Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresa.*

Orazio Attanasio,[†] Costas Meghir,[‡] Ana Santiago[§]

July 2009
(First version January 2001)

Abstract

In this paper we use an economic model to analyse data from a major social experiment, namely PROGRESA in Mexico, and to evaluate its impact on school participation. In the process we also show the usefulness of using experimental data to estimate a structural economic model. The evaluation sample includes data from villages where the program was implemented and where it was not. The allocation was randomised for evaluation purposes. We estimate a structural model of education choices and argue that without such a framework it is impossible to evaluate the effect of the program and, especially, possible changes to its structure. We also argue that the randomized component of the data allows us to identify a more flexible model that is better suited to evaluate the program. We find that the program has a positive effect on the enrollment of children, especially after primary school; this result is well replicated by the parsimonious structural model. We also find that a revenue neutral change in the program that would increase the grant for secondary school children while eliminating for the primary school children would have a substantially larger effect on enrollment of the latter, while having minor effects on the former.

*This paper has benefitted from valuable comments from Joe Altonji, Gary Becker, Esther Duflo, Jim Heckman, Hide Ichimura, Paul Schultz, Miguel Székely, Petra Todd and many seminar audiences. Costas Meghir thanks the ESRC for funding under the Professorial Fellowship RES-051-27-0204. Orazio Attanasio thanks the ESRC for funding under the Professorial Fellowship RES-051-27-0135. We also thank the ESRC Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies.Responsibility for any errors is ours.

[†]UCL, IFS and NBER.
[‡]UCL, IFS, CEPR and IZA
[§]UCL and IADB

1 Introduction

In 1998 the Mexican government started a remarkable new program in rural localities. PROGRESA was one of the first and probably the most visible of a new generation of interventions whose main aim is to improve the process of human capital accumulation in the poorest communities by providing cash transfers conditional on specific types of behaviour in three key areas targeted by the program: nutrition, health and education. Arguably the largest of the three components of the program was the education one. Mothers in the poorest households in a set of targeted villages are given grants to keep their children in school. The grants start in third grade and increase until the ninth and they are conditional on school enrolment and attendance. PROGRESA was noticeable and remarkable not only for the original design but also because, when the program begun, the Mexican government started a rigorous evaluation of its effects.

The evaluation of PROGRESA is greatly helped by the existence of a high quality data set whose collection was started at the outset of the program, between 1997 and 1998. The PROGRESA administration identified 506 communities that qualified for the program and started the collection of a rich longitudinal data set in these communities. Moreover, 186 of these communities where randomized out of the program with the purpose of providing a control group that would enhance the evaluation. However, rather than being excluded from the program all together, in the control villages the program was postponed for about two years, during which period, four waves of the panel were collected. Within each community in the evaluation sample, all households, both beneficiaries and non-beneficiaries, were covered by the survey. In the control villages, it is possible to identify the would-be beneficiaries were the program to be implemented.

The aim of this paper is to analyze the impact of monetary incentives on education choices in rural Mexico and to discuss the most effective design of educational interventions aimed at improving educational participation in developing countries. To achieve this goal we combine the information provided by the randomized allocation of PROGRESA across localities with a simple structural model of education choices. The main contributions of the paper are therefore twofold: on the one hand, we estimate the effect of PROGRESA on school enrolment and predict the effect of slightly different programs; on the other, we see our approach to this particular evaluation problem as making a methodological contribution to the evaluation literature. From the substantive point of view, we estimate the effect of PROGRESA within a structural model that allows us to simulate the effect of changes to some of the parameters of the programs. We are also able to disentangle the general equilibrium effects and the attenuation effect that they might have on the program impact. From the methodological point of view, we show that even when data from a randomized trial are available, the use of a structural model is necessary if one wants to use the evaluation evidence to improve the design of the program.

By all accounts and evidence the evaluation of PROGRESA, based on the large randomized experiment described above, was highly successful (see Schultz, 2003). Hence the program impacts could be estimated by comparing mean outcomes between treatment and control villages. However, to get more out of this exceptional experiment, we need to provide a credible model of individual behavior, which can be estimated exploiting the rich data set and the experimental variation induced by the randomization.

PROGRESA effectively changes the relative price of education and child labour in a controlled and exogenous fashion. A tightly parameterized model under suitable restrictions (see below) could be used, together with data on wages and enrollment, to predict the effect of the program even before its implementation using variation in wages across communities where the program is not available. This is the strategy followed, for instance by Todd and Wolpin (2006). Whether the predictions of the structural model line up with the experimental estimates constitutes then a strong test of the former and is in it self an important question.

Like Todd and Wolpin (2006), we estimate a structural model but our approach and objective are different. We exploit the exogenous variation induced by the randomization to estimate a more flexible specification than would be possible without the programme: critically, we do not restrict the effect of the grant to be the same as that of wages.

While it is true that variation in the conditional grant has an effect similar to changes in child wages, as it modifies the relative price of school versus work, it is plausible that the marginal utility of income differs depending on whether the child attends school or not; this is just a manifestation of preferences being nonseparable in consumption and schooling. It is therefore possible that the changes in the grant and in wages have different effects on school enrollment. The experiment allows us to test whether this the case empirically and if necessary to relax this separability restriction.

In what follows, we estimate a dynamic structural model of education choice using survey data collected to evaluate the experiment. Our model is similar to the Willis and Rosen (1979) model where individuals base their choice on comparing the costs and benefits of additional schooling depending on their comparative advantage. Our estimation approach is similar to that of a much simplified version of the model by Keane and Wolpin (1997), although their problem is a more complete study of careers. A distinguishing feature of our model is the use of the randomised experiment and the fact that we can allow the program effect to differ from the effect of the wage, which is the usual opportunity cost of education. This is potentially very important for understanding the extent to which *ex-ante* evaluation can be helpful in predicting the effects of a planned intervention of this sort.

A study of this sort and a better understanding of the effectiveness of such a policy is important as deficits in the accumulation of human capital have been identified by several commentators as one of the main reasons for the relatively modest growth performance of Latin American economies in comparison, for instance, with some of the South East Asian countries (see, for instance, Behrman, 1999, Behrman, Duryea and Székely, 1999,2000). For this reason, the program we study and similar ones have received considerable attention in Latin America.

A randomised experiment such as the one designed to evaluate PROGRESA, offers a purely exogenous variation in some well identified dimensions, such as the opportunity cost of schooling. On its own it can answer only a limited question (albeit without relying on any parametric or functional form assumption), namely how did the specific program implemented in the experiment affect the outcomes of interest. However, policy may require answers to much more refined questions, such as extrapolating to different groups or altering the parameters of the initial program. Thus, we can think of using experimental variation and indeed designing experiments to generate such variability so as to more credibly estimate structural models capable of rich policy analysis. Thus, our work draws from the tradition of Orcutt and Orcutt (1968) who advocate precisely this approach.

Our results show that, in addition to the non parametric identification of the average treatment effect, the experiment allows us to estimate a richer and more flexible structural model that can shed light on the mechanisms that generate the observed impacts. They also motivate both further theoretical work to better understand the observed outcomes and the design of more complex, but still realistic experiments, to improve measurement of effects and extrapolation. At the same time, the approach we advocate, while using the experiment, can also offer a framework for extrapolating beyond the experiment and redesigning policies. Thus our approach combines the virtues of experiments with the benefits of an explicitly specified framework for policy simulation.

Todd and Wolpin (2006) use the experiment to validate their model, which they estimate on the control group as if no experiment ever took place; they do not offer an approach that requires experimental variation other than as a check to their model. Thus, while both papers use the same data, the idea underlying the two research programs are different. Ultimately we believe that it is important to try and use as much genuinely exogenous variation to identify structural relationships and designing or using existing experiments for this purpose is likely to lead to important advances in structural modelling.[1]

[1]There are also differences in the estimation approach.as well as in the specification of the models. On the estimation side we exploit the increasing availability of schooling as an instrument to control for the initial conditions problem so as to better disentangle state dependence from unobserved heterogeneity. As for the specification, Todd and Wolpin's model solves a family decision problem, including tradeoffs between children and fertility, which we do not. What has informed our modelling choices is the focus on using the incentives introduced by the programme in some localities to identify in a credible fashion our model. Considering fertility effects is potentially interesting. However, it should be pointed out that

As mentioned above, PROGRESA was randomized across localities, rather across households. As these localities are isolated from each other, this experimental design also affords the possibility of estimating general equilibrium effect induced by the program. In the context of the educational grant to promote school enrolment and attendance, one could imagine that, if the program is effective, a reduction in the supply of child labour could result in an increase in children wages which, in turn, would result in an attenuation of the program's impact.

Although some papers in the literature (see, for instance, Angelucci and De Giorgi, 2008) have looked at the impact of PROGRESA on some prices and other village level variables, perhaps surprisingly, no study has considered, as far as we know, the effect that the program has had on children wages. In our exercise we estimate this impact (taking into account the fact that children wages is observed only for the selected subset of children who actually work) and use these general equilibrium impacts within our model and in our simulations. To the best of our knowledge this is the first structural model that includes explicitly general equilibrium effects both at the estimation and the simulation level.

The rest of the paper is organized as follows. In Section 2, we present the main features of the program and of the evaluation survey we use. In section 3, we present some simple results on the effectiveness of the program based on the comparison of treatment and control villages and discuss the limitations of this evidence. In section 4, we present a structural dynamic model of education choices that we estimate by Maximum Likelihood. Section 5 presents the results we obtain from the estimation and uses the model to perform a number of policy simulations that could help to fine-tune the program. Finally, Section 6 concludes the paper with some thoughts about open issues and future research.

the programme did not have any effects on fertility. Todd and Wolpin's model of fertility is identified only using the observed cross sectional variation which may not necessarily reflect exogenous differences in incentives.

2 The PROGRESA program.

In 1997, the Mexican government started a large program to reduce poverty in rural Mexico. The program proposed by the Zedillo administration was innovative in that introduced a number of incentives and conditions with which participant households had to comply to keep receiving the program's benefits. When the program was started, the administration decided to collect a large longitudinal survey with the scope of evaluating the effectiveness of the program, In this section, we describe the main features of the program and of the evaluation survey.

2.1 The specifics of the PROGRESA program

PROGRESA is the Spanish acronym for "Health, Nutrition and Education", that are the three main areas of the program. PROGRESA is one of the first and probably the best known of the so-called 'conditional cash transfers', which aim at alleviating poverty in the short run while at the same time fostering the accumulation of human capital to reduce it in the long run. This is achieved by transferring cash to poor households under the condition that they engage in behaviours that are consistent with the accumulation of human capital: the nutritional subsidy is paid to mothers that register the children for growth and development check ups and vaccinate them as well as attend courses on hygiene, nutrition and contraception. The education grants are paid to mothers if their school age children attend school regularly. Interestingly such conditional cash transfers have become quite popular. In 1998 the British government has piloted a similar program targeted to children aged 16-18 (see Dearden et al 2008). The program has received considerable attention and publicity. More recently programs similar to and inspired by PROGRESA were implemented in Colombia, Honduras, Nicaragua, Argentina, Brazil, Turkey and other countries. Rawlings (2004) contains a survey of some of these programs. Skoufias (2001) provides additional details on PROGRESA and its evaluation.

PROGRESA is first targeted at the locality level. In 1997, a number of poor communities in rural Mexico were declared eligible for the program. Roughly speaking, the two criteria communities had to satisfy to qualify for the program were a certain degree of poverty (as measured by what is called an 'index of marginalization', basically the first principal component of a certain number of village level variables routinely collected by the government) and access to certain basic structures (schools and health centers). The reason for the second criterion is the conditional nature of the program: without some basic structures within a certain distance, beneficiary households could not comply with the conditions for retaining the beneficiary status (participation in vaccination and check-up visits for the health and nutrition components and school attendance for the education component). As a consequence of these eligibility criteria the PROGRESA communities, while poor, are not the poorest in Mexico.

Within each community, then the program is targeted by proxy means testing. Once a locality qualifies, individual households could qualify or not for the program, depending on a single indicator, once again the first principal component of a number of variables (such as income, house type, presence of running water, and so on). Eligibility was determined in two steps. First, a general census of the PROGRESA localities measured the variables needed to compute the indicator and each household was defined as 'poor' or 'not-poor' (where 'poor' is equivalent to eligibility). Subsequently, in March 1998, an additional survey was carried out and some households were added to the list of beneficiaries. This second set of beneficiary households are called 'densificados'. Fortunately, the re-classification survey was operated both in treatment and control towns.

The largest component of the program is the education one. Beneficiary households with school age children receive grants conditional on school attendance. The size of the grant increases with the grade and, for secondary education, is slightly higher for girls than for boys. In Table 1, we report the grant structure. All the figures are in current pesos, and can be converted in US dollars at approximately an exchange rate of 10 pesos per dollar. In addition to the (bi) monthly payments, beneficiaries with children in school age receive a small annual grant for school supplies.

For logistic and budgetary reasons, the program was phased in slowly but is currently very large. In 1998 it was started in less than 10,000 localities. However, at the end of 1999 it was implemented in more than 50,000 localities and had a budget of about US$777 million or 0.2% of Mexican GDP. At that time, about 2.6 million households, or 40% of all rural families and one ninth of all households in Mexico, were included in the program. Subsequently the program was further expanded and, in 2002-2003 was extended to some urban areas.

The program represents a substantial help for the beneficiaries. The nutritional component of 100 pesos per month (or 10 US dollars) in the second semester of 1998, corresponded to 8% of the beneficiaries' income in the evaluation sample.

As mentioned above, the education grants are conditional to school enrolment and attendance of children, and can be cumulated within a household up to a maximum of 625 pesos (or 62.5 dollars) per month per household or 52% of the average beneficiary's income. The average grant per household in the sample we use was 348 pesos per month for households with children and 250 for all beneficiaries or 21% of the beneficiaries income. To keep the grant, children have to attend at least 85% of classes. Upon not passing a grade, a child is still entitled to the grant for the same grade. However, if the child fails the grade again, it looses eligibility.

2.2 The evaluation sample

Before starting the program, the agency running it decided to start the collection of a large data set to evaluate its effectiveness. Among the beneficiaries localities, 506 where chosen randomly and included in the evaluation sample. The 1997 survey was supplemented, in March 1998, by a richer survey in these villages, located in 7 of the 31 Mexican states. All households in these villages where interviewed, for a total of roughly 25,000 households. Using the information of the 1997 survey and that in the March 1998 survey, each household can be classified as poor or non-poor, that is, each household can be identified as being entitled or not to the program.

PROGRESA bi-monthly monetary benefits				
Type of benefit	1998 1st sem.	1998 2nd sem.	1999 1st sem.	1999 2nd sem
Nutrition support	190	200	230	250
Primary school				
3	130	140	150	160
4	150	160	180	190
5	190	200	230	250
6	260	270	300	330
secondary school				
1st year				
boys	380	400	440	480
girls	400	410	470	500
2nd year				
boys	400	400	470	500
girls	440	470	520	560
3rd year				
boys	420	440	490	530
girls	480	510	570	610
maximum support	1,170	1,250	1,390	1,500

Table 1: Table Caption

One of the most interesting aspects of the evaluation sample is the fact that it contains a randomization component. The agency running PROGRESA used the fact that, for logistic reasons, the program could not be started everywhere simultaneously, to allocate randomly the villages in the evaluation sample to 'treatment' and 'control' groups. In particular, in 320 randomly chosen villages of the evaluation sample were assigned to the communities where the program started early, that is in May 1998. The remaining 186 villages were assigned to the communities where the program started almost two years later (December 1999 rather than May 1998).

An extensive survey was carried out in the evaluation sample: after the initial data collection between the end of 1997 and the beginning of 1998, an additional 4 instruments were collected in November 1998, March 1999, November 1999 and April 2000. Within each village in the evaluation sample, the survey covers all the households and collects extensive information on consumption, income, transfers and a variety of other variables. For each household member, including each child, there is information about age, gender, education, current labour supply, earnings, school enrolment, and health status. The household survey is supplemented by a locality questionnaire that provides information on prices of various commodities, average agricultural wages (both for males and females) as well as institutions present in the village and distance of the village from the closest primary and secondary school (in kilometers and minutes).

In what follows we make an extensive use of both the household and the locality survey. In particular, we use the household questionnaire to get information on each child's age, completed last grade, school enrolment, parental background, state of residence, school costs. We use the locality questionnaire to get information on distance from schools and prevailing wages.

3 Measuring the impact of the program by comparing treatment and control villages.

In this section, we discuss the evidence that can be obtained by comparing treatment and control villages. The main purpose of this exercise is to show how the random allocation of the program between treatment and control localities can be used to evaluate the effect of the program. However, such an exercise would estimate the impact of the program as a whole, without specifying the mechanisms through which it operates. Schultz (2003) presents a complete set of evaluation results, which are substantially similar to those presented here. Here our focus is on introducing some aspects of the data that are pertinent to our model and to the sample we use to estimate it. And more importantly, by describing the effect of the program in this sample we use to estimate our structural model, we set the mark against which it will be fitted.

If the allocation of the program across localities is truly random and treatment and control villages are statistically identical, comparing them provides a reliable and simple estimate of the program's impact. The quality of the randomization can be evaluated by testing for the presence of significant difference in measured variables in the baseline surveys in treatment and control villages. An extensive study of this type has been performed by Behrman and Todd (1999). They find that, for the large majority of a very wide set of variables, there are no statistical differences between treatment and control villages.

Given the results in Behrman and Todd (1999), we proceed to evaluate the difference in enrollment rates between treatment and control variables for eligible and ineligible children. As our structural model will be estimated on boys, we report only the results for them. The effects for girls are slightly higher but not substantially different from those reported here for boys. As we will be interested in how the effect of the grant is affected by age, we also report the results for different age groups. In particular, we consider boys less than 10, boys aged 10 to 13 and older than 13. Finally we need to take into account two possible definitions of 'eligibility'. While, in principle, children belonging to households that were re-defined as eligible in 1998, just before the start of the program, should all have got the program (conditional on school attendance), it turns out that a substantial fraction of them, probably for administrative errors, did not. In Table 2, where we report our results, therefore, we have three columns. The first refers to the difference in school enrolments for children who were defined as eligible in 1997. Most of these children did get the program conditional on attendance. In the second column, we add the children who were classified as eligible in 1998. As mentioned above, a substantial fraction of the 'reclassified' children did not receive the program. The estimates in this column should measure what in the evaluation literature is sometimes referred to as the intention to treat . In the third column we report the difference in enrolment rates for ineligible children.

The average treatment effect is of 0.039 points on the children belonging to the original eligible households. The effect is concentrated among the oldest children, being not significantly different from zero for the youngest group and almost 9% for the oldest. These effects are slightly smaller when we consider the intent to treat estimates on the group that includes the new eligible households, some of which, did not get the program.

Post-program differences in educational attendance between treatment and control communities				
Age Group	Participation Control	Eligible 97	Eligible 97 + 98	non eligible
6-17	0.751	0.039	0.034	0.049
		(0.012)	(0.012)	(0.021)
6-9	0.937	0.008	0.003	-0.014
		(0.005)	(0.005)	(0.009)
10-13	0.870	0.037	0.032	0.030
		(0.012)	(0.011)	(0.022)
14-17	0.352	0.089	0.084	0.099
		(0.033)	(0.031)	(0.0432)
Standard errors in parentheses are clustered at the locality level.				

Table 2: Experimental Results October 1998

The results in the third column are the most surprising, as they would indicate a substantive effect of the program on the enrollment of ineligible children. There are two possible alternative explanations of such a result. First, it is possible that the effect is genuine and reflects spill-overs from eligible families to non eligible ones. The alternative explanation, instead, would point out to pre-program differences between treatment and control villages. Given the evidence in Behrman and Todd (1999), the latter explanation is surprising. However, the size of the effect seems to be too large to be plausible: for older children (aged 14-17) the effect on ineligible children would be actually larger than for the eligible ones.

To investigate the matter further, we turn to the analysis of pre-program differences. In Table 3 we re-do the exercise presented in Table 2 but on the baseline data collected in August 1997, before the program started operating. From this table, it emerges that the pre-program differences between attendance rates of eligible children in treatment and control communities are very small indeed and never statistically different from zero. Instead, especially for older children, we observe sizeable differences in pre-program enrollment rates for non- eligible children. While the estimates are not extremely precise, and an explanation of these differences is not obvious, it is important, in evaluating the effect of the program, to take this difference into account.

Having used the baseline to verify the existence of pre-program differences in enroll-

Pre-program differences in educational attendance between treatment and control communities			
age group	eligible 97	eligible 97 + 98	ineligible
6-17	0.006	0.006	0.033
	(0.011)	(0.011)	(0.021)
6-9	-0.004	-0.003	-0.003
	(0.003)	(0.003)	(0.009)
10-13	0.014	0.013	0.034
	(0.012)	(0.011)	(0.022)
14-17	0.014	0.015	0.048
	(0.030)	(0.028)	(0.035)
Standard errors in parentheses are clustered at the locality level.			

Table 3: Baseline Results August 1997

Difference-in-differences estimates of program impact on educational attendance			
age group	eligible 97	eligible 97 + 98	ineligible
6-17	0.033	0.028	0.016
	(0.009)	(0.008)	(0.019)
6-9	0.012	0.007	-0.010
	(0.006)	(0.006)	(0.011)
10-13	0.024	0.020	-0.003
	(0.011)	(0.011)	(0.023)
14-17	0.075	0.069	0.051
	(0.025)	(0.021)	(0.037)
Standard errors in parentheses are clustered at the locality level.			

Table 4: Experimental Results October 1998- August 1997

ment rates, we can also use it to control for them. In particular, to estimate the effect of the program, we can use a difference in difference estimator. Such an approach, however, requires assumptions not usually necessary in a randomized experiment: that is that trends between treatment and control groups are the same. We report the estimates of the effect of the program that take into account the differences observed in Table 3 in Table 4.

The results in Table 4 indicate that the program has, on average, an effect of about 3% on school enrollment of boys aged 6 to 17. There are, however, some large differences in the effect of the program by age. The effect is much larger (around 7.5%) for older boys and virtually zero for boys younger than 9. Finally, on ineligible boys we find no significant effect of the program. Only for the older group the point estimate indicates a somewhat positive effect, but not as large as the one obtained by simple differences. Unfortunately, for this group, which is much smaller than that of the beneficiaries, the estimates are not extremely precise.

What do we learn from this exercise that is relevant for our structural approach? First and foremost, it is clear that there were some pre-program differences between treatment and control villages which should be taken into account. These differences make a difference for the size of the estimated program's impact and, although we do not have a convincing explanation for them, they will have somehow enter in our structural model. It is somewhat comforting that the pre-program differences are concentrated among the ineligible families. A simple strategy, therefore, would be to restrict our estimation sample to the eligible households. We come back to these issues when we discuss the identification of our model. Second, the effect of the program is different for different age groups. As we will see, our model will assume that the program works by changing the relative price of schooling relative to working. While we consider a reasonably flexible specification, it is not completely obvious that the model should reproduce the pattern of different age effects we see in the data, as we estimate a unique coefficient for the grant. The pattern of results that emerges from Table 4 therefore, constitutes a useful benchmark against which to evaluate the goodness of fit of our model.

4 The model

We use a simple dynamic school participation model. Each child, (or his/her parents) decide whether to attend school or to work taking into account the economic incentives involved with such choices. Parents are assumed here to act in the best interest of the child and consequently we do not admit any interactions between children. We assume that children have the possibility of going to school up to age 17. All formal schooling ends by that time. In the data, almost no individuals above age 17 are in school. We assume that children who go to school do not work and vice-versa. We also assume that children necessarily choose one of these two options. If they decide to work they receive a village/education/age specific wage. If they go to school, they incur a (utility) cost (which might depend on various observable and unobservable characteristics) and, with a certain probability, progress a grade. At 18, everybody ends formal schooling and reaps the value of schooling investments in the form of a terminal value function that depends on the highest grade passed. The PROGRESA grant is easily introduced as an additional monetary reward to schooling, that would be compared to that of working.

The model we consider is dynamic for two main reasons. First, the fact that one cannot attend regular school past age 17 means that going to school now provides the option of completing some grades in the future: that is a six year old child who wants to complete secondary education has to go to school (and pass the grade) every single year, starting from the current. This source of dynamics becomes particularly important when we consider the impact of the PROGRESA grants, since children,as we saw above, are only eligible for six grades: the last three years of primary school and the first three of secondary. Going through primary school (by age 14), therefore, also 'buys' eligibility for the secondary school grants. Second, we allow for state dependence: The number of years of schooling affects the utility of attending in this period. We explicitly address the initial conditions problem that arises from such a consideration and discuss the related identification issues at length below. State dependence is important because it may be a mechanism that reinforces the effect of the grant.

Before discussing the details of the model it is worth mentioning that using a structural approach allows us to address the issue of anticipation effects and the assumptions required for their identification. PROGRESA as well as other randomized experiments or pilot studies create a control group by delaying the implementation of the program in some areas, rather than excluding them completely. It is therefore possible that the control villages react to the program prior to its implementation, depending on the degree to which they believe they will eventually receive it. A straight comparison between treatment and control areas may then underestimate the impact of the program. A structural model that exploits other sources of variation, such as the variation of the grant with age may be able to estimate the extent of anticipation effects. We investigated this with our model by examining its fit under different assumptions about when the controls are expecting to receive payment. As it turns out we find no evidence of anticipation effects in our data. This is not surprising because there was no explicit policy announcing the future availability of the grants. The absence of evidence on anticipation effects, however, is consistent both with no information about the future availability of the program and with an inability to take advantage of future availability due, for instance, to liquidity constraints.

4.1 Instantaneous utilities from schooling and work

The version of the model we use assumes linear utility. In each period, going to school involves instantaneous pecuniary and non-pecuniary costs, in addition to losing the opportunity of working for a wage. The current benefits come from the utility of attending school and possibly, as far as the parents are concerned, by the childcare services that the school provides during the working day. As mentioned above, the benefits are also assumed to be a function of past attendance. The costs of attending school are the costs of buying books etc. as well as clothing items such as shoes. There are also transport costs to the extent that the village does not have a secondary school. For households who are entitled to PROGRESA and live in a treatment village, going to school involves receiving the grade and gender specific grant.

As we are using a single cross section, we use the notation t to signify the age of the child in the year of the survey. Variables with a subscript t may be varying with age. Denote the utility of attending school for individual i in period t who has already attended ed_{it} years as u_{it}^s.We posit:

$$
\begin{aligned}
u_{it}^s &= Y_{it}^s + \alpha g_{it} \qquad (1)\\
Y_{it}^s &= \mu_i^s + a^{s\prime} z_{it} + b^s ed_{it} + 1(p_{it} = 1)\beta^p x_{it}^p + 1(s_{it} = 1)\beta^s x_{it}^s + \varepsilon_{it}^s
\end{aligned}
$$

where g_{it} is the amount of the grant an individual is entitled to and it will be equal to zero for non-eligible individuals and for control localities. Y_{it}^s represents the remaining pecuniary and non pecuniary costs or gains that one gets from to going to school. z_{it} is a vector of taste shifter variables, including parental background, age and state dummies. The variable $1(p_{it} = 1)$ denotes attendance in primary school, while the variable $1(s_{it} = 1)$ denotes attendance in secondary school. x_{it}^p and x_{it}^s represent factors affecting the costs of attending primary school and secondary school respectively. The term ε_{it}^s represents an extreme value error term which is assumed independently and identically distributed over time and individuals Notice that the presence of ed_{it} introduces an important element of dynamics we alluded to above: schooling choices affect future grades and, therefore, the utility cost of schooling. Finally, the term μ_i^s represents unobservables which we assume have a constant impact over time.[2]

The utility of not attending school is denoted by

$$\begin{aligned} u_{it}^w &= Y_{it}^w + \delta w_{it} \\ Y_{it}^w &= \mu_i^w + a^{w\prime} z_{it} + b^w ed_{it} + \varepsilon_{it}^w \end{aligned} \tag{2}$$

where w_{it} are (potential) earnings when out of school. The wage is a function (estimated from data) of age and education attainment as well as village of residence that we discuss below. Notice that, while the grant involves a monetary payment, just like the wage, we allow the coefficient on the two variables to be different, possibly reflecting disutility from work. On the other hand, we can only identify the difference between the coefficients on the variables that enter both the utility of work and that of school. We can therefore, without loss of generality, re-write equations 1 and 2 as follows:

[2]We have employed a one factor model of unobserved heterogeneity, where the unobservables affects only the costs of education. When we attempted a richer specification, allowing a second factor to affect the impact of the wage we got no improvement in the likelihood. There would be other options such as allowing for heterogeneity in the discount factor. However, in terms of fit, this is likely to act very much like the heterogeneous costs of education and overall the model did not seem to require any further unobserved factors to fit the data.

$$u_{it}^{s} = \gamma\delta g_{it} + \mu_i + a' z_{it} + bed_{it} + 1(p_{it}=1)\beta^p x_{it}^p + 1(s_{it}=1)\beta^s x_{it}^s + \varepsilon_{it} \quad (3)$$

$$u_{it}^{w} = \delta w_{it} \quad (4)$$

where $a = a^s - a^w$, $b = b^s - b^w$, $\gamma = \alpha/\delta$, $\mu_i = \mu_i^s - \mu_i^w$, $\varepsilon_{it} = \varepsilon_{it}^s - \varepsilon_{it}^w$. The error term ε_{it}, the difference between two extreme value distributed random variables and as such is distributed as a logistic. We will assume that μ_i is a discrete random variables whose points of support and probabilities will be estimated empirically. Finally note that all time-varying exogenous variables are assumed to be perfectly foreseen when individuals consider tradeoffs between the present and the future.

The coefficient γ measures the impact of the grant as a proportion of the impact of the wage on the education decision. The grant (which is a function of the school grade currently attended -as in Table 1) is suitably scaled so as to be comparable to the wage. If $\gamma = 1$, the effect of the grant on utility and therefore on schooling choices, would be the same as that of the wage. In some standard economic model they should have the same effect. If this was the case, the effect of the program could be estimated even using data only from the control communities in which it does not operate, from estimates of δ. This is the strategy used by Todd and Wolpin (2006). However, one can think of many simple models in which there is every reason to expect that the impact of the grant will be different from that of the wage.

The issue can be illustrated easily within a simple static model. As in our framework, we assume that utility depends on whether the child goes to school or not. Moreover we assume that this decision affects the budget constraint. In particular we have:

$$U^s = c^s = Y + g$$

$$U^w = \theta c^w - \alpha = \theta(Y + w) - \alpha$$

where c^s and c^w represent consumption conditional on the child going to school or to work, respectively; Y is other sources of income, α is the disutility of work and the parameter θ captures non-separabilities between consumption and child labour; in other words it allows for the possibility that the marginal utility of income depends on whether the child is working or attending school. The difference in utilities between school and work will then be given by:

$$U^s - U^w = (1-\theta)Y + g - \theta w$$

From this equation we can see that only if $\theta = 1$ the grant and the wage have the same effect on the decision to go to school. The same reasoning generalizes, *a fortiori*, to a dynamic setting.

The reason for this non-separability may be just because of the structure of preferences or because of the structure of intrahousehold decisions and allocations: PROGRESA cheques are actually handed out to the mother, while we do not know who receives the child's wage. Depending on the age of the child, wages are either received by the child or by one of the parents. Depending on who receives it, a standard collective model will predict different effects because the distribution of power will change in the household.[3]

Therefore, whether changes in the grant have the same effect as changes in child wages, is an empirical matter. Using the experiment we are able to test whether the grant and the wage have the same effect on school enrolment. The design of the experiment allows us to address this important issue.

This leads us back to the inevitable comparison with the paper by Todd and Wolpin (2006). Our approach has been to use the experiment to estimate our model for two important reasons: first, to have a source of genuine exogenous variation for the estimation of a structural model; second by using the experimental variation we are able to estimate a model that allows for a richer structure of preferences than is possible with standard observational data. In addition, our approach also allows us to account for general equilibrium effects, using direct information of the impact of the experiment on wages. The points just discussed distinguish our approach from that of Todd and Wolpin (2006). By

[3] See Blundell, Chiappori and Meghir (2005) on how spending on children depends on individual preferences and relative bargaining power.

demonstrating the scope of combining experimental data with structural models we hope to make it standard both to analyse experiments using structural models and to design experiments so as to enable the estimation of richer models.

Our sample includes both eligible and ineligible individuals. Eligibility is determined on the basis of a number of observable variables that might affect schooling costs and utility. To take into account the possibility of these systematic differences, we also include in equation 3 (among the z variables a dummy for eligibility (which obviously is operative both in treatment and control localities).

As we discussed in Section 3, there seems to be some differences in pre-program enrolment rates between treatment and control localities. As we do not have an obvious explanation for these differences, we use two alternative strategies. First we control for them by adding to the equation for the schooling utility (3) a dummy for treatment villages. Obviously such a dummy will absorb some of the exogenous variability induced by the experiment. We discuss this issue when we tackle the identification question in the next section. A less extreme approach, justified by the fact that most of the unexplained differences in pre-program enrollment is observed among non-eligible household, we introduce a dummy for this group only.

4.2 Uncertainty

There are two sources of uncertainty in our model. The first is an iid shock to schooling costs, modelled by the (logistic) random term ε_{it}. Given the structure of the model, having a logistic error in the cost of going to school is equivalent to having two extreme value errors, one in the cost of going to school and one in the utility of work. Although the individual knows ε_{it} in the current period,[4] she does not know its value in the future. Since future costs will affect future schooling choices, indirectly they affect current choices. Notice that the term μ_i, while known (and constant) for the individual, is unobserved by the econometrician.

[4]We could have introduced an additional residual term ε_{it}^{w} in equation 2. Because what matters for the fit of the model is only the difference between the current (and future) utility of schooling and working, assuming that both ε_{it} and ε_{it}^{w} were distributed as an extreme value distribution is equivalent to assuming a single logistic residual.

The second source of uncertainty originates from the fact that the pupil may not be successful in completing the grade. If a grade is not completed successfully, we assume that the level of education does not increase. We assume that the probability of failing to complete a grade is exogenous and does not depend on the willingness to continue schooling. We allow however this probability to vary with the grade in question and with the age of the individual and we assume it known to the individual.[5] We estimate the probability of failure for each grade as the ratio of individuals who are in the same grade as the year before at a particular age. Since we know the completed grade for those not attending school we include these in the calculation - this may be important since failure may discourage school attendance. In the appendix we provide a Table with our estimated probabilities of passing a grade.

4.3 The return to education and the terminal value function

As mentioned above, after age 17, we assume individuals work and earn wages depending on their level of education. In principle, one could try to measure the returns to education investment from the data on the wages received by adults in the village with different level of educations. However, the number of choices open to the individual after school include working in the village, migrating to the closest town or even migrating to another state. Since we do not have data that would allow us to model these choices (and schooling as a function of these) we model the terminal value function in the following fashion:

$$V(ed_{i,18}) = \frac{\alpha_1}{1 + exp(-\alpha_2 * ed_{i,18})}$$

where $ed_{i,18}$ is the education level achieved by age 18. The parameters α_1 and α_2 of this function will be estimated alongside the other parameters of the model and will be constrained to be non-negative.[6] Implicit in this specification is the assumption that the

[5]Since we estimate this probability from the data we could also allow for dependence on other characteristics.

[6]We have used some information on urban and rural returns to education at the state level along with some information on migration in each state to try to model such a relationship. Unfortunately, we have no information on migration patterns and the data on the returns to education are very noisy. This situation has motivated our choice of estimating the returns to education that best fit our education choices.

only thing that matters for lifetime values is the level of education achieved. All other characteristics, which we include in the model, are assumed to affect the achieved level of education and not its return. Finally, to check whether our estimate make sense we compare the implied returns to education with observed wage differentials in Mexico.

4.4 Value functions

Since the problem is not separable overtime, schooling choices involve comparing the costs of schooling now to its future and current benefits. The latter are intangible preferences for attending school including the potential child care benefits that parents may enjoy.

We denote by $I \in \{0,1\}$ the random increment to the grade which results from attending school at present. If successful, then $I = 1$, otherwise $I = 0$. We denote the probability of success at age t for grade ed as $p_t^s(ed_{it})$.

Thus the value of attending school for someone who has completed successfully ed_i years in school and is of age t already and has characteristics z_{it} is

$$\begin{aligned} V_{it}^s(ed_{it}|\Upsilon_{it}) &= u_{it}^s + \beta\{p_t^s(ed_{it}+1)E\max\left[V_{it+1}^s\left(ed_{it}+1\right), V_{it+1}^w\left(ed_{it}+1\right)\right] \\ &\quad +(1-p_t^s(ed_{it}+1))E\max\left[V_{it+1}^s\left(ed_{it}\right), V_{it+1}^w\left(ed_{it}\right)\right]\} \end{aligned}$$

where the expectation is taken over the possible outcomes of the random shock ε_{it} and where Υ_{it} is the entire set of variables known to the individual at period t and affecting preferences and expectations the costs of education and labour market opportunities. The value of working is similarly written as

$$V_{it}^w(ed_{it}|\Upsilon_{it}) = u_{it}^w + \beta E\max\left\{V_{it+1}^s\left(ed_{it}\right), V_{it+1}^w\left(ed_{it}\right)\right\}$$

The difference between the first terms of the two equations reflects the current costs of attending, while the difference between the second two terms reflects the future benefits and costs of schooling. The parameter β represents the discount factor. In practice, since we do not model savings and borrowing explicitly this will reflect liquidity constraints or other factors that lead the households to disregard more or less the future.

Given the terminal value function described above, these equations can be used to compute the value of school and work for each child in the sample recursively. These formulae will be used to build the likelihood function used to estimate the parameters of this model.

4.5 Wages and General Equilibrium Responses

Wages are the opportunity cost of education. In our model, an increase in wages will reduce school participation. Since such wages may be determined within the local labour market, they may also be affected by the program because the latter reduces the supply of labour children. This is going to be even more so the case if child labour is not highly substitutable with other types of labour. With our data we can estimate the effect of the program on wages and thus establish whether the change in the supply of labour does indeed affect them.

As wages are not observed for children who are not working, we need to estimate a wage equation so as to predict the missing values. We also need to correct for the measurement error of wages we do observe, as well as for their possible endogeneity.

One possibility would be to incorporate the wage equation in our school and work choice model and estimate it simultaneously. Here we follow a computationally simpler procedure: we estimate the wage equation separately, correcting for selection using the standard Heckman (1979) selectivity estimator; we then predict wages for all children in the sample, whether their wage is observed or not. The predictions do not include the Mills ratio used to correct for selection when estimating the coefficients.

For such an equation to be non-parametrically identified it is necessary to have variables that enter the participation equation and do not enter directly the wage equation and have broad enough support. The variables we use for such purpose are consistent with our model and include measures of the availability and cost of schools in the locality where the child lives. The determinants of child log wages are, in addition to a constant, the log of the adult agricultural wages in the village $\ln w_j^{ag}$, the education and the age of the child and an indicator of whether the village received the PROGRESA benefits P. With this specification we obtain :

$$\ln w_{ij} = \underset{(0.384)}{-0.983} + \underset{(0.028)}{0.0605}\, P + \underset{(0.049)}{0.883}\, \ln w_j^{ag} + \underset{(0.027)}{0.066}\, age_i + \underset{(0.0065)}{0.0116}\, educ - \underset{(0.053)}{0.056}\, Mills_i + \varpi_{ijt} \tag{5}$$

where $Mills_i$ represents the correction for selection which is obtained from the estimation of a reduced form school participation equation[7] and ϖ_{ijt} random factors including measurement error assumed independent of the observable characteristics. We report in parentheses below the estimates of the coefficients their standard errors. Although the $Mills$ ratio coefficient has the expected sign, implying that those who go to school tend to have lower wages, it is not significant, reflecting the probable fact that child labour is quite homogeneous given age and education. The age effect is significant and large as expected. The effect of education is very small, probably reflecting the limited types of jobs available in these villages.[8]

To interpret the equation above, let us suppose that production involves the use of adult and child labour as well as other inputs and that the elasticity of substitution between the two types of labour is given by ρ ($\rho > 0$). Suppose also that the price of labour is determined in the local labour market. Then the price of a unit of child labour in a locality can be written as

$$\log w_{child} = \frac{\rho + \gamma_{adult}}{\rho + \gamma_{child}} \log w_{adult} - \left[\frac{1}{\rho + \gamma_{child}} \log \left(\frac{L_{child}}{L_{adult}} \right) + \kappa \right] \tag{6}$$

In the above the $\gamma_k > 0$ (k = adult, child) are the adult and the child labour supply elasticities respectively and the L_k ($k = adult$, $child$) represent the level of labour supply of each group in the village.[9] The fact that the coefficient on the adult wage is smaller than one implies that the child labour supply elasticity is larger than the adult one. The adult agricultural wage (w_j^{ag}) is a sufficient statistic for the overall level of demand for goods in the local area and can thus explain in part the price of human capital provid-

[7]Hence the discrete dependent variable is zero for work and 1 for school.

[8]Overall the returns to education in Mexico are substantial, but they are obtained by the adults migrating to urban centres. we expect the children who progress in education to leave the village as adults so as to reap the benefits.

[9]This expression for the wage has been derived using as labour supply $H_k = L_k w_k^{\gamma_k}$ and production function $Q = \left[\delta H_{child}^{\sigma} + (1-\delta) H_{adult}^{\sigma}\right]^{\frac{1}{\sigma}}$ with $\sigma = \frac{\rho - 1}{\rho} < 1$,where $\rho > 0$ is the substitution elasticity.

ing the necessary exclusion restriction for identifying the wage effect in the education choice model. The term in square brackets in 6 is unobserved and reflects preferences for labour supply and technology (ρ and κ). These will, in general, be correlated with $\ln w_j^{ag}$through the determination of local equilibrium. Identification requires us to assume that L_{child}/L_{adult} as well as technological parameters are constant across localities, other than through the effect of the program, which will shift L_{child} resulting in the general equilibrium effects we are measuring. The additional regressors in 5 reflect differences in the amount of human capital supplied by each child.

We now turn to the effect of the program. This is captured by the "treatment" dummy P in equation (5), which decreases L_{child}. This pushes up the wage as a new local equilibrium is established. The program decreased child labour (increased schooling) by 3.3% and increased the wage rate by 6%, Taking into account that average participation is about 62% at baseline for our group, this implies an elasticity of wages with respect to participation (labour supply) of about -1.2. Thus, allowing for the general equilibrium effect of the policy can be potentially important, particularly if ρ is small. The estimation of the model will inform us, among other things on the sensitivity of school participation (and hence labour supply) to the child wage.

4.6 Habits and Initial Conditions

The presence of ed_{it} in equation 1 creates an important initial conditions problem because we do not observe the entire history of schooling for the children in the sample as we use a single cross section. We cannot assume that the random variable μ_i in equation 3 is independent of past school decisions as reflected in the current level of schooling ed_{it}.

To solve this problem we specify a reduced form for educational attainment up to the current date. We model the level of schooling already attained by an ordered probit with index function $h_i'\zeta + \xi\mu_i$ where we have assumed that the same heterogeneity term μ_i enters the prior decision multiplied by a loading factor ξ_i. The ordered choice model allows for thresholds that change with age, and is thus more general that the standard specification; we use this as an approximation to the sequential choices made before the programme.[10] The vector h_i includes variables reflecting past schooling costs such as the distance from the closest secondary schools in pre-experimental years. Since school availability, as measured by variables such as distance, changes over time, it can be used as an instrument in the initial conditions model that is excluded from the subsequent (current) attendance choice, which depends on the school availability during the experiment. We write the probability of $ed_{it} = e$ and of child i attending school as

$$\begin{gathered} P(ed_{it} = e, Attend_{it} = 1 | z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it}, \mu_i) = \\ P(Attend_{it} = 1 | z_{it}, x_{it}^p, x_{it}^s, wage_{it}, ed_{it}, \mu_i) \times P(ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \mu_i) \end{gathered} \tag{7}$$

This will be the key component of the likelihood function presented below. The endogeneity of the number of passed grades (the stock of schooling) is therefore captured by the common heterogeneity factor μ_i affecting both decisions The loading factor ξ governs the covariance between the two equations. It is important to stress the role played in identification by the variables that capture lagged availability of schools as variables that enter the initial condition equations but not the current participation equation.

5 Identification and Estimation

5.1 Identification

In the context of our model, there are two identification issues. The first concerns the identification of the model in the presence of habits and unobserved heterogeneity. In the previous section we stressed that identification can as far as this aspect is concerned, be achieved because of the presence of lagged school availability variables in the initial condition equation and the increase in school availability over time.

The second issues is about the identification of the effect of the program and the sources of variation in the data that allow us to measure such a parameter separately from the effects of the wage, which constitutes the usual opportunity cost of education. We already mentioned that a version of the model in which the effect of the grant can be identified from the control sample alone without using the program data at all: if we assume that the effect of the grant is the same (but opposite in sign) as the effect of the

[10] See Cameron and Heckman (1998) and Cunha, Heckman and Vavarro (2007) on the conditions under which a sequential dynamic optimisation problem can be represented as an ordered choice model.

wage, we can exploit variation across villages in child wages to estimate this parameter, under the conditions discussed in the section on wages. Here, however, we are interested in estimating a version of the model that allows these two parameters to be different. We therefore need variation in the grant received by individuals in our sample.

The grant available (or not) to the children in our sample varies for several reasons. First, some individuals live in treatment villages, while others live in control villages. This variation, given the random assignment of the program, is, by construction, exogenous and can help identify the parameter of interest. However, as we discussed in Section 3, there seems to be some differences in pre-program school enrollment between treatment and control villages, especially among ineligible families. We therefore might want to allow for such (unexplained) differences in pre-program enrollment between treatment and control via a 'treatment' dummy. Such a strategy sacrifices some of the variability induced by the experiment to measure the effect of the program, although the effect of the grant is still identified, in principle, by the fact that it varies with grade as well as by the comparison between eligible and ineligible individuals. However we do not wish to rely on this latter source of variation and clearly we want to put more weight on the exogenous experimental variation. We thus introduce a dummy variable for eligibility (whether in a treatment or control village). We then also note that the main source of pre-program differences are due to differences in the enrolment of the ineligibles across treatment and control villages. We therefore remove the treatment dummy an replace it with a dummy for an ineligible household living in a treatment village. In this version of the model the effect of the grant is identified by the comparison between eligible households in treatment and control villages and by the age variation of the grant, over and above the linear age effect we allow to affect preferences. Thus to summarize we present results of three specifications: In the first we only include a dummy for eligibility (only i.e. whether the household is poor or not); in the second we add a dummy for living in a treatment village; in the third we remove the treatment dummy and add a dummy for treatment village interacted with being ineligible, i.e. not among the poorest households.

The discount factor in this model as well as reflecting individual time preference also summarizes the impact of liquidity constraints, to the extent that these are important. Thus a low discount factor may be a reflection of an inability of households to transfer funds between periods. This would limit their behavioral response to promises of future grants. However, since the model relates to a discrete binary choice, the actual value of the discount factor is not identified - rather just the ratio of the discount factor to the standard deviation of the error is identified. This ratio, which we keep referring to as the discount factor in what follows, is identified by the variation induced in future opportunities by the age at which the policy is introduced. It need not be lower than 1 but it has to be positive.

We conclude this discussion stressing that, while functional form assumptions play a role as in all structural models, the experimental information is key to both the estimation of the model and the validation of the results. Without such variation we would need to rely exclusively on wage variation and we would not have this explicitly exogenous source of variation for the costs of schooling.

5.2 Estimation

We estimate the model by maximum likelihood.[11] Denote by $F(\cdot)$ the distribution of the *iid* preference shock ε_{it}, assumed logistic. Assume the distribution of unobservables μ_i is independent of all observables in the population and approximate it by a discrete distribution with M points of support s_m each with probability p_m, all of which need to be estimated (Heckman and Singer, 1984). The joint probability of attendance and

[11] To achieve the maximum we combine a grid search for the discount factor with a Gauss Newton method for the rest of the parameters. We did this because often in dynamic models the discount factor is not well determined. However in our case the likelifood function had plenty of curvature around the optimal value and there was no difficulty in identifying the optimum.

having already achieved e years of education ($ed_{it} = e$) can then be written as

$$\begin{aligned} P_i = P(Attend_{it} = 1, ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, wage_{it}) = \\ \sum_{m=1}^{M} p_m \{ F\{ u_{it}^s + \beta E \max \left(V_{it+1}^s (ed_{it} + I), V_{it+1}^w (ed_{it} + I) \right) \\ - \left(u_{it}^w + \beta E \max \left(V_{it+1}^s (ed_{it}), V_{it+1}^w (ed_{it}) \right) \right) | \mu_i = s_m \} \\ \times P(ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \mu_i = s_m) \} \end{aligned} \tag{8}$$

The probability of attending school conditional on μ_i and $ed_{it} = e$ is given by the $F(\cdot)$; the expectations inside the probability are taken over future uncertain realizations of the shock ε_{it} and the success of passing a grade; the sum is taken over all points of support of the distribution of unobserved heterogeneity. The expressions for the value functions at subsequent ages are computed recursively starting from age 18 and working backwards until the current age using the formulae in Section 4.4. The logistic assumption on the ε_{it} allows us to derive a closed form expression for the terms in equation 8

We should stress that the wage we use in the estimation is the value predicted by equation 5 (with the exclusion of the Mills ratio). Such an equation accounts for endogenous selection and takes into account the effect that PROGRESA had on child wages, so that it imputes a higher values for treatment villages.[12]

The log-likelihood function is then based on this probability and takes the form

$$\sum_{i=1}^{N} \{ D_i \log P_i + (1 - D_i) \log(1 - P_i) \} \tag{9}$$

where $D_i = 1$ denotes a school attendee in the sample of N children.

6 Estimation results

In this section, we report the results we obtain estimating different versions of the dynamic programming model we discussed above. We start by presenting estimates of the basic model and then discuss extensions and robustness checks we have performed.

[12] We do not correct the standard errors to take into account that the wage is a generated regressor.

In Tables 5 to 7, we present estimates of the three versions of the basic model we mentioned above: the first column of each table refers to the version that ignores differences in pre-program school enrollment between treatment and control villages, in the second column such differences are accounted for by a 'treatment dummy', while in the third they are accounted for by a dummy for non-eligible households in treatment localities. The second and third specifications aim at taking into account the observed pre-program differences between treatment and control variables. Given the nature of the evidence discussed in Section 3, which identified pre-programme differences in enrolment between treatment and control villages only among non-eligible households, we believe the most plausible specification is the third column that allows for such differences. For all specifications the discount factor was estimated to be 0.89. The standard errors we report are conditional on this value.

We estimate all the versions of the model on the sample of boys older than 9. All specifications include, both in the initial conditions equation and in the cost of education equation, state dummies, whose estimates are not reported for the sake or brevity. In addition, we have variables reflecting parental education (the excluded groups are heads and spouses with less than completed primary) and parents' ethnicity. We also include the distance from secondary school as well as the cost of attending such school, which in some cases includes fees. Finally all specifications include a dummy for eligibility (potential if in a control village- poor). This is, effectively, just a measure of wealth.

As mentioned above, we allow for unobserved heterogeneity that is modelled as a discrete variable with three points of support. The same variable enters, with different loading factors, both the utility of going to school equation and the initial condition equations. Such a variable, therefore, plays an important role in that it allows for a flexible specification of unobserved heterogeneity and determines the degree of correlation between the utility of schooling and completed schooling, which, by entering the equation for the current utility of schooling, introduces an important dynamic effect into the model. We therefore start reporting, in Table 5, the estimates of the points of support

	A	B	C
Point of Support 1	-16.064	-16.082	-15.622
	0.991	*0.994*	*0.975*
Point of Support 2	-19.915	-19.943	-19.457
	1.236	*1.241*	*1.218*
Point of Support 3	-12.003	-12.012	-11.616
	0.768	*0.769*	*0.758*
probability of 1	0.519	0.520	0.521
	0.024	*0.024*	*0.024*
probability of 2	0.405	0.405	0.403
	0.025	*0.025*	*0.025*
probability of 3	0.075	0.076	0.076
	-	-	-
load factor for Initial condition	-0.119	-0.118	-0.124
	0.023	*0.023*	*0.023*

Notes: Column A: Eligible dummy only; B: Treatment Village and eligible dummies C: Eligible dummy and non- Eligible in treatment village dummy. Asymptotic standard errors in italics

Table 5: The distribution of unobserved heterogeneity

of the unobserved heterogeneity terms, and that of the loading factor of the unobserved heterogeneity terms in the initial condition equation. Three points of support seemed to be enough to capture the observed heterogeneity in our sample.

The estimates in Table 5 reveal that we have three types of children, of which one is very unlikely to go to school and accounts for roughly 7.6% of the sample. Given that attendance rates at young ages are close to 90%, it is likely that these are the children that have not been attending primary school and, for some reason, would be very difficult to attract to school. Another group, which accounts for about 40.3% of the sample is much more likely to be in school. The largest group, accounting for 52.1% of the sample, is the middle one.

The loading factor of the unobserved heterogeneity term is negative (as expected) and precisely estimated. It implies that individuals more likely to have completed a higher level of schooling by 1997 are also more likely to be attending in 1998 due to unobserved factors.

The initial condition ed_{it} is modelled as an ordered probit with age specific cutoff points reflecting the fact that different ages will have very different probabilities to have completed a certain grade. Indeed, even the number of cutoff points is age specific, to allow for the fact that relatively young children could not have completed more than a certain grade. To save space, we do not report the estimates of the cut-off points.[13]

In addition to the variables considered in the specification for school utility, we include among the regressors the presence of a primary and a secondary school and the distance from the nearest secondary school in 1997. It is important to stress that these variables are included in the initial condition model only, in addition to the equivalent variables for 1998, included in both equations. As discussed above, these 1997 variables effectively identify the dynamic effect of schooling on preferences. It is therefore comforting that they are strongly significant, even after controlling for the 1998 variables. This indicates that in our there is enough variability in the availability of school between 1997 and 1998. Taking all coefficients together it seems that the most discriminating variable is the presence of a primary school in 1997.

The results also make sense: children living in villages with greater availability of schools in 1997 are better educated, children of better educated parents have, on average, reached higher grades, while children of indigenous households have typically completed fewer grades. Children from poor households have, on average, lower levels of schooling. As for the state dummies, which are not reported, all the six states listed seem to have better education outcomes than Guerrero, one of the poorest states in Mexico, and particularly so Hidalgo and Queretaro.

We now turn to the variables included in the education choice model, reported in the top panel of Table 7. All the variables, except for the grant and the wage are expressed as determinants of the cost of schooling, so that a positive sign on a given variable, decreases the probability of currently attending school. The wage is expressed as a determinant of the utility of work (so that an increase in wages decreases school attendance) and

[13]The discrete random term representing unobserved heterogeneity is added to the normally distributed random variable of the ordered probit, effectively yielding a mixture of normals.

	A	B	C
poor	-0.281	-0.283	-0.246
	0.027	*0.027*	*0.040*
PROGRESA Village	-	0.069	-
	-	*0.020*	-
PROGRESA village	-	-	0.060
non-eligible	-	-	*0.040*
Father's education			
Primary	0.188	0.188	0.189
	0.024	*0.024*	*0.024*
Secondary	0.298	0.297	0.298
	0.027	*0.028*	*0.028*
Preparatoria	0.596	0.597	0.596
	0.052	*0.052*	*0.052*
Mother's education			
Primary	0.168	0.167	0.169
	0.024	*0.024*	*0.024*
Secondary	0.328	0.316	0.329
	0.028	*0.028*	*0.028*
Preparatoria	0.296	0.297	0.286
	0.058	*0.058*	*0.056*
indigenous	0.002	0.003	0.001
	0.024	*0.024*	*0.024*
Availability of Primary 1997	0.365	0.378	0.367
	0.070	*0.070*	*0.070*
Availability of Secondary 1997	0.744	0.726	0.741
	0.173	*0.171*	*0.173*
Km to closest secondary school 97	0.001	0.001	0.001
	0.003	*0.003*	*0.003*
Availability of Primary 1998	-0.134	-0.173	-0.144
	0.113	*0.117*	*0.117*
Availability of Secondary 1998	-0.789	-0.762	-0.783
	0.172	*0.171*	*0.172*
Km to closest secondary school 98	-0.007	-0.006	-0.006
	0.003	*0.003*	*0.003*
Cost of attending secondary	0.005	0.008	0.004
	0.022	*0.022*	*0.022*

Notes: as in Table 5. State dummies included
Availability means school in the village.

Table 6: Equation for Initial conditions

the grant is a determinant of the utility of schooling, so that an increase in it, increases school attendance. In addition, the coefficient on the grant is expressed as a ratio to the coefficient on the wage, so that a coefficient of 1 indicates that a unitary increase in the grant has the same effect on the utility of school as an increase in the wage has on the utility of work.[14]

The key parameters of the model from a policy perspective are the wage coefficient and the coefficient of the grant itself. An increase in the wage decreases the probability of attending school. On average, the effect of reducing the wage by 44% (which would roughly give a reduction similar to the average grant to beneficiaries) increases the probability of attending school by 2.1%. This effect cannot be inferred directly from the value of the parameter alone and has been obtained from the simulations of the model that we discuss in detail below.

The value of the grant varies by treatment and control villages (where of course it is zero) and by grade the child could be attending.[15] As mentioned above, the coefficient of the grant is expressed as a fraction of the wage coefficient. The values between 3 and 4 for this coefficient reported in Table 7 indicate that the effect of the grant on school attendance is considerably larger than the effect of the wage.[16] This contrasts to the assumption made in the model by Todd and Wolpin (2006), where the effect of the grant is assumed to be equal and opposite to the effect of the wage (suitably scaled) [17]

In terms of background characteristics, belonging to a household with less educated parents leads to lower attendance. This may be a reflection of liquidity constraints or of different costs of schooling. Perhaps surprisingly, the coefficient on poor (eligible)

[14]The wage has been scaled to be interpreted as the earnings corresponding to the period covered by the grant. Thus the effects are comparable.

[15]As mentioned above, in secondary school, the grant is also higher for girls than for boys. However, as we only estimate the model for the latter, we do not exploit this variation.

[16]The non-linearity of the model implies that the effect of enrolment is not 3 or 4 times larger than the effect of enrolment of the wage.

[17]The likelihood ratio test is 30 and is distributed $\chi^2_{(1)}$ under the null. Since this hypothesis relates to a non-linear restriction and subject to a variety of different normalisations the Wald test is not appropriate. See Gregory and Veall (1985)

	A	B	C
wage	0.103	0.109	0.110
	0.033	*0.033*	*0.033*
PROGRESA Grant	4.154	3.218	3.869
	1.386	*1.129*	*1.036*
parameter in terminal function $\ln(\alpha_1)$	5.798	5.811	5.797
	0.099	*0.099*	*0.099*
parameter in terminal function $\ln(\alpha_2)$	-1.104	-1.113	-1.103
	0.020	*0.021*	*0.020*
Poor	0.159	0.106	-0.128
	0.118	*0.120*	*0.164*
In PROGRESA village		-0.171	
		0.106	
In PROGRESA village & ineligible			-0.492
			0.198
Father's Education - Default is less than primary			
Primary	-0.131	-0.137	-0.133
	0.093	*0.093*	*0.092*
Secondary	-0.271	-0.270	-0.267
	0.113	*0.113*	*0.112*
Preparatoria	-0.751	-0.755	-0.735
	0.253	*0.253*	*0.250*
Mother's Education - Default is less than primary			
Primary	-0.136	-0.131	-0.131
	0.096	*0.096*	*0.096*
Secondary	-0.137	-0.139	-0.129
	0.113	*0.114*	*0.113*
Preparatoria	-0.920	-0.918	-0.893
	0.284	*0.286*	*0.283*
indigenous	-0.605	-0.598	-0.591
	0.107	*0.108*	*0.107*
Availability of Primary 1998	2.734	2.701	2.744
	0.220	*0.220*	*0.219*
Availability of Secondary 1998	-0.160	-0.164	-0.189
	0.152	*0.153*	*0.152*
Km to closest secondary school 98	0.001	0.001	0.001
	0.004	*0.004*	*0.004*
Cost of attending secondary	0.005	0.005	0.005
	0.001	*0.001*	*0.001*
Age	2.772	2.790	2.758
	0.180	*0.183*	*0.180*
Prior Years of education	-2.995	-3.005	-2.996
	0.214	*0.216*	*0.214*
log-Likelihood	-26676.676	-26668.927	-26672.064

Notes as in Table 5. Discount rate $\beta = 0.89$ State dummies included

Table 7: Parameter estimates for the Education choice model

households is not significantly different from zero, while that on indigenous households indicates that, ceteris paribus, they are more likely to send their children to school. The states that exhibit the higher costs are Queretaro and Puebla, followed by San Luis Potosi and Michoacan, while Veracruz and Guerrero are the cheapest. The costs of attending secondary measured in either distance (since many secondary schools are located in neighbouring village) or in terms of money have a significant and negative effect on attendance.

Both age and grade have a very important effect on the cost of schooling, age increasing and grade decreasing it. The coefficient on age is the way in which the model fits the decline in enrolment rates by age across both treatment and control villages. The effect on the grade captures the dynamics of the model and, as we discussed, is identified by the presence of lagged supply of school infrastructure. We find a very strong effect of state dependence despite the fact that we include age as one of the explanatory variables. The pre-existing level of education is a critical determinant of choice, with increased levels of education having a substantially positive effect on further participation. This is of course an important point since it provides an additional mechanism by which the subsidy may increase educational participation and is likely to be key for simulating alternative policies. It also implies different effects depending on the amount of prior educational participation. The likelihood ratio test for the null that past education does not matter has a p-value of zero.[18] Moreover, eliminating habits leads to an overprediction of the program effects.

As we discussed above, we have scant information on the returns to education, so that we include the returns to education among the parameters to be inferred from the differential attendance rates in school. It is therefore important to check what are the returns to education implied by this model. When we compute the return to education implied by the coefficients on the terminal value function, we obtain an estimate of the average return to education of 5% per year, maybe a bit low in the Mexican context, but above the average return to education observed in our villages.[19]

[18] The likelihood ratio statistic is 300.

[19] The return is calculated as $r = \frac{\partial V_T}{V_T \partial ed}$ where $V_T = V(ed_{i,18})$ and where the parameters α_1 and α_2 are

Cost variables have the expected sign and are significant. For instance, an increase in the distance from the nearest secondary school, significantly decreases the probability of attending school. Likewise for an increase in the average cost of attending secondary school. As several of the children in our sample are still attending primary school we also tried to include variables reflecting the cost and availability of primary schools, but we could not identify any significant effect.

The three versions of the model we present differ because of the presence, in column B and C of Table 7 of a dummy for PROGRESA villages (col B) and one for PROGRESA non-eligible. As we discussed above, these are meant to capture pre-program differences that were identified to be present in enrolment rates despite the random assignment of the program. In Section 3, we showed that most of the pre-program differences are explained by differences among non eligible households. This evidence would justify the specification in Column C. The specification in Column B takes a more conservative approach and controls for the fact of living in a PROGRESA village. This could be too conservative as it absorbs much of the experimental variability of the grant.

Comparing the results across specifications, we see that most coefficients do not change much. Not surprisingly, the only coefficient that exhibits some variation is the one on the grant. And even that does not change much between Column A and C. The size of the coefficient on the grant, however, is somewhat reduced in column B, which, therefore, provides a lower bound on the effect of the program.

The model is complex and non linear. To quantify the effect of the main variables of interest from a policy point of view and to assess how the grant operates and what its effect are we now turn to the analysis of some simulations.

7 Simulations

We now use the model to simulate school participation under different scenarios. The first exercise is simply aimed at evaluating the effect of the PROGRESA grant as designed. For

given in their logs in column 3 of Table 7.

such a purpose we compute, for each child in the sample, the probability of enrolment with and without the grant. The effect is then computed averaging the differences between these probabilities. We perform the exercise twice, first without taking into account the general equilibrium effects (that is, eliminating the grant, but without changing the wage) and then incorporating such effects. The latter is done by reducing children wages in treatment communities when the grant is removed. The amount by which the wage is reduced, can be read directly from equation 5.

In Figure 1, the continuous line plots these averages by age for each of the three versions of the model estimated in Table 7, keeping the wage equal to the value observed following the experiment. The dotted line adjusts the wage for the effect of the program by lowering the wage in the non-grant state. This effect is smaller, but not by much because the wage effect in the model is small (albeit significant). The dotted line is the one that is most comparable to what is obtained in the experimental evaluation, and indeed it is closer to it. The effect of the PROGRESA grant is very small at the youngest ages and peaks, for the three figures age 15. Perhaps not surprisingly, the effect is smaller when we introduce a 'treatment dummy' in the model. Notice that in this case the effect of the grant is effectively identified exclusively by its variation across grades and by the non-linearities embedded in the model. The other two models, however, (the one without any 'treatment dummy' - A in the table and the one with a 'treatment non-eligible' dummy -C in the table) yield similar results. Reassuringly, these results are very similar to the difference-in-differences results described in Section 3. This indicates that the model fits the data reasonably well: our dynamic model is able to fit with a parsimonious specification the pattern of effects by age groups we obtained in Section 3. It should be stressed that the impacts are driven by a single coefficient, the one on the grant. And yet our impacts line up remarkably well with the several parameters estimated in Section 3.

One of the main advantages of having a structural model, however, is the possibility of performing policy experiments. The first we perform is to compare the effect of the current program with that of a similar program that differs in the way in which the grant

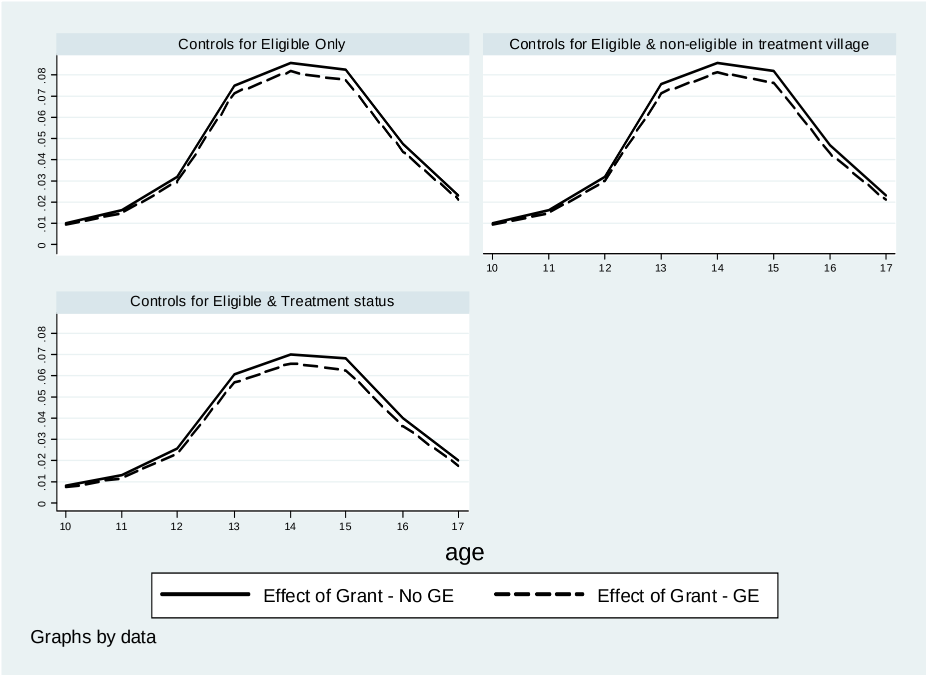

Figure 1: The Effect of the Grant in partial and General Equilibrium

varies with the grade attended.

The aim of this exercise is to compare the impact of the current grant, as predicted by our model to the impact that is obtained when the structure of the grant is changed so as to target the program to those in the most responsive ages. In particular, we focus on a 'balanced budget' reform of the program. That is, we increase the grant for children above grade 6 and set it to zero for children below that grade. The increase for the older children is calibrated so that, taking into account the response of children schooling choices to the change, the overall cost of the program is left unchanged (at least for our sample).

We plot the result of this exercise in the same Figure 2 where again we show the results with no wage adjustment (continuous line) and with a GE adjustment. Performing the GE adjustment is now a bit more complicated than in the previous exercise. The amount by which children wages would change with the counterfactual grant structure has to be extrapolated. We do that by using the elasticities discussed in Section 4.5.

The graph shows that the impact on school participation can almost double by targeting the grant to the older children, with no effect on the school participation of the younger primary age children. This is not surprising since the grant hardly changes their behavior in the first place. However, it does show that the older children are sufficiently sensitive to economic incentives for a redesign of the policy to have major effects. Obviously the grant on the younger children has other effects, for example on health and cognitive as well as physical development. Of course, one should note that almost all children go to school below grade 6, which means that the early part of the grant does not really have a conditional nature. In removing the grant form these children we are assuming that the sole purpose of the policy is to induce children into school. If this was considered as a general anti-poverty measure disguised as a schooling subsidy this policy trade-off, whose effect we are simulating may not be available. Notice however, that the overall amount of resources transferred to eligible households is the same in the two cases. Nevertheless, it is interesting to consider the cost such a policy in terms of future schooling participation by restructuring the grant, offering it to the group where behavioral effects are most important. Allowing for GE effects here makes a difference of only about 1/2 a percentage point.

We next consider a number of alternative experiments. In particular we consider the effect of the wage on school participation, the effect of distance to school and the impact of unobserved costs. All three experiments are summarized in figure 3. In all cases we use the model which includes a PROGRESA dummy for the eligible individuals only. No grant is our baseline.

First we decrease the wage by an amount equivalent to the grant.[20] We see that the effect of the wage is estimated to be lower than the grant; for example at age 15 the incentive effect is half the one in Figure 1. This is an important point and puts into question the possibility to evaluate programs of this kind using an ex ante evaluation

[20]The reduction is proportional so as to give an average amount equivalent to the grant. The grant however is additive. So we would not expect the effects of the wage to be distributed equally because the wage varies with age much more steeply than the grant.

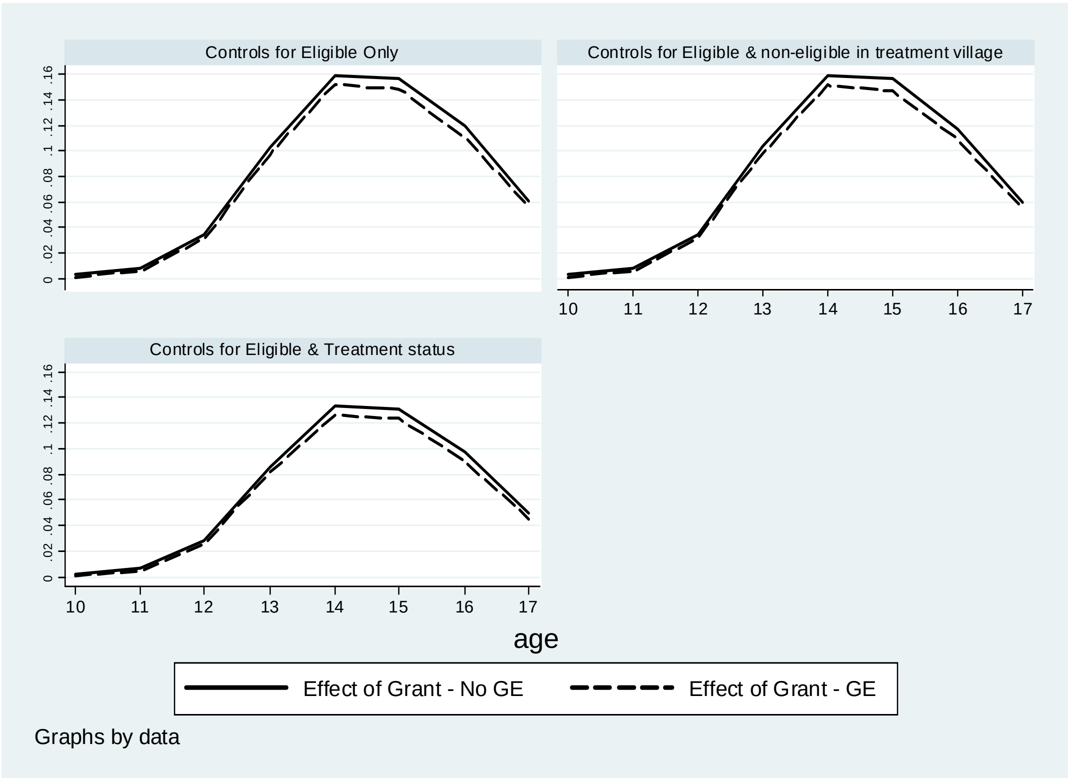

Figure 2: The Effect of redistributing the Grant to children above grade 6 only

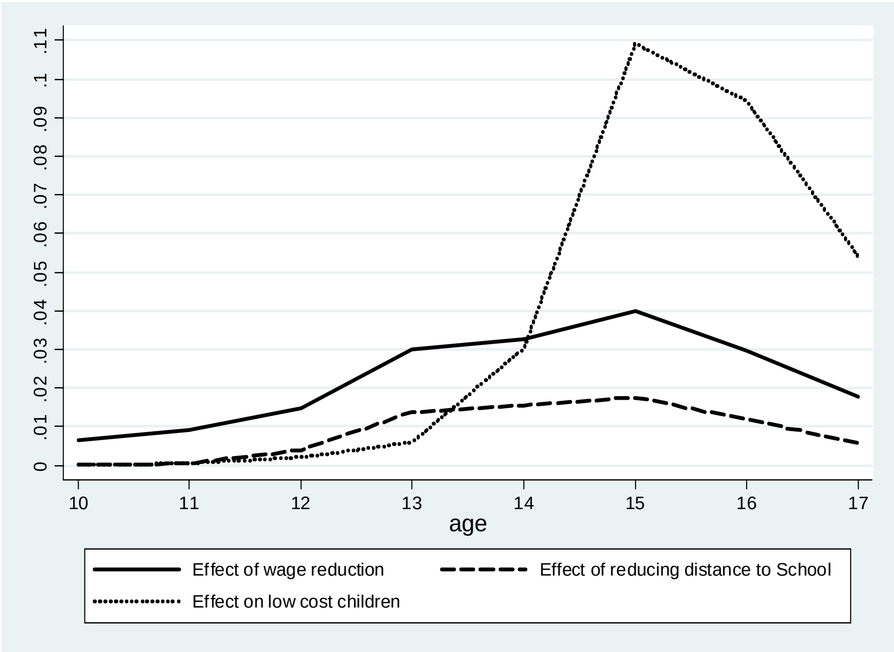

Figure 3: A Comparison of the Effects of Alternative Policies

model that relies on observable market data, rather than on experimental/pilot data, unless we are able to understand at a deeper level the reason why the wage and the grant have different effects. We believe this is because of the different ways these two incentives, the wage and the grant, act on intrahousehold

In the next experiment we demonstrate the effects of a potential school building program that would reduce the distance of secondary schools to no more than 3 km. We consider this because it would constitute an alternative policy to subsidizing participation (although we do not claim that this policy is equivalent in terms of cost or in terms of other benefits such as better nutrition and its impact). The effect is modest as it would increase participation by just below 2 percentage points at age 15.

Finally in order to demonstrate the potential importance of targeting we show the effects of the policy on individuals with "low cost" of schooling; i.e. we evaluate the effect at the point of support of the unobserved heterogeneity with the highest school participation. As shown in the graph for these individuals there is no effect of the program before age 14. Beyond that the effects are higher than average. Conversely, for the high cost of schooling individuals (not shown) the effects are higher than average at younger ages and lower later.

8 Conclusions

In this paper we demonstrate the power of using an economic model to analyse data from a major social experiment, namely PROGRESA in Mexico. Conversely, we also show the usefulness of using experimental data to estimate a structural economic model. The welfare program we consider was originally applied to rural Mexico and aimed, among other things, at increasing school enrolment rates among poor children.

We start by showing some of the effects of this program on school participation. As the program was randomised across villages, inducing truly exogenous variation in the incentives to attend school, by offering a subsidy for attendance, estimation of its impact is relatively straightforward. After presenting some evidence based on simple treatment/control comparisons and differences in differences, we argue that many questions are left unanswered by this type of techniques and propose the use of a structural model.

By interpreting the data through the viewpoint of a dynamic model one can investigate how the impacts of the program would be different if one were to change some of its parameters. In particular, we analyze what would happen to the program's impact on enrolment by changing the structure of the grants with age. This type of analysis is instrumental to the design of effective interventions and cannot be performed without a well specified behavioural model. We can also compare the impacts of the program with alternative policies, such as a program that would reduce the distance of the households in our sample from the nearest secondary school. Therefore we show that only through this kind of modelling approach is it possible to answer questions that are more general than the specific focus of the evaluation.

But this is not a one way street: we also argue that the use of the experimental data and the genuine exogenous variation it induces, allows us to identify models that are much richer than those that could be identified based on standard observational data. In our specific context, we show that, although the program operates by changing the relative price of education and its opportunity cost, the economic incentives it provides are much larger than those provided by changes in children wages. This is important from a policy point of view, but also for modelling if one notes that without the experimental variation this type of model could not be estimated. Our approach provides a rich set of results and at the same time points to the importance of the interaction between economic models and social experiments for the purposes of evaluation and understanding behaviour. In this sense our approach is in the spirit of Orcutt and Orcutt (1968).

The experimental design of the evaluation data, where the program was randomized across communities, and the fact that the localities in the sample are relatively isolated, allows one to estimate the 'general equilibrium' effect that the program has on children wages. Having estimated these effects, we incorporate this additional channel in our structural model, both at the estimation stage (where our estimated wage equation incorporates the g.e. effects) and in performing the simulation of alternative interventions.

The model we estimate fits the data reasonably well and predicts impacts that match the results obtained by the experimental evaluation closely. They indicate that the program is quite effective in increasing the enrolment of children at the end of their primary education, a fact that has been noticed in several evaluations of conditional cash transfers in Latin America. On the other hand, the program does not have a big impact on children of primary school age, partly because enrolment rates for these children are already quite high.

We identify relatively large 'general equilibrium' effects of the program on child wages: in the treatment localities child wages are about 6% higher than in control localities. Remarkably, this fact had not been noticed in the large empirical literature on PROGRESA. Although these effects and the impact that the program is observed to have on school enrolment (and child labour supply) imply relatively large elasticities of wages to changes in enrolment, the attenuating effect on the impact of the grant is not large.

The heterogeneity of the program impacts by age has suggested the possibility of changing the structure of the grant, reducing or eliminating the primary school grant and increasing the size of the grant for secondary school. This proposal, that has been considered in urban Mexico and in Colombia, can be analyzed with the help of our structural model. The simulations we perform indicate that the effect on school participation could be much improved by offering more resources to older children and less to relatively younger one. By taking into account behavioural responses, we simulate changes to the program that result in the same amount of resources spent by the program and yet obtain much larger impacts on school enrolment of older children.

Some words of caution are obviously in order when considering the results of our simulations. It should be pointed out, for instance, that our model is silent as to the effect of the grant on other dimensions of child development. Since most young children go to school anyway, the grant for them is effectively unconditional and can have other effect such as on their nutrition and resulting physical and cognitive development. Although the reformulated grant could transfer the same resources to the families whose children continued school attendance, albeit with different timing the effects on these other dimensions we do not model could be quite different in the presence of liquidity constraints; moreover fewer resources would end up with those who despite the increased grant drop out of school anyway. Nevertheless, the point remains that in terms of school participation, the joint use of the experimental data and the model suggests that a different age structure of the grant would achieve substantially different results.

9 References

Behrman, J. (1999): "Education, Health and Demography in Latin America Around the End of the Century: What Do We Know? What Questions Should Be Explored?", IADB Mimeo.

Behrman, J., Duryea, S. and M. Székely (1999): "Schooling Investments and Aggregate Conditions: A Household Survey-Based Approach for Latin America and the Caribbean", OCE Working Paper No 407, IADB, Washington:

http://www.iadb.org/res/publications/pubfiles/pubWP-407.pdf.

Behrman, J., Duryea, S. and M. Székely (2000): "Households and Economic Growth in Latin America and the Caribbean", GDN, Mimeo.

Behrman, J. and P. Todd (1999); "Randomness in the Experimental Samples of PROGRESA (Education,Health, and Nutrition Program)", IFPRI. downloadable from: http://www.ifpri.org/themes/PROGRESA/pdf/BehrmanTodd_random.pdf .

Blundell Richard, Costas Meghir and Pierre Andre Chiappori (2005) Collective Labour Supply with Children Journal of Political Economy, Volume 113, Number 6, December 2005

Cameron, Stephen V. and James J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males" *The Journal of Political Economy*, Vol. 106, No. 2 (Apr., 1998), pp. 262-333

Cunha, Flavio, James. J. Heckman and Salvador Navarro (2007) "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds" *International Economic Review* Vol. 48, No. 4, November 2007

Dearden, L. C. Emmerson, C. Frayne and C. Meghir (2008) "Can Education Subsidies stop School Drop-outs? An evaluation of Education Maintenance Allowances in England", IFS discussion paper forthcoming *Journal of Human Resources*.

Gregory, A.W. and M.R. Veall (1985), "Formulating Wald Tests of Nonlinear Restrictions", *Econometrica*, vol. 54, no. 6, November (1985), 1465-1468.

Heckman, J.J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*

Heckman, J.J and B. Singer (1984) "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data"

Econometrica Vol. 52, No. 2 (Mar., 1984), pp. 271-320

Keane Michael P. and Kenneth I. Wolpin (1997) The Career Decisions of Young Men *The Journal of Political Economy* Vol. 105, No. 3 (June 1997), pp. 473-522.

Orcutt, G. H., and A. G. Orcutt (1968) Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes," *American Economic Review*, 58, 754-772.

Rawlings, L. (2004): “A New Approach to Social Assistance: Latin America’s Experience with Conditional Cash Transfer Programs”, World Bank Social Protection Discussion Paper No 0416.

Schultz, T.P. (2003): “School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program” , *Journal of Development Economics,*

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.”, IFPRI, downloadable from: http://www.ifpri.org/themes/PROGRESA/pdf/Skoufias_finalsyn.pdf

Todd , P. and K. Wolpin (2006): “Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico”, *American Economic Review*, 96(5): 1384–1417

Willis R. and S. Rosen (1979) “Education and Self Selection”, *Journal of Political Economy,* 87, pp S7-S36.