of the program ensures that grade attainment is the same in PROGRESA villages and control villages. Moreover, this criterion rules out children who have completed grades 1-3 of secondary school in the year prior to the start of the program. However, note that (i) for many of these children we have no information on school attendance in October 1998 (this item is only available for children aged 6-16), and (ii) many villages do not have a local secondary school – implying that the children in the local village are only a part of the social network.

[17] This criterion restricts the age limit to be between 6 years and 16 years in 1998 because the outcome indicator – school attendance – was only collected for children in this age group in 1998. These age limits are not binding, however, since we are focusing on a sample that has completed grade 3 to 6 in 1997.

children who are still enrolled in school as well as children who have already left school either temporarily or permanently. This is advantageous since our focus is to study the effect of social interactions on overall schooling attendance decisions.

The focus of this paper is on grade specific social interactions. The peer group of a child consists of all children who are living in the same village and have completed the same grade in October 1997 (pre-program). We measure social interactions between a child that has, for instance, completed 3rd grade in October 1997 and all other children living in the same village who have completed 3rd grade in 1997.[18] All children in this peer group have reached the same schooling level and are therefore at the same stage in their schooling career. However, note that these peer groups may be heterogeneous with respect to age due to temporary school exits or grade repetition.[19] We therefore investigate the sensitivity of our main results with respect to heterogeneity in age and school progression in a supplementary analysis that focuses on children whose school career is uninterrupted – children on grade level.

3.2 Descriptive Evidence

Table 1 provides descriptive statistics on the background characteristics of the children in the sample. Panel A in Table 1 reports statistics for the 10,484 children living in poor households. The distribution of children across grades is slightly skewed to the right. The fraction of children having completed grades 3, 4, and 5 is about 6 percentage points lower than the fraction of children having completed grade 6. This is due to the fact that some children for whom the highest grade attained is the sixth grade in primary school have already left the education system. Parental education – a powerful predictor of household income – is very poor.[20] Roughly 16 percent of all children have a mother or father who has completed primary school, i.e. has reached grade 6 of primary school or a higher grade level.[21] Children in poor households also do not tend to live in dwellings with a cement floor or firm roof. (The omitted categories concerning the roof type refer to roofs made of cardboard or palm leaves.) Whereas all villages have a local primary school, many villages do not have a local secondary school.

[18] We use pre-program information from the year 1997 rather than during program information from the year 1998 to define peer groups to rule out that endogenous grade progression or moving is affecting results. Also, we allow for social interactions between all children in the village because these villages are small.

[19] The average within peer group standard deviation of age is 1.36 years.

[20] We have constructed a direct measure of household income. It turns out that this measure does not add to the empirical analysis we report below. We do not report the constructed household income measure for two reasons. First, household income is much more likely subject to reporting bias. Second, it is well known that measuring household income is difficult in agricultural societies due to the importance of home production.

[21] Note that parental education is affected by non-response. For about 33 percent of all children, there is no information regarding parental education. Further analysis (not shown) of this fact indicates that non-response increases strongly with age. This suggests that non-response refers to no schooling at all. Results are not sensitive to adding an indicator that reflects non-response concerning parental education.

Thus, the costs of attending secondary school are large and the incentive to finish primary school is weak. Peer groups are relatively small with about 11 to 12 children in each grade per village.[22]

Table 1 about here

Comparing Panel A and Panel B in Table 1 allows inferring to what extent our proxies for household income predict poverty status of the household. There is strong evidence that poverty status is related to parental education, and dwelling characteristics. The percentage of children with a mother having completed primary education is about 16 percent among poor children (Panel A), and 22 percent among non-poor children (Panel B). Father primary school completion follows a similar pattern. With respect to dwelling characteristics, we find that whereas only about 25 percent of all poor children live in a dwelling that has a cement floor (as opposed to a dirt floor), the corresponding figure is roughly 60 percent among non-poor children. The fraction of children living in a dwelling with a permanent roof type is much higher among non-poor children than among poor children. There is also an interesting difference between poor and non-poor children with respect to grade. We find that the fraction of children having completed grade 6 is 34 percent among children from non-poor households but only 28 percent among children from poor households. This is consistent with a higher fraction of children dropping out of school before completing primary school among poor children.

The main result in Table 1, however, is that there is no difference between villages with PROGRESA and control villages with respect to any of these important background characteristics of poor and non-poor children. This suggests that randomization successfully generated independence between PROGRESA status of the village and observed (and potentially also unobserved) characteristics. Thus, the effects of PROGRESA on school attendance can be identified convincingly because treatment differences in terms of school attendance are likely due to implementing the program rather than due to differences in terms of the average poverty level.

Table 2 reports descriptive evidence on the effect of the cash subsidy on school attendance. School attendance is a binary indicator variable taking the value 1 if the child attends school at the date of the interview, and zero otherwise. School attendance reflects both attendance and enrollment. Panel A in Table 2 reports effects for the eligible children living in poor households. On average, only about 77 % of all children in grades 3-6 attend school in control villages in October 1997. In treated villages, school attendance is slightly higher, 78 %, one

[22]There are 16 children with an empty peer group (living in small villages). About 10 % of all children have a peer group smaller or equal to 4 children and 90 % of all children have a peer group that is smaller or equal to 24 children.

year prior to the start of the program. However, the treatment contrast is not significant at any conventional level of significance. One year later, in October 1998, school attendance is 69 % in control villages – 8 percentage points lower than the year before. This means that a substantial fraction of children in our sample have dropped out of school in control villages. In contrast, in treated villages school attendance is 76 % – only 2 percentage points lower than the year before. This means that the program increased school attendance by 6 percentage points – a significant impact both in the economic and statistical sense.

Table 2 about here

Panel B in Table 2 discusses the spillover effect of the cash subsidy on the ineligible children whose household was not classified as poor. The idea is that children from non-poor households in villages with PROGRESA do not receive a cash subsidy but they are living in the same village as children whose school attendance has been strongly increased. The control villages provide information on the counterfactual situation without PROGRESA. The data indicate that school attendance is about 76 % in control villages, and about 78 % in treated villages about one year before the program was introduced – the treatment contrast being insignificant. By October 1998, school attendance has dropped by 7 percentage points in control villages but only by 5 percentage points for ineligible children in treated villages. Thus, the program appears to have reduced school drop out sightly, by 2 percentage points, among ineligible children. However, this "spillover" effect is not significantly different from zero. Thus, results in Table 2 indicate that PROGRESA strongly increases schooling for the eligible children but only very weakly for the ineligible children suggesting weak or no social spillovers of the program.

Table 3 reports the effect of PROGRESA on poor and non-poor children's change in school attendance using linear regression analysis that controls for all the observed characteristics of children. The change in school attendance is the difference between the school attendance indicator between October 1998 and October 1997 for each child. This implies that Table 3 reports a within individual difference-in-difference analysis of the effect of PROGRESA on school attendance. It is advisable to use the difference-in-difference strategy since evidence in Table 2 indicates that school attendance levels are slightly higher in treated villages compared to control villages before the program was implemented. Inference is based on robust standard errors allowing for clustering at the village level.

Results indicate that there is a statistically significant and quantitatively important increase in school attendance trends among poor children (Table 3 Panel A). Villages with program experience a 5.8 percentage points weaker downward trend in school attendance compared to

villages without the program. Second, controlling for observed characteristics has virtually no impact on the estimated effect of PROGRESA among poor children. There is a strong grade related pattern in school attendance trends. Children who have completed grade 4 are affected by a slightly higher drop in school attendance than the reference – children who have completed grade 3. Children in grade 6 (having completed grade 5) experience a 19 percentage points stronger reduction in school attendance than children in grade 4 (having completed grade 3). This shows that the decision to leave school is primarily taken in grade 6. Interestingly, upon primary school completion (children who have completed grade 6) there is a slightly more favorable trend in school attendance than for children having completed grade 3. A more detailed analysis (not shown) indicates that there is an equal proportion of students attending school in October 1997 but not attending school in October 1998 (school leavers) and students with exactly the reverse pattern (school entrants) among the children who have completed grade 6. This shows that PROGRESA not only affects those who are currently in school but also those who do not attend school. In addition to grade level, missing father information reduces the trend in school attendance and local presence of the secondary school increases the trend in school attendance. There are no further statistically important determinants of the trend in school attendance.[23]

Table 3 about here

The second Column in Table 3 reports results for the non-poor. Implementing PROGRESA increases the trend in school attendance in villages with program by 2.1 percentage points compared to the villages without the program. This estimate is slightly higher than the estimate reported in Table 2, and it is statistically significant at the 10 percent level. The results concerning the correlation between trends in school attendance are similar among the non-poor as among the poor, except for grade 6 (not significantly different from zero), and missing father information (not significant). Moreover, father education status is positively correlated with the trend in school attendance. Children whose father has completed primary school tend to have 3.3 percentage points higher change in school attendance.

An interesting first result emerges from this discussion. In the context of the PROGRESA experiment, we not only find a treatment effect among eligible children but a weaker spillover effect is also present among ineligible children. In the following section we discuss how to use information on this spillover effect to identify endogenous social interactions.

[23]Note, however, that our control variables are important predictors of the level of school attendance. Our finding that control variables are weak predictors of trends in school attendance implies that their effects on the level of schooling are time invariant.

4 Identification

The purpose of this section is to discuss identification of social interactions and to discuss how to measure the direct effect of a targeted intervention in a situation with endogenous social interactions.

4.1 PROGRESA as an Instrument

Let S_{igv} denote the change in school attendance between October 1997 and October 1998 of child i in peer group g in village v.[24] Let $P_{igv} = 1$ if the child lives in a household classified as poor by PROGRESA, and $P_{igv} = 0$ otherwise. Let $T_v = 1$ if PROGRESA has been implemented in the village of residence of child i in August 1998, and $T_v = 0$ otherwise.[25] Table 3 reports the results of the following regression

$$S_{igv} = \pi_0 + \pi_1 T_v + \mu_{igv} \tag{1}$$

where $E(\mu_{igv}|T_v) = 0$ by virtue of randomized assignment of PROGRESA to villages when estimating this model in the sub-groups of the non-poor and the poor.

Estimating regression (1) among the poor yields an estimate of the total effect of the program on the poor, π_1^P. As we show below, this overall effect is the sum of a direct effect applying only to eligible (= poor) children and an indirect effect due to endogenous social interactions. Clearly, estimates of π_1^P do not provide evidence on the role of social interactions in schooling.

Estimating regression (1) among the non-poor yields an estimate of the spillover effect of the program on the non-poor, π_1^N. This spillover effect can potentially be informative on endogenous social interactions.

Equation (1) is a reduced form model of schooling decisions. The treatment effects on the poor, π_1^P, and on the non-poor, π_1^N, can be interpreted by considering the following structural model (Manski, 1993) of social interactions.[26]

$$S_{igv} = \alpha_{gv} + \beta P_{gv} + \gamma S_{gv} + \delta T_v P_{igv} + \eta P_{igv} + \epsilon_{igv} \tag{2}$$

where P_{gv} is the fraction of poor children in the peer group of child i (all children in same

[24] The empirical analysis focuses on the change in school attendance because the level of school attendance is not perfectly balanced between control and treated villages in October 1997 (Table 2). Analyzing school attendance trends addresses time-invariant unobserved individual and village level heterogeneity.

[25] Note that this section omits background characteristics for ease of exposition. All estimates do take the observed characteristics into account.

[26] We discuss identification in a linear model of social interactions. See Brock and Durlauf (2001) for a discussion of identification in non-linear models.

village and grade as child i, excluding i), S_{gv} is the mean change in school attendance in the peer group of child i (all children in same village and grade as child i, excluding i). We also assume that $E(\epsilon_{igv}|g,v) = 0$ – a normalization that defines the unobserved α_g to capture all unobserved group level differences.

The are three sets of parameters that capture possible group level effects on individual behavior. The parameter α_{gv} is a peer group specific unobserved effect (capturing "correlated" interactions in Manski (1993)'s terminology). The parameter β captures "exogenous" social interactions that may arise, for instance, if child i has lower access to books in a classroom where many other children are classified as poor (implying that $\beta < 0$). The parameter γ measures "endogenous" social interactions, i.e. the extent to which individual change in school attendance is affected by peer group average change in school attendance.

There are also two parameters that capture individual behavior. The parameter δ measures the direct effect of the schooling subsidy. The parameter η measures differences in school attendance trends between poor and non-poor children.

The structural model (2) can be used to discuss the components of the overall treatment effect, π_1^P, that we report in Table 3. This parameter is estimated by comparing mean school attendance trends in treated and control villages. Collecting terms that are specific to children from poor households in treated villages shows that

$$\pi_1^P = \delta + \gamma(E[S_{gv}|T_v = 1, P_{igv} = 1] - E[S_{gv}|T_v = 0, P_{igv} = 1]) \tag{3}$$

(Note that unobserved group effects α_{gv} and eligible fraction P_{gv} is balanced between treated and control villages due to random assignment of PROGRESA to villages.)

Equation (3) shows that the overall treatment effect on poor children is the sum of the direct effect (δ) of the program and the indirect effect of the program. The indirect effect is the product of the social interactions parameter γ with the effect of the program on peer group average schooling. The overall effect π_1^P is larger than the direct effect if social interactions are important ($\gamma \neq 0$) and if the targeted intervention changes peer group schooling ($E[S_{gv}|T_v = 1, P_{igv} = 1] - E[S_{gv}|T_v = 0, P_{igv} = 1]$). Equation (3) also shows that social interactions can not be identified from the eligible children because there is both a direct effect and a social spillover effect.

Collecting terms that are specific to ineligible children in treated villages in the structural model (2) shows that the "treatment" effect on the ineligible children, π_1^N is

$$\pi_1^N = \gamma(E[S_{gv}|T_v = 1, P_{igv} = 0] - E[S_{gv}|T_v = 0, P_{igv} = 0]) \quad (4)$$

(provided that unobserved group effects α_{gv} and peer group eligible fraction P_{gv} are balanced with respect to treatment).

Equation (4) shows that the "treatment" effect of the program on the ineligible children reflects social interactions. This social spillover effect is important if the social interactions parameter $\gamma \neq 0$ and if the targeted program is changing peer group schooling.

Equation (4) also suggests an identification strategy. The ratio of the "treatment" effect of the program on the ineligible children to the "treatment" effect on of the program on their peer groups identifies γ.[27] The "treatment" effect of the program on peer group schooling can be identified in the following regression

$$S_{gv} = \pi_{0g} + \pi_{1g} T_v + \mu_{gv} \quad (5)$$

Equations (1) and (5) are the reduced form equations defining the "IV1" identification strategy. Intuitively, PROGRESA is taken as an instrument for peer group schooling for ineligible children. This instrument provides a shock to peer group average schooling (measured in equation 5). Assuming that PROGRESA does not directly change schooling choices of ineligible children (i.e. excluding T_v from the structural model (2)), school attendance of ineligible children will therefore only increase if social interactions are present (measured in equation 1).

The IV1 identification strategy has been used in the existing literature on social interactions in schooling decisions (Bobonis and Finan, 2005). There are two important problems with this identification strategy. First, this strategy is not robust to PROGRESA village specific shocks to schooling. Such shocks can arise if eligible and ineligible households pool their incomes or due to improved credit arrangements, etc.. Second, this identification strategy is not derived from the underlying structural model of schooling (2). We therefore proceed to discuss a second identification strategy that addresses these two issues.

[27] Note that in this case of a binary instrument, the IV estimator is a Wald (1940) estimator.

4.2 Eligible Fraction in Treated Village as Instrument

The "IV2 strategy" recognizes that there is tremendous pre-program variation in the percentage of children from poor households within a non-poor child's peer group.[28] This means that the effect of PROGRESA on peer group average schooling is stronger for grades with many children in PROGRESA than for grades with few children in PROGRESA, i.e. it depends on the peer group's eligible composition. Taking the expectation of equation (5) at the peer group level, shows that the model for peer group average school attendance is in social equilibrium (assuming that $|\gamma < 1|$)

$$S_{gv} = \frac{\alpha_{gv}}{1-\gamma} + \frac{\beta+\eta}{1-\gamma}P_{gv} + \frac{\delta}{1-\gamma}T_v P_{gv} \tag{6}$$

Clearly, peer group eligible fraction in treated villages, $T_v P_{gv}$, is a predictor of peer group average school attendance, provided the individual effect of the policy is not zero, i.e. $\delta \neq 0$. Interestingly, the peer group response to the eligible fraction, $\delta(1-\gamma)^{-1}$ is stronger than the individual effect of the program, δ, if positive endogenous social interactions are present, i.e. $1 > \gamma > 0$. This is because endogenous social interactions amplify the effects of policies that affect individuals.

Inserting peer group average school attendance as given in equation (6) into the structural model (2) of school attendance yields

$$S_{igv} = \frac{\alpha_{gv}}{1-\gamma} + \frac{\gamma\eta+\beta}{1-\gamma}P_{gv} + \frac{\gamma\delta}{1-\gamma}T_v P_{gv} + \delta T_v P_{igv} + \eta P_{igv} + \epsilon_{igv} \tag{7}$$

Equation (7) shows that children will go to school more frequently when many children in their peer group are eligible for PROGRESA cash transfers. This effect is captured by the positive reduced form coefficient on the regressor "eligible share in treated villages" $T_v P_{gv}$. Thus, equation (7) suggests analyzing the spillover effect of PROGRESA on ineligible children as a function of the fraction of peers eligible for transfers.

Equations (6) and (7) are the reduced form equations for the "IV2 identification strategy" that uses peer group eligible fraction in treated village as an instrument for peer group average school attendance. The social interactions parameter is identified from the ratio of the coefficient on $T_v P_{gv}$ in the individual model of school attendance (7) to the corresponding coefficient in the model for peer group average school attendance (6). The peer group eligible fraction

[28] Essentially, peer groups differ with respect to the eligible fraction because PROGRESA applied a state poverty line to a set of villages which differ with respect to their location relative to the poverty line.

in treated villages is a valid instrument because it is a predictor of peer group average school attendance and it is excluded from the structural model (2).

Is it valid to exclude peer group eligible fraction in treated villages? It is useful to discuss validity of the central exclusion restriction in two steps: validity without the program (concerns with omitted variables), and validity with the program (concerns with exclusion of $T_v P_{gv}$ from the structural model (2)).

In a first step, we discuss the validity of exclusion supposing that the program had not been implemented. In this situation, the peer group's eligible fraction can clearly be excluded from the structural model due to randomized assignment of the program.[29] Randomized assignment implies that treated and control villages will have the same exogenous social interactions and that unobserved and observed determinants of school attendance are correlated in the same way with the eligible fraction in the peer group. Interestingly, it is possible to provide evidence on the correlation between school attendance and the eligible fraction of the peer group *before* the program was introduced.

Figure 1 about here

Figure 1 shows kernel regression estimates of the probability of attending school as a function of the eligible share in each child's peer group, separately for treated and control villages. Interestingly, results indicate that the probability of attending school does not change as the eligible share in the peer group increases. Moreover, the probability of attending school varies in the same way in treated and control villages (Table A2). This evidence shows that school attendance is balanced conditional on the eligible fraction in the peer group *before* the program.

Are children also comparable in treated villages and control villages with respect to observed background characteristics?[30] We can address this question either by testing each background characteristic separately, or by forming a linear index of background characteristics. A meaningful linear index of background characteristics can be constructed in two steps. In the first step, we fit a linear model of the change in school attendance for ineligible children in control villages. In the second step, we use this model to predict the change in school attendance for ineligible children in treated and control villages. This procedure creates an index of observed

[29] The eligible share in the peer group might be endogenous because individuals move. This concern is not warranted because we rely on the eligible fraction in the peer group as it is observed in October 1997, i.e. 9 months before the program was implemented. The pre-program composition of peer groups can not be affected by relocation before the treatment status of the village was determined. Note that using the pre-program (October 1997) eligible fraction rather than the during-program (October 1998) eligible fraction leads to a weaker instrument because the pre-program eligible fraction proxies for during-program eligible fraction. Arguably, the advantage in terms of the pre-program eligible fraction being exogenous greatly outweighs this cost.

[30] These characteristics are grade completed in 1997, gender, mother education, father education, father missing, floor type, roof type, and secondary school in village.

characteristics where elements of this index are weighted with respect to their relevance in explaining school attendance trends. Figure 2 shows predicted change in school attendance, separately for villages with PROGRESA and control villages.

Figure 2 about here

Results in Figure 2 show that children in villages without PROGRESA are expected to leave school more likely in grades with a high eligible fraction compared to villages with a low eligible fraction. Whereas about 5 percent of children in the sample leave school in a grade with no poor children, the corresponding number is 10 percent in grades with a 100 percent share of eligible children. The picture is quite similar for children in treated villages. The probability of leaving school is predicted to be lower in grades with a low eligible share compared to grades with a high eligible share. A formal test of the comparability of the two groups of children consists in regressing predicted school attendance on the eligible fraction, a dummy for treated village, and the interaction term eligible fraction in treated village. Results indicate that neither the interaction term nor the dummy for treated villages is significantly different from zero (Table A2). This result suggests that children from treated and control villages are comparable with respect to their observed characteristics.

In a second step, we discuss validity of excluding the peer group's eligible fraction with PROGRESA. Exclusion may fail for three reasons. First, the direct effect of the program (δ) may vary across peer groups in ways which are correlated with the eligible fraction in the peer group. We address this by focusing on non-poor children only, who are not receiving conditional cash transfers.[31]

Second, children in treated villages may tend to go to school more frequently because their village is part of the PROGRESA program. Recall that the program does not only have an education but also a health and nutrition component which may spill over to all inhabitants of a village. Thus, villages with PROGRESA may experience an increase in school attendance that can not be explained by endogenous social interactions. This concern can not be addressed directly in the IV1 identification strategy because this strategy uses the treatment status of the village as an instrument. Interestingly, because the "IV2" identification only uses the eligible share in treated villages, we can test directly whether there is a general spillover effect

[31] This means that the peer group average reduced form equation (6) remains the same but the individual reduced form equation (7) is

$$S_{igv} = \frac{\alpha_{gv}}{1-\gamma} + \frac{\gamma\eta+\beta}{1-\gamma}P_{gv} + \frac{\gamma\delta}{1-\gamma}T_v P_{gv} + \epsilon_{igv} \tag{8}$$

because there is no variation in P_{igv}.

by adding the treatment status of the village, T_v, as an additional regressor in the structural model (2). This is an important advantage of the IV2 analysis compared to the IV1 analysis used in the existing literature.

The third reason why exclusion may fail is that income pooling is more likely to arise in villages where many children are eligible for PROGRESA transfers. The idea is that the program is injecting more cash into villages with many eligible children. This suggests that alternative mechanisms for spillover effects could also be more relevant in these villages. It is, again, not possible to address this issue in the IV1 strategy. This concern can be addressed by exploiting the restriction that peer effects are grade specific.[32] The IV2 strategy exploits an instrument, the eligible share in treated villages, that varies within villages. It is therefore possible to identify social interactions from within village variation in the peer group eligible fraction, i.e. by comparing ineligible children who are living in the *same village* but who are enrolled in different grades.This is an IV2 analysis that uses deviations from the village mean of the peer groups eligible share as an instrument for peer group average schooling. The deviation of village mean eligible fraction is orthogonal to village specific shocks but still correlated with peer group average schooling.[33]

4.3 Different Social Interactions Parameters

The implicit assumption in equation (2) is that the social interactions are equally important among the poor and the non-poor. This assumption can be justified by in-depth studies of community social interactions (Adato, 2000). Nevertheless, is the IV2 identification strategy valid if the social interactions parameter γ is group specific? An extension of the basic model (2) of social interactions allows for separate social interactions parameters for the poor and the non-poor, i.e. γ_P and γ_N. Interestingly, even if the social interactions parameter differs between poor and non-poor children, an IV2 strategy that identifies social interactions in the sub-sample of non-poor children will measure γ_N (see appendix).

However, note that identification fails if there are heterogeneous social interaction parameters within the sample. For instance, suppose that boys react to peer group average schooling whereas girls do not. Heterogeneity in the social interactions parameter introduces a specification error into the error term of model (2). Regression analysis only provides an estimate of

[32] We thank a referee for this suggestion.

[33] The problem with this analysis is that most of the variance in the eligible share arises across villages rather than within villages. This means that the precision of the IV2 analysis will be greatly reduced. We address lack of precision using the program status of the village as an additional instrument. An alternative to the "deviation from village mean approach" is to add village fixed effects to model (2) as in Lalive and Cattaneo (2006). The main advantage of the "deviation from village mean" approach is that it allows sharpening the precision of the instrument.

the mean social interactions parameter if treatment response heterogeneity is orthogonal to the instrument used. Below, we address heterogeneous social interactions by analyzing separately, social interactions for girls, for boys, and for children in lower and upper grades of primary school finding no support for the hypothesis of heterogeneous social interactions.

4.4 Identifying the Direct Effect

The crucial missing part of the discussion is how to identify the direct effect of the program. The direct effect can be measured in two ways. The first strategy identifies the direct effect by estimating model (2) pooling data on children in poor households and non-poor households. Equation (6) and (7) are the reduced form equations defining an IV2 strategy in the overall sample. This identification strategy is implicitly comparing children in eligible households to their classroom peers in households that are not eligible. This contrast identifies the direct effect of the program since the social interactions component of schooling choices is held constant.

A second way to identify the direct effect is within peer groups. Subtracting peer group average schooling choices from individual schooling choices allows measuring the direct effect of the program since social interactions are held constant. Specifically, subtracting group average schooling from the structural model (2) shows that[34]

$$S_{igv} - S_{gv} = \eta(P_{igv} - P_{gv}) + \delta T_v(P_{igv} - P_{gv}) + \epsilon_{igv} \tag{9}$$

This identification strategy measures the direct effect controlling for social interactions by subtracting peer group average schooling.[35]

Note that the IV2 identification strategy is valid if there is no correlation between the direct effect of PROGRESA and the eligible fraction in the peer group and if the social interactions parameter does not vary across individuals. In contrast, the within peer group identification strategy is not sensitive to violation of the assumption of homogeneous effects of the program or homogenous social interactions.

We can use information on the direct effect of the program on an individual child, δ, to assess the magnitude of the social multiplier. The idea is to relate the peer group level "treatment" effect that we get when regressing peer group average school attendance on the peer group's

[34] Note that we obtain this expression by subtracting from the structural model (2) the model for peer group average schooling before solving for the social equilibrium, i.e. $S_{gv} = \alpha_g + (\beta+\eta)P_{gv} + \gamma S_{gv} + \delta T_v P_{gv}$. Subtracting peer group average schooling from individual schooling clearly eliminates all peer group effects.

[35] An alternative procedure is to identify social interactions by adding peer group fixed effects (see Lalive and Cattaneo (2006) or Graham and Hahn (2005)). The peer group fixed effects identification strategy is an approximation to the strategy reported here since peer group average schooling is not constant within peer groups.

eligible fraction to the individual level effect of the program. The effect of the program on the peer group average schooling is $\delta(1-\gamma)^{-1}$. The effect of the program on individual schooling is δ. This means that the ratio of the peer group level effect of PROGRESA to the individual level effect of PROGRESA is the social multiplier.

Information on the direct effect of PROGRESA and on the importance of social interactions can be combined to decompose the total effect of the program on eligible children (Table 3, Column A). The total effect is

$$\begin{aligned} \pi_1^P &= \delta + \gamma * (E[S_{gv}|T_v = 1, P_{igv} = 1] - E[S_{gv}|T_v = 0, P_{igv} = 1]) \\ &= \delta \left(1 + \frac{\gamma}{1-\gamma} E[P_{gv}|T_v = 1, P_{igv} = 1]\right) \end{aligned} \tag{10}$$

This contrast therefore captures both, the direct effect of PROGRESA that is due to transfers as well as the social effect of PROGRESA that arises because the program changes the average change in school attendance of other children in the same grade. Thus, comparing the direct effect of PROGRESA with the contrast (10) is informative on the relative importance of the individual and social determinants of school attendance decisions.

Equation (10) also highlights why information on the relevance of social interactions – the parameter γ – is important for policy design. Granting access to the transfer for only one household in each village ($E(P_{gv}|T_v = 1, P_{igv} = 1) = 0$) will generate the effect δ among eligible children. In contrast, granting access to the program to all households within the village ($E(P_{gv}|T_v = 1, P_{igv} = 1) = 1$) generates an additional social effect because the social environment of an individual changes. This additional effect is equal to $\gamma(1-\gamma)^{-1}$ or the social interactions parameter times the social multiplier. The intuition for this effect is that the initial increase in schooling for eligible children δ increases peer group average schooling by δ if everyone is eligible for treatment. Because everybody in the peer group is now more likely to be in school, there is an indirect effect on eligible children's school attendance decisions on the order of $\gamma\delta$. This indirect effect in turn generates an additional indirect effect on peer group average schooling of magnitude $\gamma^2\delta$, etc. The resulting total effect on peer group average schooling is $\delta(1-\gamma)^{-1}$ – the direct effect amplified by the social multiplier. The resulting effect of this increase in peer group average schooling is $\gamma\delta(1-\gamma)^{-1}$ as given in equation (10).

5 Results

This section first reports our main estimates of the social interactions parameter. The section then discusses the sensitivity of the main estimates with respect to important concerns with identification. The section closes by analyzing the relative importance of financial incentives and social interactions in shaping school attendance decisions.

5.1 Main Result on Social Interactions

Table 4 reports the main results of the empirical analysis.[36] The first Column in Table 4 provides the estimate of π_1^N (equation 1) – the reduced form effect of the PROGRESA status of the village on school attendance of children whose families are not part of the program. The analysis indicates that school attendance drops by 2.1 percentage points less strongly among non-poor children whose peers have access to PROGRESA compared to similar children in villages that were excluded from PROGRESA.

Table 4 about here

The second Column in Table 4 reports the effect of PROGRESA on peer group average change in school attendance (equation (5)) – the reduced form effect of the instrument on the endogenous regressor. The cash transfer increases the trend in school attendance among peers by 3.9 percentage points. This effect is statistically significant and quantitatively important. Taken together, the evidence in Table 4 suggests that a 3.9 percentage points increase in peer's school attendance trend generates a 2.1 percentage points increase in ineligible students' school attendance. According to equation (4), this is a social interactions parameter of .534.

Indeed, the IV1 strategy in Column 3 of Table 4 that identifies social interactions from the ratio of the individual reduced form effect and the peer group average effect finds that γ is .534. The parameter estimate is significantly different from zero at the 5 % level, and quantitatively important. A 10 percentage point increase in peer group school attendance leads to a 5 percentage point increase in individual school attendance. The IV1 strategy therefore provides strong evidence for social interactions.[37]

Recall, however, that the IV1 strategy is not robust to village specific shocks. The IV2 analysis can address the relevance of these shocks. Before we turn to the estimates of the social interactions parameter we discuss at a descriptive level whether the eligible fraction in the peer

[36] Note that all estimates control for the full set of control variables shown in Table 3. Inference is based on robust standard errors that allow for clustering at the village level. The analysis is confined to the 5,143 children for whom we find at least one other classmate within the village. See Table A3 for detailed results.

[37] Note that the Cragg and Donald (1993) weak identification statistic is larger than the Stock and Yogo (2005) critical value at 10 % maximal size. This means that we can reject the null hypothesis of weak instruments.

group predicts the change in school attendance. The central prediction from the identification analysis in section 4 is that the spillover effect among the poor should be stronger the higher is the eligible fraction in the peer group (equation (7)). Figure 3 investigates the spillover effect among non-poor children conditional on the eligible fraction in the peer group.[38] Figure 3 shows that school exits are more frequent for children in villages without PROGRESA whose classroom is characterized by a high eligible fraction. In contrast, school attendance even appears to increase slightly in villages with PROGRESA. This evidence is consistent with a positive reduced form effect of the eligible fraction in program villages.

Figure 3 about here

The Columns D, E, and F in table 4 show the results of the IV2 analysis. Column D reports estimates of the response of non-poor children to introducing the program in their peer group (equation 7). Findings indicate that individual schooling increases as the eligible fraction in their peer group increases. Individual school attendance increases as much as 4.1 percentage points due to an increase in the eligible fraction from 0 to 1 in villages with PROGRESA. This is in line with equation 7. Column E reports the peer group average schooling equation (6). Results indicate that peer group average schooling increases as the eligible fraction increases. Importantly, the effect of PROGRESA on peer group average schooling is almost twice as strong as the spillover effect of PROGRESA on individual school attendance. Column F shows the IV2 estimates of the social interactions parameter γ. Findings indicate that individual schooling increases very strongly as peer group average schooling increases. A 10 percentage point increase in individual schooling is estimated to increase individual schooling by 5.4 percentage points. The point estimate of the IV2 effect is very much in line with the point estimate of the IV1 effect. However, the IV2 effect is estimated more precisely than the IV1 effect. The IV2 effect is significantly different from zero at the 1 percent level whereas the IV1 effect is not significantly different from zero at the same level.[39]

5.2 Sensitivity Analysis on Social Interactions

Table 5 discusses concerns with the identification strategy. Column A in Table 5 reproduces the baseline result from Table 4, Column F.

[38] Specifically, in preparing the evidence in Figure 1 we regress the change in school attendance S_i among ineligible children on the eligible fraction in the peer group P_{ig} (excluding individual i), separately for children living in villages with the program and for children in villages without the program. The non-parametric kernel regressions use an Epanechikov kernel with bandwidth of .5.

[39] The Cragg and Donald (1993) weak identification statistic indicates that the predictive power of the eligible fraction in treated villages is, indeed, considerably stronger than the predictive power of the village with program indicator. This explains the increased precision of the IV2 estimate.

Table 5 about here

The first concern is that PROGRESA might have generated more traditional health or nutritional spillover effects on all families in PROGRESA villages. Interestingly, we can test for a PROGRESA wide spillover in the IV2 analysis but not in the IV1 identification strategy. The idea is to add the indicator village with PROGRESA, T_v, to the set of regressors in model (2). Column B in Table 5 reports the resulting estimate of the social interactions parameter as well as general PROGRESA village effects. Results indicate that the social interactions parameter point estimate is 0.572 (significantly different from zero) which is very much in line with the baseline estimate of 0.542. Moreover, there is no evidence of a spillover effect affecting all PROGRESA villages.

The second important concern with the identification strategy IV2 is that eligible families may share PROGRESA transfers or school resources (such as books, school uniforms, etc.) with ineligible families. This concern differs from the first concern because such sharing will be absent in villages with few eligible families but sharing will be strong in villages with many eligible families. We address non-social spillover effects by using the deviation of the eligible share in a classroom from the village mean as an instrument. This instrument is orthogonal to village specific shocks by construction but it still allows identifying grade specific social interactions. Column C in Table 5 reports the results we get when we identify using within village variation. The resulting point estimate of the social interactions parameter is 0.599 which is qualitatively similar as the baseline estimate. The downside to the within village identification strategy is that there is little within village variation in the eligible fraction of a grade, leading to a very imprecise estimate of the social interactions parameter.

One way to improve precision is to add the PROGRESA status of the village as a second instrument. Use of this second instrument can improve the precision of the estimate if the PROGRESA status of the village predicts peer group average school attendance. Moreover, because the model is overidentified, we can assess the validity of the instruments using Hansen (1982)'s test of over-identifying restrictions. Column D in Table 5 reports the result. The over-identification test does not reject the null of valid instruments. The point estimate of the social interactions parameter is 0.541 – almost precisely the same estimate we get in the baseline analysis. Moreover, the precision of the estimate improves greatly allowing us to reject the null hypothesis of absent social interactions at the 5 percent level of significance.

Taken together, these findings indicate that village specific responses to PROGRESA can not explain the spillover effect on ineligible children. Indeed, we can corroborate this interpretation using direct measures of consumption. Comparing ineligible (non-poor) household's

monthly adult equivalent consumption in treated villages to control villages we do not find any evidence of a consumption spillover effect.[40]

A further concern with the baseline analysis is that children in ineligible households may anticipate extension of program eligibility criteria that took place in the second year of the program. Moreover, it is possible that the marginally ineligible families within villages were not aware of the fact that they are not eligible for cash transfers. Recall that eligibility to PROGRESA transfers was based on a poverty index. We therefore discard about 25 % of all households that are closest to the poverty threshold ending up with 3809 children instead of 5143 children. These households are very close to the poverty threshold and may have expected to be eligible for the program more strongly than the other households. The idea is that excluding these children from the analysis will lead to a reduced estimate of the social interactions parameter if anticipation effects are important. Results suggest, however, that anticipation effects are not driving results (Column E in Table 5). The analysis identifies a social interactions parameter of .596 with standard error .249. These results are very much in line with the baseline estimates.

Column F in Table 5 focuses on children who have progressed through primary school without repeating a grade – who are on grade level. There are two reasons for this analysis. We first find that the higher rate of school attendance in October 1997 – before the program started – in villages with program compared to villages without the program arises among children who are not on grade level. In the grade level sample, the difference in school attendance between treated and control villages is -.7 percentage points. This compares favorably with the overall difference of 2.4 percentage points in the entire sample that we report in Table 2, Column B (first row). Furthermore, focusing on the grade level sample reduces the heterogeneity of children in terms of age. Findings in Column G indicate that the social interactions parameter is .554 (with standard error .229) which is very much in line with the baseline result of .546.

So far the empirical analysis has assumed that the social interactions parameter γ is identical across individuals. This assumption motivates a standard constant coefficient regression model for identification. Table 6 explores the extent to which this assumption is true along the two important dimensions gender and grade level. Panel A in Table 6 reports the social interactions coefficient by completed grade level. Results indicate that social interactions are

[40] See Angelucci and De Giorgi (2006) for further analyzes of the effects of PROGRESA on consumption. We also investigate whether adding controls for consumption affects our estimate of the social interactions parameter. We find that consumption is an important predictor of school attendance trends. Households with higher consumption also see their children leaving school to a lesser extent. However, adding total expenditure to the regression does not affect the result concerning the social interactions parameter (estimate is 0.554 with standard error 0.213). We also investigate the sensitivity of our results to adding regional fixed effects at the municipality level. These results indicate that the social interactions parameter is 0.560 (with standard error 0.274).

significantly different from zero and almost equally important for children in primary school (grade level 3-4) and children transiting progressing from primary school to secondary school (grade level 5-6). This finding is important in the sense that social interactions not only affect the drop out decision (captured in grades 5 and 6) but also the regular attendance decision (captured in grades 3 and 4).

Table 6 about here

Gender is a second important dimension for two reasons. First, the labor market opportunities are expected to differ strongly between boys and girls. Second, the secondary school PROGRESA transfer is higher for girls than for boys. Panel B in Table 6 therefore provides separate results for girls and for boys. Results for girls suggest that the social interactions parameter point estimate is slightly lower than in the baseline estimate but it is significantly different from zero. Results for boys are significantly different from zero and slightly larger than the baseline estimate. Thus, results by gender do not suggest any important differences in the extent to which social forces affect human capital decisions.

To complete our analysis of the effects of PROGRESA on schooling decisions, we need to identify the direct effect of PROGRESA.

5.3 The Direct Effect of PROGRESA

Section 4 discusses that there are two ways to identify the direct effect of PROGRESA on eligible children. The first identification strategy is the IV2 identification strategy applied to the full sample of eligible and ineligible children. The second identification strategy is the within peer group strategy that differences out social interactions. Table 7 reports the resulting two sets of estimates of the direct effect of PROGRESA transfers.

Table 7 about here

Column A of Table 7 shows the estimates of the reduced form individual equation (7, for all children). Results indicate that there is a 3.5 percentage point increase in school attendance that is specific to eligible children in villages with PROGRESA. Moreover, there is also an effect of the eligible fraction in the peer group on individual school attendance. Increasing the eligible share in the peer group from 0 to 1 will increase individual school attendance by 3.1 percent.

Column B in Table 7 shows the estimates of equation (6). Results indicate that increasing the eligible fraction in the peer group from 0 to 1 increases school attendance by 6.2 percentage points.

Column C in Table 7 reports the IV2 estimates of model 2 that identify the social interactions parameter from the ratio of the individual to the peer group response to peer group composition. Results regarding the social interactions parameter indicate that a 10 percentage point increase in peer group school attendance raises individual school attendance by 5 percentage points. Note that the point estimate of the social interactions parameter γ is almost identical to the point estimate in the ineligible children sub-sample (Table 4). Results regarding the direct effect of the policy indicate that individual school attendance increases by 3.3 percentage points due to conditional cash transfers.

Column D in Table 7 shows results based on the "within peer group" approach (section 4). This identification strategy regresses the difference between individual and peer group average school attendance on the difference between the individual eligibility indicator and the eligible fraction in the peer group and control variables as in model (9). Results indicate that the direct effect of PROGRESA results in raising individual school attendance by 3.1 percentage points. This effect is statistically different from zero and slightly smaller than the effect implied by the IV2 identification strategy in Column C of Table 7. This suggests that heterogeneity in the direct treatment effect is not important.

The crucial implication of endogenous social interactions is that the effects of individual policies get amplified at the group level. Table 7 allows discussing amplification. We find that PROGRESA increases individual school attendance by 3.1 percentage points. This individual effect would translate into differences of the same order of magnitude across classrooms. Specifically, suppose we compare a classroom in which all children are eligible to the program to a classroom with not eligible child. The individual level analysis would lead us to expect that school attendance rates between these classrooms would be expected to differ by 3.1 percentage points. Table 7 Column B shows that this is not the case. Increasing the eligible fraction from 0 to 1 leads to an increase by 6.2 percentage points. This means that the difference in school attendance rates between eligible and ineligible classrooms is twice as large as the difference in school attendance rates between eligible and ineligible children – the social multiplier is 2.

We close this discussion by decomposing the total effect of PROGRESA on children who live in eligible families into a direct effect that arises due to the financial incentive and an indirect effect arising due to the effect of PROGRESA on the social environment of the children (Table 8).

Table 8 about here

The direct effect of the program on eligible children is 3.1 percentage points (Table 7). The indirect effect is the product of the social interactions parameter (.505) with the peer

group response to PROGRESA (equation 3). The peer group response to PROGRESA can be identified contrasting average school attendance in treated and control villages (equation (5) for the poor). It turns out that PROGRESA led to an increase in peer group average school attendance of about 4.8 percentage points. Thus, the total predicted effect of the PROGRESA program on eligible children is 5.5 percentage points. This shows that the direct effect of the monetary transfer due to PROGRESA is roughly of equal size as the indirect effect of PROGRESA via the change in peer group decisions.

Moreover, Table 8 shows that the indirect effect of PROGRESA is larger for eligible children than for ineligible children. This is due to the fact that the typical peer group of an eligible child is treated at a much larger rate (73 % treated) than the typical peer group of an ineligible child (55 % treated) leading to a stronger peer group impact of PROGRESA for eligible children.[41] This highlights the main contribution of our results to the literature concerned with targeted interventions.[42] If social interactions are relevant, targeting is an important determinant of the average effect of the program among eligible individuals. Thus, a program that reaches the same number of individuals will generate a different effect on treated individuals when spread out over individuals who are not connected to each other compared with a strategy of strong concentration of the program within tightly defined groups rather than individuals. Indeed, we find that the total treatment effect of PROGRESA is 0.043 for children whose classroom is minority eligible (eligible fraction < 0.5) whereas the total effect of PROGRESA is 0.059 for children in classrooms that are majority eligible (eligible fraction ≥ 0.5).[43]

6 Conclusions

Individual schooling decisions may be affected by peer group schooling decisions. This paper studies the empirical relevance of social interactions in the context of a randomized experiment that grants a cash transfer to poor children who are attending school regularly. This targeted subsidy encourages eligible students to remain in school. Moreover, if social interactions are important, the ineligible students may also decide to acquire more schooling.

Results indeed indicate first that eligible children attend school more regularly when their families are paid to send them to school. Second, we find that ineligible children's school enrollment also increases, especially so if many of their peers – children in the same grade living in the same village – are eligible for the cash subsidy. Third, we find that a child will

[41] Note that the social spillovers predicted by equation (10) $\gamma(1-\gamma)^{-1}E(P_{ig}|T_i = 1, P_i)$ are .025 for children in eligible families, and .019 for children living in ineligible households. These estimates are in line with the predictions in Table 8.

[42] See Behrman *et al.* (1999) on a discussion of targeting issues in the PROGRESA context.

[43] Results are available upon request.

attend school 5 percentage points more likely if school attendance in the peer group increases by 10 percentage points. Fourth, we identify the direct effect of the cash subsidy program on the eligible students. Comparing this direct effect of the program with the indirect effect of the program due to social interactions, we find that both are equally relevant among eligible children. Fifth, we document strong amplification of the effects of individual policies due to endogenous social interactions. Taken together, these findings suggest that social interactions are empirically relevant for schooling decisions.

There are several reasons why social interactions may be relevant. First, as the opening quote suggests, there may be important complementarities in school attendance. For example, transportation costs (in terms of money and safety) may be lower when children go to school together or children might enjoy spending time with other children. Second, parents who are deciding about sending their child to school face uncertainty regarding the costs, benefits, and ability of their children. Parents may thus consider other similar parents' decisions to learn about the optimal choice for their child (Manski, 2004). Third, ineligible parents may also anticipate future admission of their child to the program. Discriminating between these competing explanations for social interactions in schooling is beyond the scope of this paper but a worthwhile topic for future research.

Regardless of the exact nature of social interactions, knowledge on the strength of social interactions is important for at least two reasons. If social interactions are present, targeting of the policy becomes important. We document that the overall effect of a targeted intervention on eligible children is stronger if the policy targets all children in a few peer groups rather than just a few children in many peer groups. Second, we show that the individual effects of targeted interventions are amplified due to endogenous social interactions. Amplification of the effects of targeted policies can be important for policy because it leads to strong inequality in schooling outcomes for observationally similar peer groups.

References

Adato, M. (2000). The Impact of PROGRESA on Community Social Relationships. Report, IFPRI.

Adato, M., Coady, D., and Ruel, M. (2000). An operations evaluation of PROGRESA from the perspective of beneficiaries, promotoras, school directors, and health staff. Report, IFPRI.

Angelucci, M. and De Giorgi, G. (2006). Indirect Effects of an Aid Program: The Case of Progresa and Consumption. Discussion paper 1955, IZA.

Angrist, J. D. and Lang, K. (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, **94**(5), 1613–1634.

Attanasio, O., Meghir, C., and Szekely, M. (2003). Using Randomised Experiments And Structural Models For 'Scaling Up': Evidence From The PROGRESA Evaluation. Working paper no ewp03/05, Edepo.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. NY: Columbia University Press.

Behrman, J. R., Davis, B., and Skoufias, E. (1999). An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico. Report, IFPRI.

Behrman, J. R., Sengupta, P., and Todd, P. E. (2001). Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment. Working paper, University of Pennsylvania.

Betts, J. R. and Zau, A. (2004). Peer Groups and Academic Achievement: Panel Evidence from Administrative Data. Working paper, University of California.

Bobonis, G. J. (2004). Income Transfers, Marital Dissolution and Intra-Household Resource Allocation: Evidence from Rural Mexico. Working paper, University of California at Berkeley.

Bobonis, G. J. and Finan, F. S. (2005). Endoenous Peer Effects in School Participation. Working paper, UC at Berkeley.

Brock, W. A. and Durlauf, S. N. (2001). Interactions-based models. In J. Heckman and E. Leamer, editors, *Handbook of Econometrics, vol 5*, pages 3297–3380. North Holland: Amsterdam.

Buddelmeyer, H. and Skoufias, E. (2003). An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA. Working paper, University of Melbourne.

Card, D. E. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter and D. E. Card, editors, *Handbook of Labor Economics, Vol. 3*, pages 1801– 1863. North-Holland.

Case, A. C. and Katz, L. F. (1991). The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Youth. Working paper 3708, NBER.

Cipollone, P. and Rosolia, A. (2007). Social Interactions in High School: Lessons from an Earth Quake. *American Economic Review*, page forthcoming.

Coleman, J. S. (1961). *The Adolescent Society: The Social Life of the Teenagers and Its Impact on Education*. NY: The Free Press.

Cragg, J. and Donald, S. (1993). Testing Identfiability and Specification in Instrumental Variables Models. *Econometric Theory*, **9**, 222–240.

Duflo, E. and Saez, E. (2003). The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *Quarterly Journal of Economics*, **118**(3), 815–842.

Gertler, P. (2000). The Impact of Progresa on Health. Final Report, IFPRI.

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

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and Social Interactions. *Quarterly Journal of Economics*, **111**(2), 507–548.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (2003). The Social Multiplier. *Journal of the European Economic Association*, **1**(1-2), 345–353.

Graham, B. S. and Hahn, J. (2005). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters*, **88**, 1–6.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(3), 1029–1054.

Hanson, G. E. (2002). Education and Educational Attainment in Mexico. Working paper, World Bank.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2001). Does Peer Ability Affect Student Achievement. Working paper 8502, NBER.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. Working paper 7867, NBER.

Lalive, R. (2003). Social Interactions in Unemployment. Discussion paper 803, IZA.

Lalive, R. and Cattaneo, A. (2006). Social Interactions and Schooling Decisions. Working paper no. 1787, CESifo.

Manski, C. F. (1993). Identification of Endogenous Social Interactions: The Reflection Problem. *Review of Economic Studies*, **60**(3), 531–542.

Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

Manski, C. F. (2000). Economic Analysis of Social Interactions. *Journal of Economic Perspectives*, **14**(3), 115–136.

Manski, C. F. (2004). Social Learning from Private Experiences: The Dynamics of the Selection Problem. *Review of Economic Studies*, **71**(2), 443–458.

Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, **72**(1), 159–217.

Miguel, E., Kremer, M., and Thornton, R. (2004). Incentives to Learn. Working paper, Harvard University.

Moffitt, R. A. (2001). Policy Interventions, Low Level Equilibria, and Social Interactions. In S. N. Durlauf and H. P. Young, editors, *Social Dynamics*, pages 45–82. Cambridge, Massachusetts: MIT Press.

OECD (2004). *Education at a Glance 2004 – Mexico*. Paris: OECD.

Sacerdote, B. (2001). Peer Effects with Random Assignment. *Quarterly Journal of Economics*, **116**(2), 681–704.

Skoufias, E. (2001). PROGRESA and its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI. Report, IFPRI.

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. Andrews and J. H. Stock, editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, page 80108. Cambridge: Cambridge University Press.

Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment. *Review of Economic Studies*, **68**(2), 261–295.

Topa, G. and Conley, T. G. (2002). Socio-Economic Distance and Spatial Patterns in Unemployment. *Journal of Applied Econometrics*, **17**(4), 303–327.

Wald, A. (1940). The Fitting of Straight Lines in Both Variables are Subject to Error. *Annals of Mathematical Statistics*, **11**(3), 284–300.

Appendix: Heterogeneous Social Interactions

Suppose that the structural model of schooling is

$$S_{igv} = \alpha_{gv} + \beta P_{gv} + \gamma_N (1 - P_{igv}) S_{gv} + \gamma_P P_{igv} S_{gv} + \delta T_v P_{igv} + \eta P_{igv} + \epsilon_{igv} \tag{11}$$

where γ_N is the social interactions parameter for the non-poor and γ_P is the social interactions parameter for the poor. Peer group average school attendance is

$$S_{gv} = \frac{\alpha_{gv}}{1 - \tilde{\gamma}} + \frac{\beta + \eta}{1 - \tilde{\gamma}} P_{gv} + \frac{\delta}{1 - \tilde{\gamma}} T_v P_{gv} \tag{12}$$

where $\tilde{\gamma} = \gamma_N + (\gamma_P - \gamma_N) P_{gv}$ is the average social interactions parameter in a group with a poor share of P_{gv}. This shows that the effect of the peer group eligible fraction on peer group average schooling is heterogeneous.

The empirical analysis assumes the peer group average response to PROGRESA to be homogeneous

$$S_{gv} = \lambda_0 + \lambda_1 P_{gv} + \lambda_2 T_v P_{gv} + \xi_{gv} \tag{13}$$

Thus, the model in the empirical analysis is characterized by a specification error ξ_{gv}, where ξ_{gv} captures the difference between the estimated model (13) and true model (12). Specification error introduces bias in the estimates of the reduced form parameters $\lambda_0, \lambda_1, \lambda_2$. For instance, the probability limit of the least squares estimator of λ_2 is

$$plim(\hat{\lambda}_2) = \lambda_2 + \frac{Cov(\xi_{gv}, \tilde{T}_v \tilde{P}_{gv})}{Var(\tilde{T}_v \tilde{P}_{gv})}$$

where $\tilde{T}_v \tilde{P}_{gv}$ is the component of the eligible share in treated villages that can not be predicted by the other regressors in the model (13). The key question is whether this bias introduces a bias in the IV estimator of γ_N.

Inserting the estimated model (13) into the structural model of individual choices (11) yields the following reduced form model for non-poor individuals

$$S_{igv} = \alpha_{gv} + \gamma_N \lambda_0 + (\beta + \gamma_N \lambda_1) P_{gv} + \gamma_N \lambda_2 T_v P_{gv} + \epsilon_{igv} + \gamma_N \xi_{gv} \tag{14}$$

The probability limit of the least squares estimator of $\gamma_N \lambda_2$ is

$$plim(\widehat{\gamma_N \lambda_2}) = \gamma_N \left(\lambda_2 + \frac{Cov(\xi, \tilde{T}_v \tilde{P}_{gv})}{Var(\tilde{T}_v \tilde{P}_{gv})} \right)$$

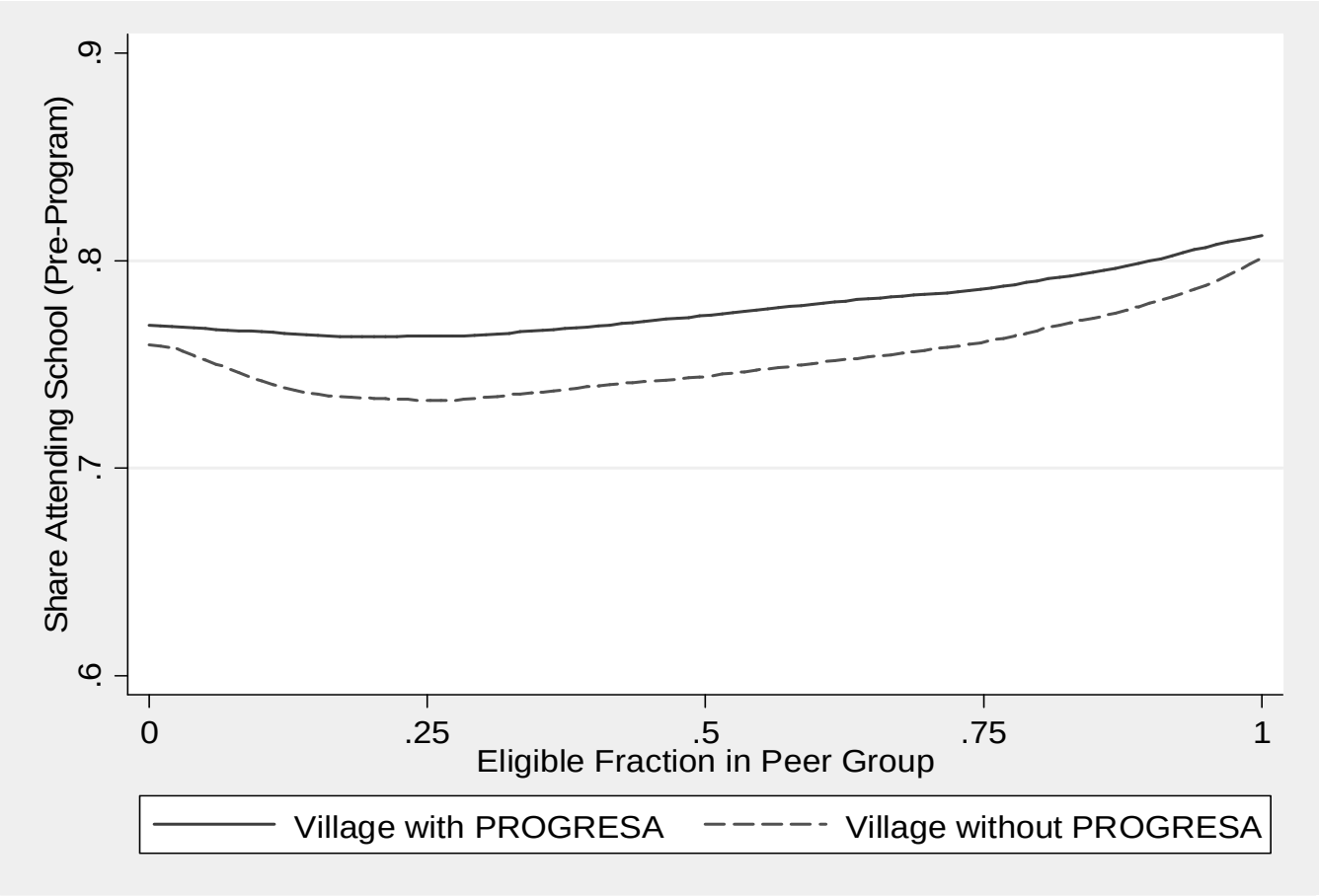
The IV2 estimator identifies γ_N from the ratio of $\widehat{\gamma_N \lambda_2}$ to $\hat{\lambda}_2$. The probability limit of the IV2 estimator is

$$plim\left(\frac{\widehat{\gamma_N \lambda_2}}{\hat{\lambda}_2}\right) = \gamma_N$$

The IV2 estimator is a consistent estimator of γ_N.

Intuitively, heterogeneous peer group response to PROGRESA does generate a specification error. But because this is the same specification error in both reduced form equations, the specification error does not affect the consistency of the IV2 estimator.

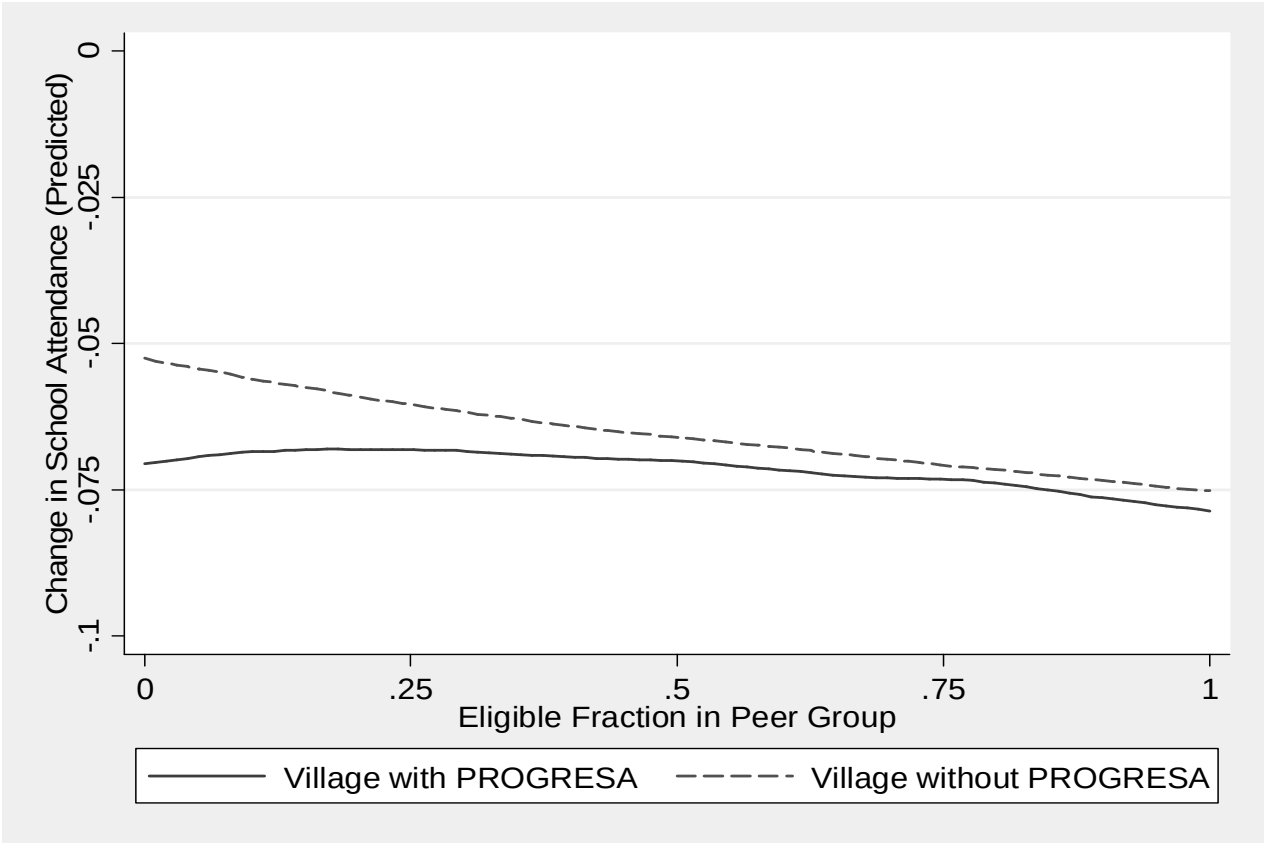
Figure 1:
School Attendance and Eligible Fraction (Before PROGRESA)



Notes: Peer group = all children attending the same grade in the same village. Village with PROGRESA = village with conditional cash transfers for eligible children. Epanechnikov-Kernel Regression (bandwith .5) of school attendance in October 1997 -- before PROGRESA -- on the eligible (=poor) fraction in the peer group. The Regressions show that school attendance changes in the same way with eligible fraction in villages with PROGRESA and villages without PROGRESA.

Source: Own Calculation, Based on Progresa Evaluation Data.

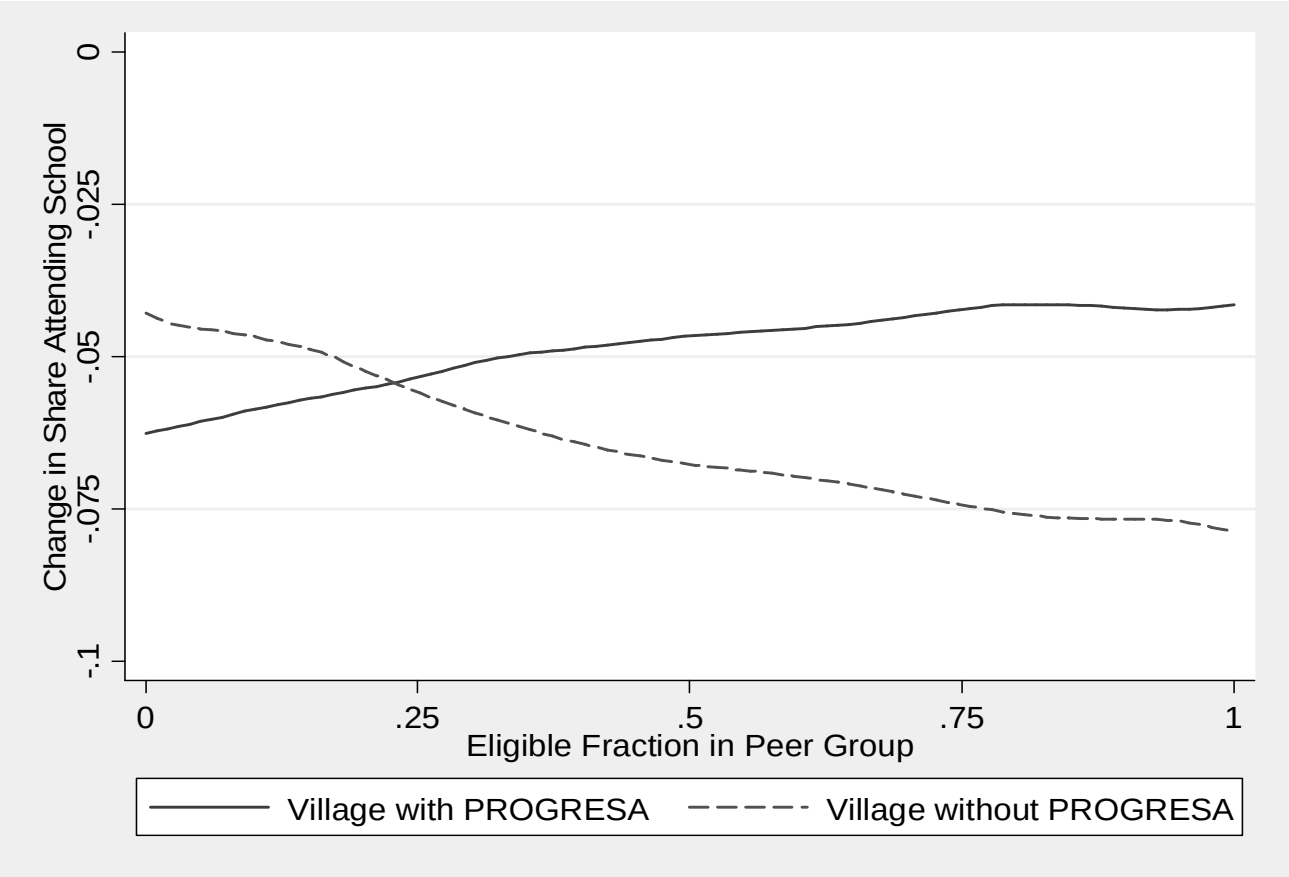
Figure 2:
Change in School Attendance as Predicted from Village without PROGRESA



Notes: Epanechnikov-Kernel Regression (bandwith .5) of predicted change in school attendance on the eligible (=poor) fraction in children attending same grade in same village (=peer group). Change in school attendance predicted in two steps. 1) linear model of school attendance using all control variables listed in Table 1 and fit on ineligible children in villages without PROGRESA. 2) use this model to predict change in school attendance for all ineligible children. The figure shows that the predicted change in school attendance is similar for children in villages with PROGRESA and villages without PROGRESA.

Source: Own Calculation, Based on Progresa Evaluation Data.

Figure 3:
Explaining the Spillover Effect on *Ineligible* Children



Notes: Epanechnikov-Kernel Regression (bandwith .5) of the change in school attendance between October 1998 (During-program) and October 1997 (Pre-program) for children who are not eligible for cash transfers on the eligible fraction in the peer group (other children in same grade and same village). The figure shows that change in school attendance deteriorates for ineligible children living in villages without PROGRESA but improves for ineligible children living in villages with PROGRESA.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 1:

Descriptive Statistics:
Pre-Program Characteristics of Children

	A. Eligible			*B. Ineligible*		
Village with PROGRESA	No	Yes	Difference (abs. z-Value)	No	Yes	Difference (abs. z-Value)
Completed grade 4	0.246	0.235	-0.011	0.227	0.218	-0.009
			(1.25)			(0.78)
Completed grade 5	0.224	0.212	-0.012	0.227	0.239	0.013
			(1.45)			(0.99)
Completed grade 6	0.272	0.282	0.010	0.337	0.337	0.000
			(0.87)			(0.01)
Girl	0.502	0.492	-0.010	0.497	0.507	0.010
			(0.87)			(0.65)
Mother Ed.: Primary School or Higher	0.163	0.160	-0.003	0.217	0.218	0.002
			(0.19)			(0.07)
Father Ed.: Primary School or Higher	0.165	0.169	0.004	0.192	0.202	0.011
			(0.22)			(0.50)
Father Information Missing	0.099	0.095	-0.004	0.117	0.119	0.001
			(0.41)			(0.13)
Floor: Cement	0.262	0.288	0.026	0.626	0.598	-0.027
			(0.92)			(0.86)
Roof: Tin	0.263	0.286	0.023	0.315	0.304	-0.011
			(0.78)			(0.30)
Roof: Asbest	0.129	0.131	0.002	0.167	0.163	-0.004
			(0.11)			(0.12)
Roof: Tiles	0.093	0.127	0.034	0.084	0.092	0.008
			(1.32)			(0.39)
Roof: Cement Blocks	0.106	0.086	-0.020	0.241	0.251	0.010
			(1.15)			(0.31)
Secondary School in Village	0.260	0.256	-0.004	0.340	0.265	-0.075
			(0.07)			(1.15)
Number of Children in Peer Group	12.602	11.931	-0.671	12.992	12.139	-0.853
			(0.52)			(0.51)
Observations	3880	6604		2083	3086	

Notes: Sample refers to children who have completed grades 3 to 6 of primary school in October 1997, who are living with their mother. Eligible means household is classified as poor. Mothers in eligible households receive a cash grant for each child attending school more than 85 % of each quarter in grades 3 to 6 of primary school and grades 1 to 3 of secondary school. All characteristics are measured in October 1997, i.e. abouth 8 months before PROGRESA started. Peer group consists of all children in same grade in same village. Village with PROGRESA are the villages selected for the PROGRESA demonstration in October 1997. Village without PROGRESA is the set of villages who were randomized out of the program.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 2:
Descriptive Evidence on the Effect of Cash Subsidies on School Attendance

	A. Eligible Children			*B. Ineligible Children*		
Village with PROGRESA	No	Yes	Difference (abs. z-Value)	No	Yes	Difference (abs. z-Value)
School Attendance in 1997 (Pre Program)	0.769	0.782	0.013 (0.897)	0.758	0.782	0.024 (1.305)
School Attendance in 1998 (During Program)	0.685	0.757	0.072 (3.798)	0.691	0.733	0.042 (1.907)
Change in School Attendance 1998-1997	-0.084	-0.025	0.059 (5.511)	-0.067	-0.049	0.018 (1.484)
Observations	3880	6604		2083	3086	

Notes: School attendance is the item "Does your child currently attend school?" that was posed to the parents. "Yes" is coded as a value of 1, "No" is coded as a value of 0. October 1997 is 8 months before PROGRESA started, October 1998 is 2 months after PROGRESA started. See notes to Table 1 for definition of the program and eligiblity.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 3:
The Effect of PROGRESA Transfers on Eligible and Ineligible Children
Dependent Variable: Change in School Attendance 1998-1997

	A. Eligible Children	*B. Ineligible Children*
Village with PROGRESA	0.058	0.021
	(0.010)***	(0.012)*
Completed grade 4	-0.018	-0.017
	(0.007)**	(0.012)
Completed grade 5	-0.185	-0.176
	(0.014)***	(0.017)***
Completed grade 6	0.045	0.019
	(0.012)***	(0.014)
Girl	-0.005	-0.008
	(0.008)	(0.010)
Mother Ed.: Primary School or Higher	0.003	0.013
	(0.010)	(0.011)
Father Ed.: Primary School or Higher	0.015	0.033
	(0.010)	(0.013)**
Father Information Missing	-0.043	-0.022
	(0.014)***	(0.015)
Floor: Cement	-0.000	0.004
	(0.010)	(0.013)
Roof: Tin	0.005	0.018
	(0.011)	(0.017)
Roof: Asbest	0.015	0.032
	(0.014)	(0.020)
Roof: Tiles	0.005	0.036
	(0.017)	(0.024)
Roof: Cement Blocks	0.003	0.026
	(0.015)	(0.018)
Secondary School in Village	0.036	0.030
	(0.009)***	(0.014)**
Constant	-0.083	-0.072
	(0.019)***	(0.046)
State Effects (7)	Yes	Yes
Observations	10484	5169

Notes: Robust standard errors in parentheses (allow for clustering at village level). * significant at 10%; ** significant at 5%; *** significant at 1%.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 4:
The Effect of Peer Group Schooling on Individual Schooling
Sample: *Ineligible* Children

	IV1 identification strategy			IV2 identification strategy		
Dependent variable: Change in school Attendance of	Child	Peer Group	Child	Child	Peer Group	Child
Estimation Method	*A. OLS*	*B. OLS*	*C. IV*	*D. OLS*	*E. OLS*	*F. IV*
Change in Peer Group School Attendance	-	-	0.534			0.542
			(0.235)**	-	-	(0.208)***
Village with PROGRESA	0.021	0.039	-	-	-	-
	(0.012)*	(0.010)***				
Village with PROGRESA * Eligible fraction	-	-	-	0.041	0.076	-
				(0.019)**	(0.015)***	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
State Effects (7)	Yes	Yes	Yes	Yes	Yes	Yes
F-Statistic	7.40***	9.59***	-	7.30***	9.78***	-
Cragg-Donald (1993) weak identification statistic	-	-	62.74	-	-	90.23
Observations	5143	5143	5143	5143	5143	5143

Notes: Robust standard errors in parentheses (allow for clustering at village level). * significant at 10%; ** significant at 5%; *** significant at 1%. Peer group consists of all children in same grade in same village.

Change in school attendance of child is school attendance in 1998 (during) - school attendance in 1997 (pre). Change in school attendance of peer group is peer group mean school attendance in 1998 - peer group mean school attendance in 1997.

IV1 uses "Village with PROGRESA" as instrument for peer group average schooling, IV2 uses "Village with PROGRESA * Eligible Fraction in Peer Group" as instrument for peer group average schooling. Control Variables are those listed in Table 3. Stock-Yogo (2005) critical value for Cragg-Donald (1993) statistic at 10 % maximal IV size is 16.38.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 5:

Sensitivity Analysis
Dependent Variable: Change in School Attendance
Sample: *Ineligible* Children

	A. Baseline Result	*B. Testing Alternatives*	*C. Within Village*	*D. Within villages, more precision*	*E. Anticipation*	*F. On Grade Level*
Change in Peer Group School Attendance	0.542	0.572	0.599	0.541	0.583	0.551
	(0.208)***	(0.331)*	(0.752)	(0.233)**	(0.249)**	(0.229)**
Village with PROGRESA	-	-0.001	-	-	-	-
		(0.014)				
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
State Effects (7)	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Progresa * Eligible Fraction	Progresa * Eligible Fraction	Progresa * Deviation from Village Mean Eligible Fraction	Progresa * Deviation from Village Mean Eligible Fraction and Progresa	Progresa * Eligible Fraction	Progresa * Eligible Fraction
Cragg-Donald (1993) weak identification statistic	90.23	30.512	7.406	35.155	61.813	55.032
Hansen J statistic	-	-	-	0.007	-	-
Observations	5143	5143	5143	5143	3809	2246

Notes: Robust standard errors in parentheses (allow for clustering at village level). * significant at 10%; ** significant at 5%; *** significant at 1%. Column A repeats the baseline result (Table 4, column F)
Column B adds PROGRESA status of the village as a regressor to test wheter there are PROGRESA effects on all villages. Column C identifies using within village variation in the eligible fraction per grade.
Column D adds PROGRESA status of the village as a second instrument to the analysis in column C. Column E discards all children who are close to the poverty threshold.
Column F focuses on children who are on grade level, i.e. who have progressed through school without repeating a grade.
Stock and Yogo (2005) critical value for Cragg-Donald (1993) statistic is 16.38 for 10 % maximal IV size in columns A, B, C, E, and F; critical value is 19.93 in column D.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 6:

The Effect of Peer Group Schooling on Individual Schooling in Subgroups
Dependent Variable: Change in School Attendance
Sample: *Ineligible* Children

	A. Grades 3-4	*B. Grades 5-6*	*C. Girls*	*D. Boys*
Change in Peer Group School Attendance	0.621	0.481	0.531	0.599
	(0.300)**	(0.270)*	(0.284)*	(0.305)*
Control Variables	Yes	Yes	Yes	Yes
State Effects (7)	Yes	Yes	Yes	Yes
Cragg-Donald (1993) Weak Identifiation Statistic	48.004	53.63	53.063	38.269
Observations	2205	2938	2589	2554

Notes: Robust standard errors in parentheses (allow for clustering at village level). * significant at 10%; ** significant at 5%; *** significant at 1%
Peer group contains all children in same grade in same village. See Table 3 for a list of all control variables.
Stock-Yogo (2005) critical value for Cragg-Donald (1993) statistic at 10 % maximal IV size is 16.38.
Estimates based on IV2 strategy that uses "Eligible fraction * village with PROGRESA" as instrument.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 7:

Measuring Direct and Social Interaction Effects of PROGRESA
Dependent Variable: Change in School Attendance
Sample: Eligible and Ineligible Children

Dependent Variable	Change in Individual School Attendance	Change in Peer Group School Attendance	Change in Individual School Attendance	Change in Individual School Attendance - Change in Peer Group School Attendance
	A. OLS	*B. OLS*	*C. IV2*	*D. Within Peer Groups*
Change in Peer Group School Attendance	-	-	0.505	-
			(0.253)**	
Treated*Poor (Direct Effect)	0.035	0.004	0.033	-
	(0.014)**	(0.004)	(0.016)**	
Village with PROGRESA * Eligible fraction	0.031	0.062	-	-
	(0.017)*	(0.011)***		
Treated * (Poor - Eligible fraction)	-	-	-	0.031
				(0.016)*
Control Variables	Yes	Yes	Yes	Yes
State Effects (7)	Yes	Yes	Yes	Yes
Cragg-Donald (1993) Weak Identifiation Statistic	-	-	79.095	-
Observations	15574	15574	15574	15574

Notes: Robust standard errors in parentheses (allow for clustering at village level). * significant at 10%; ** significant at 5%; *** significant at 1%. Peer group consists of all children in same grade in same village.
IV2 uses"Village with PROGRESA * Eligible Fraction in Peer Group" as instrument. Control Variables are those listed in Table 3.
Stock-Yogo (2005) critical value for Cragg-Donald (1993) statistic at 10 % maximal IV size is 16.38.
Change in school attendance of child is school attendance in 1998 (during) - school attendance in 1997 (pre). Change in school attendance of peer group is peer group mean school attendance in 1998 - peer group mean school attendance in 1997.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table 8:

Comparing Direct and Indirect Effects of PROGRESA
Based on Results in Table 7

	A. Eligible Children	*B. Ineligible Children*
I. Direct Treatment Effect of Cash Grant	0.031	0.000
	(0.016)**	-
II. Social Interactions Parameter	0.505	0.505
	(0.253)**	(0.253)**
III. Effect of Cash Grant on Peer Group Schooling	0.048	0.039
	(0.009)***	(0.010)***
IV. Indirect Social Interactions Effect (II*III)	0.024	0.020
	(0.011)*	(0.013)*
V. Total Treatment Effect (I+IV)	0.055	0.020
	(0.019)***	(0.013)*

Notes: I and II taken from Table 7, IIIA taken from a regression of peer group average schooling on "village with PROGRESA" and controls, IIIB taken from Table 4. Standard errors for IV and V calculated using delta method.

Source: Own Calculation, Based on Progresa Evaluation Data.

Table A1:
PROGRESA cash subsidy in August-December 1998, per child

	A. Girl	B. Boy
Primary School		
3	70	70
4	80	80
5	95	95
6	115	115
Secondary School		
1	210	200
2	235	210
3	255	225

Source: Table 1 in Behrman, Sengupta, Todd (2001)

Table A2:

Dependent Variable:	School Attendance October 1997	Predicted Change in School Attendance
Village with PROGRESA	0.029	-0.018
	(0.052)	(0.013)
Village with PROGRESA * Eligible Fraction in Peer Group	-0.011	0.022
	(0.077)	(0.019)
Eligible Fraction in Peer Group	0.100	-0.036
	(0.062)	(0.013)***
Constant	0.703	-0.046
	(0.042)***	(0.008)***
Observations	5143	5143
R2	0.00	0.01

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A3:
Detailed Results for Table 4

	IV1 identification strategy			IV2 identification strategy		
DV: Change in school Attendance of	Child	Peer Group	Child	Child	Peer Group	Child
Change in Peer Group School Attendance			0.534			0.542
			(0.235)**			(0.208)***
Village with PROGRESA	0.021	0.039				
	(0.012)*	(0.010)***				
Village with PROGRESA * Eligible fraction				0.041	0.076	
				(0.019)**	(0.015)***	
Eligible fraction				-0.008	-0.054	0.022
				(0.027)	(0.022)**	(0.019)
Completed grade 4	-0.016	-0.021	-0.005	-0.015	-0.021	-0.004
	(0.012)	(0.008)***	(0.011)	(0.012)	(0.008)***	(0.011)
Completed grade 5	-0.174	-0.171	-0.083	-0.172	-0.170	-0.080
	(0.017)***	(0.015)***	(0.043)*	(0.017)***	(0.014)***	(0.038)**
Completed grade 6	0.020	0.020	0.009	0.022	0.020	0.011
	(0.014)	(0.010)*	(0.012)	(0.014)	(0.011)*	(0.012)
Girl	-0.008	0.006	-0.011	-0.008	0.006	-0.011
	(0.010)	(0.005)	(0.010)	(0.010)	(0.005)	(0.010)
Mother Ed.: Primary School or Higher	0.013	0.011	0.007	0.013	0.011	0.008
	(0.011)	(0.007)	(0.012)	(0.011)	(0.007)	(0.012)
Father Ed.: Primary School or Higher	0.033	0.006	0.029	0.034	0.007	0.030
	(0.013)**	(0.007)	(0.013)**	(0.013)**	(0.007)	(0.013)**
Father Information Missing	-0.022	0.019	-0.032	-0.022	0.018	-0.032
	(0.015)	(0.008)**	(0.016)*	(0.015)	(0.008)**	(0.016)**
Floor: Cement	0.005	-0.005	0.007	0.005	-0.006	0.008
	(0.013)	(0.007)	(0.012)	(0.013)	(0.007)	(0.012)
Roof: Tin	0.017	0.009	0.012	0.016	0.008	0.012
	(0.017)	(0.009)	(0.016)	(0.017)	(0.009)	(0.016)
Roof: Asbest	0.029	0.030	0.013	0.030	0.029	0.014
	(0.020)	(0.012)**	(0.021)	(0.020)	(0.012)**	(0.020)
Roof: Tiles	0.033	0.015	0.025	0.034	0.015	0.026
	(0.025)	(0.016)	(0.021)	(0.025)	(0.016)	(0.021)
Roof: Cement Blocks	0.025	0.008	0.020	0.026	0.008	0.021
	(0.019)	(0.010)	(0.018)	(0.019)	(0.010)	(0.018)
Secondary School in Village	0.029	0.038	0.009	0.029	0.037	0.009
	(0.014)**	(0.011)***	(0.013)	(0.014)**	(0.010)***	(0.013)
Constant	-0.071	-0.091	-0.022	-0.076	-0.065	-0.041
	(0.046)	(0.029)***	(0.040)	(0.050)	(0.032)**	(0.043)
State Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5143	5143	5143	5143	5143	5143

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%