Institute for
Fiscal Studies

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

IFS Working Paper 10/14

Orazio Attanasio
Costas Meghir
Ana Santiago

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

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

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

July 2010
(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.

1

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. In the first version of the program, which has since evolved and is now called Oportunidades, the grants started in third grade and increased until the ninth and were 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 benefi-

ciaries (i.e. the poorest) 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.

By all accounts and evidence, the evaluation of PROGRESA, based on the large randomized experiment described above, was highly successful (see Schultz, 2003). Given the evaluation design, the program impacts could be estimated by comparing mean outcomes between treatment and control villages. These estimates, on their own, 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, one can think of using experimental variation and indeed designing experiments to generate such variability so as to estimate more credibly structural models capable of richer policy analysis. Our work draws from the tradition of Orcutt and Orcutt (1968), who advocate precisely this approach, which found an early expression in the work based on the negative income tax experiments, such as in Burtless and Hausman (1978) and Moffitt (1979).

Thus, the aim of this paper is to analyze the impact of monetary incentives on education choices in rural Mexico; to discuss effective design of interventions aimed at improving educational participation in developing countries; and to illustrate the benefits of combining randomised experiments with structural models. To achieve these goals we estimate a simple structural model of education choices using the data from the PROGRESA randomised experiment. We then use the model to simulate the effect of changes to some of the parameters

3

of the program.

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, by Todd and Wolpin (2006). They then use the experiment to validate their model, which they estimate using the data from the control villages, as if no experiment ever took place. Their approach does not require experimental variation other than as a check for their model. Thus, while both papers use the same data, the idea underlying the two research programs is different.[1] 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.

Like Todd and Wolpin (2006), we estimate a structural model, but our approach and objectives are different. We use both treatment and control villages exploiting the variation in the grant induced by the randomized experiment to estimate a more flexible specification than could not be estimated without the program: critically, we do not restrict the effect of the grant to be the same

---

[1] There are also differences in the estimation approach.as well as in the specification of the models between this paper and that of Todd and Wolpin. 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 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.

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 is the case empirically and, if necessary, to relax this separability restriction. The point is important because it adds a new dimension to the discussion of the role of randomised experiments in designing policy. If indeed separability is rejected, experimental variation is useful not only because of the exogenous variation that it induces but also because it helps identify economic effects that may otherwise be unidentifiable and that play a crucial role in the design of policy. Thus, the structural model allows for a richer analysis of the experimental data and the experimental data allows for a richer economic model to be estimated.

Our model is similar to that of Willis and Rosen (1979) where individuals base their choice on comparing the costs and benefits of additional schooling depending on their comparative advantage. Our estimation approach is also 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 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.

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. Although some papers in the literature (see, for instance, Angelucci and De Giorgi, 2009) 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. One could imagine that, if the program is effective in increasing school participation, 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. Here we estimate this impact (taking into account the fact that children wages are observed only for the selected subset of children who actually work) and establish that the program led to an increase in child wages in the treatment municipalities, by decreasing the labour supply of children. We incorporate these general equilibrium impacts within our model and in our simulations.[2]

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

---

[2] The consideration of GE effects constitutes yet another difference relative to Todd and Wolpin (2006).

choices and describe its various components. Section 5 briefly discusses the estimation of the model. Section 6 presents the results we obtain from the estimation of our model. We also report the results of a version of the model that imposes the separability restriction used by Todd and Wolpin (2006). Section 7 uses the model to perform a number of policy simulations that could help to fine-tune the program. Finally, Section 8 concludes the paper with some thoughts about open issues and future research.

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 test-

ing. 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

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: The PROGRESA grants

entitled or not to the program.

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.

At the time of the 1997 survey, each household in the treatment and control villages was defined either as eligible or non eligible. Subsequently, in March 1998 before the start of the program, some of the non-eligible household were re-classified as eligible. However, a considerable fraction of the newly eligible households, due to administrative delays, did not start receiving the program until much later. In some of the results we present below, we distinguish these households. In the estimation of the structural model we consider as beneficiary a household that actually receives the program.

3   Measuring the impact of the program: treatment versus control villages.

As PROGRESA was assigned randomly between treatment and control villages during the expansion phase of the program, it is straightforward to use the evaluation sample to estimate the impact of the conditional cash transfers on school enrolment. Randomization implies that control and treatment sample are statistically identical and estimates of program impacts can be obtained by a simple comparison of means. However, such an exercise estimates the impact of the program as a whole, without specifying the mechanisms through which it operates.

The availability of baseline, pre-program data, allows one to check whether the evaluation sample is balanced between treatment and control groups both in terms of pre-program outcomes and in terms of other observable background characteristics. This exercise was performed by Behrman and Todd (1999), who explored a wide range of variables at baseline. The data includes information on programme eligibility for both treatment and control villages at baseline. This allows us to make the comparisons separately for eligible and non-eligible households. Behrman and Todd (1999) indicate that, by and large, the treatment and control samples are very well balanced. However, and annoyingly, there seem to be some pre-program differences in school enrolment among non eligible households. While it is not clear why such a difference arises, it might be important to control for these initial differences when estimating impacts.

In this section, we present some estimates of the impact of PROGRESA on school enrolment. These impacts have been widely studied: the IFPRI (2000) report estimates of program impacts on a wide set of outcomes, while Schultz

(2003) presents a complete set of results on the impact of the program on school enrolment, which are substantially similar to those presented here. Here our focus is on 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 impacts of the program in the sample we use to estimate our structural model, we set the mark against which it will be fitted.

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 varies with age, we also report the results for different age groups, although when we consider individual ages, some of the estimates are quite imprecise.

In Table 2 we report the estimated impact for the boys of each age obtained comparing treatment and control villages in October 1998. In the last two rows, we also report the average impacts on boys aged 12 to 15 (which is an age group on which Todd and Wolpin (2006) focus) and on boys aged 10 to 16, that is our entire sample. In the first column of the Table, we report enrolment rates among eligible (as of 1997) boys in control villages. In the second column, we show the estimated impacts obtained for boys from households that were declared eligible in 1997 (poor 97). In the third column, we report the results for the boys in all eligible households, including those reclassified in March 1998. Finally, in the third column, we report the impacts on the non-eligible children.

The experimental impacts show that the effect of the PROGRESA program on enrolment has a marked inverted U shape. The program impact is small and not significantly different from zero at age 10. It increases considerably past

age 10, to peak at age 14, where our point estimates indicates an impact of 14 percentage points on boys whose households were classified as eligible in 1997. The impact then declines for higher ages, probably a consequence of the fact that the grant was not available, in the first version of the program, past grade 9. The average impact for the boys in our sample (aged 10 to 16) is about 5 percentage point, while for the boys aged 12 to 15 is, on average, as high as 6.6%.

The impact on the households classified as poor in 1997 is slightly higher than on all eligible households, probably a reflection that the impact might be higher for poorer families and the fact that, some of the families that were reclassified as eligible in March 1998 (after being classified as non eligible in 1997), did not receive the program immediately, due to administrative difficulties.

A surprising feature of Table 2 is the measured impact on non eligible children. Although noisy, the estimates for some age groups indicate a large impact on non-eligible boys. Indeed, for the age group 12-15, the effect is even larger than for the eligible children, at almost 8%. While one could think of the possibility of spill-over effects that would generate positive effects on non-eligible children, the size of the impacts we measure in Table 2 is such that this type of explanation is implausible. As we mentioned above, however, if one compares school enrolment rates in 1997 between treatment and control villages, one finds that, among non-eligible households, they are significantly higher (statistically and substantially) in treatment villages than in control villages. This is particularly so for children aged 12 to 17. Instead, enrolment rates in 1997 among eligible children, are statistically identical in treatment and control villages. It is therefore possible that the observed difference in enrolment rates among non-

eligible households is driven by pre-existing differences between treatment and control towns.

The reason for the difference in enrolment rates among boys (and girls) in non eligible households between treatment and control villages is not clear. Within our structural model, we account for it by considering one specification which incorporates an unobserved cost component for non-eligible households in control villages. As for measuring the effect of the program as in Table 2, one can use the 1997 data to obtain a difference in difference estimates of its impacts. We report the results of such an exercise in Table 6 in the Appendix. In this table, the pattern of the impacts among eligible children is largely unaffected (as to be expected given the lack of significant differences between treatment and control villages at baseline for these children). The impacts on the non-eligible children, however, become insignificant. This evidence justifies the interpretation of the evidence in the last column of Table 2 as being caused by pre-existing unobservable differences for non eligible children and justifies our use of a 'non eligible control' dummy in our empirical specification.

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

Difference estimates of the impact of PROGRESA on boys school enrolment				
Age Group	enrolment rates in control villages (eligible)	Impact on Poor 97	Impact on Poor 97-98	Impact on non-elig.
10	0.951	0.0047 (0.013)	0.0026 (0.011)	0.0213 (0.021)
11	0.926	0.0287 (0.016)	0.0217 (0.015)	-0.0195 (0.019)
12	0.826	0.0613 (0.024)	0.0572 (0.022)	0.0353 (0.043)
13	0.780	0.0476 (0.030)	0.0447 (0.027)	0.0588 (0.060)
14	0.584	0.1416 (0.039)	0.1330 (0.035)	0.0672 (0.061)
15	0.455	0.0620 (0.042)	0.0484 (0.039)	0.1347 (0.063)
16	0.292	0.0304 (0.038)	0.0355 (0.036)	0.1063 (0.067)
12-15	0.629	0.0655 (0.027)	0.0720 (0.024)	0.0668 (0.022)
10-16	0.708	0.0502 (0.018)	0.0456 (0.015)	0.0810 (0.026)
Standard errors in parentheses are clustered at the locality level.				

Table 2: Experimental Results October 1998

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

18

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 child-care services that the school provides during the working

19

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:

$$u_{it}^s = Y_{it}^s + \alpha g_{it} \tag{1}$$

$$Y_{it}^s = \mu_i^s + a^{s0} z_{it} + b^s ed_{it} + 1(p_{it} = 1)\beta^p x_{it}^p + 1(s_{it} = 1)\beta^s x_{it}^s + \varepsilon_{it}^s$$

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

20

and, therefore, the utility cost of schooling. Finally, the term μ_i^s represents unobservables which we assume have a constant impact over time.[3]

The utility of not attending school is denoted by

$$
\begin{aligned}
u_{it}^w &= Y_{it}^w + \delta w_{it} & (2) \\
Y_{it}^w &= \mu_i^w + a^{w0} z_{it} + b^w ed_{it} + \varepsilon_{it}^w
\end{aligned}
$$

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:

$$
\begin{aligned}
u_{it}^s &= \gamma\delta g_{it} + \mu_i + a^0 z_{it} + bed_{it} + 1(p_{it}=1)\beta^p x_{it}^p + 1(s_{it}=1)\beta^s x_{it}^s + \varepsilon_{it} & (3) \\
u_{it}^w &= \delta w_{it} & (4)
\end{aligned}
$$

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

---

[3] 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.

to be perfectly foreseen when individuals consider trade-offs 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:

$$
\begin{aligned}
U^s &= c^s = Y + g \\
U^w &= \theta c^w - \alpha = \theta(Y + w) - \alpha
\end{aligned}
\tag{5}
$$

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.[4]

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

---

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

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 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 preprogram 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,[5] 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.

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.[6] 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.

---

[5]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.

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

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.[7] Implicit in this specification is the assumption that the 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.

---

[7] 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.

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

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

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

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

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 labour supply of children. These general equilibrium effects can be even more pronounced if child labour is not sufficiently 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.

In what follows, we need to estimate a wage equation for three reasons. First, we do not observe wages for children who are not working. Second, the dynamic programming model requires the individual to predict future wages; this is done on the basis of a model of wages perceived by the individual. Third, we wish to test for general equilibrium effects by estimating the effect of the program on wages. This is important because GE effects can dampen the effects of the program.

We thus specify a standard Mincer type wage equation, where the wage of a boy i living in community j determined by his age and education according to

$$\ln w_{ij} = q_j + a_1 age_i + a_2 educ_i + \omega_{ij} \tag{6}$$

where q_j represents the log price of human capital in the locality. We estimate this wage equation separately from the rest of the model. We then use predictions from this equation in place of actual wages. As far as future wages are

28

concerned, this approach assumes that within our model individuals use these predictions as point estimates of future wages and ignore any variance around them. Given the amount of measurement error that is likely to be present in wages this is a suitable assumption because the conditional variance will most likely overestimate the amount of risk.[8]

We assume that education is exogenous for wages. We can support this assumption with two pieces of evidence: first, as we shall show the relationship between wages and education is extremely flat within the village. This is true for both adults and children; given the limited occupations that one can undertake in these rural communities this is not surprising. Indeed the returns to education are obtained by migrating to urban centres once education is complete. Second as we shall see there are no selection effects on wages due to participation, implying that despite the fact that unobserved ability is a determinant of school participation (as we show later), it is unlikely to affect child pay rates, which are probably quite homogeneous.

Under these assumptions we could estimate the wage equation separately using OLS and use the predictions in the model. However, we use a Heckman (1979) selection correction approach that allows us to test that selection is not an issue. To construct the inverse Mills ratio, which we include as a regressor in the wage equation, we estimate a reduced form probit for school attendance as a function of the variables we include in the structural model.[9] These include measures of the availability and cost of schools in the locality where the child

---

[8]There is well documented evidence that wages in surveys suffer from substantial measurement error. With time series data and some further assumptions it is possible to identify the variance of shocks to wages, at least in the absence of transitory shocks (see Meghir and Pistaferri, 2004). However, with just one or two observations on individual wages over time one cannot distinguish measurement error from wage shocks.

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

lives, which controls among other for whether this is an experimental locality or not.

Finally, we model q_j in equation 6 as a function of the male agricultural wage in the community and whether the program was implemented in that community. The male agricultural wage acts as the exclusion restriction in the education choice model that identifies the wage. The indicator for the PROGRESA community measures the impact of the program on wages. The economic justification for both these variables is given below, following the presentation of the estimates.

The resulting estimated wage equation for a boy i living in community j has the form:

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

$$(7)$$

where ϖ_{ijt} is the residual. 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 and insignificant, reflecting the limited types of jobs available in these villages, as mentioned above.[10] Thus the key determinant of wages at the individual level is age. However, the coefficients on the community level adult wage $(\ln w_j^{ag})$ and the PROGRESA dummy (P_j)

---

[10] 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.

30

are economically and statistically significant. We now explain how they may arise.

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 in equilibrium the price of a unit of child labour in a locality can be written as

$$\log w_{child} = \frac{\rho + \gamma_{adult}}{\rho + \gamma_{child}} \log w_{adult} - \left[\frac{1}{\rho + \gamma_{child}} \log \left(\frac{L_{child}}{L_{adult}} \right) + \kappa \right] \quad (8)$$

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.[11] 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 providing the necessary exclusion restriction for identifying the wage effect in the education choice model. The term in square brackets in 8 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. In other words to identify the effects of wages on schooling/labour supply we need to

---

[11] 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.

assume that wages vary because of differences in the demand for labour, as reflected in $\ln w_j^{ag}$ and not because of differences in preferences that we do not control for.[12]

We now turn to the effect of the program. This is captured by the "treatment" dummy P_j in equation (7), which decreases L_{child}. This pushes up the wage as a new local equilibrium is established. Our estimates imply that 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.

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 $h_i^0 \zeta + \xi \mu_i$ 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,

---

[12]This assumption is central to identifying wage effects in the cross section and is implicit in the Todd and Wolpin (2006) paper as well.

and is thus more general that the standard specification; we use this as an approximation to the sequential choices made before the program.[13] 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

$$P(ed_{it} = e, Attend_{it} = 1 | z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it}, \mu_i) =$$

$$P(Attend_{it} = 1 | z_{it}, x_{it}^p, x_{it}^s, wage_{it}, ed_{it}, \mu_i) \times P(ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \mu_i) \tag{9}$$

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 Estimation

5.1 Identifying the effect of the grant

Although we estimate our model by Maximum Likelihood, which we discuss below, it is worthwhile discussing the exogenous variability in our data that

---

[13] 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.

drives our results. To estimate the effect of the size of the grant on schooling behaviour, an ideal experiment would have randomised the potential amounts offered across villages or even within villages. As it happens they did not. A village either is in the program or not. Within each PROGRESA village those classified as poor (about 70% of the population) are eligible for participation in the program. To use the variation between eligibles, while allowing for the effects of wealth on schooling, we include a "non-poor" dummy.

The comparison between treatment and control villages and between eligible and ineligible households within these villages can only identify the effect of the existence of the grant. However, the amount of the grant varies by the grade of the child. The fact that children of different ages attend the same grade offers a source of variation of the amount that can be used to identify the effect of the size of the grant. Given the demographic variables included in our model and given our treatment for initial conditions this variation can be taken as exogenous. Moreover, the way that the grant amount changes with grade varies in a nonlinear way, which also helps identify the effect.

Thus the effect of the grant is identified by comparing across treatment and control villages; by comparing across eligible and ineligible households (having controlled for being "non-poor"); and by comparing across different ages within and between grades. This is our basic model. We also estimate a version of the model where we allow for different bahaviour by the ineligible individuals in the control and treatment villages. We do this by including in our model the "non-poor" dummy interacted with being in a treatment village. This leaves the other sources of variation in the grant as identifying information. The motivation for the introduction of this additional dummy is twofold: first, as we documented

above, there were some pre-program differences between the ineligibles in the PROGRESA and the control villages. Second, there may be spillover effects of the program affecting the behaviour of ineligible individuals, which would mean that they behave differently from those in the control villages.[14]

Alternatively, the effect of the grant could be estimated imposing more structure on the data and ignoring the variation induced by the presence of the experiment. One could assume, for instance, separability between child labour and consumption and infer the effect of the grant from the observed response of schooling to child wages. We come back to this issue below.

5.2 The Likelihood Function

We estimate the model by maximum likelihood.[15] 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 having already achieved e years of ed-

---

[14] see Angelucci and DeGiorgi (2009) fo evidence on spillover effects in consumption. If the differences were due to GE effects through wages we already account for these.

[15] 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.

ucation ($ed_{it} = e$) can then be written as

$$P_i = P(Attend_{it} = 1, ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, wage_{it}) =$$

$$\sum_{m=1}^{M} p_m \{ F\{ u_{it}^s + \beta E \max \left[V_{it+1}^s(ed_{it} + I), V_{it+1}^w(ed_{it} + I) \right] \tag{10}$$

$$- \left[u_{it}^w + \beta E \max \left[V_{it+1}^s(ed_{it}), V_{it+1}^w(ed_{it}) \right] \right] | \mu_i = s_m \}$$

$$\times P(ed_{it} = e | z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \mu_i = s_m) \}$$

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 10

We should stress that the wage we use in the estimation is the value predicted by equation 7 (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.[16]

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

$$\sum_{i=1}^{N} \{ D_i \log P_i + (1 - D_i) \log(1 - P_i) \} \tag{11}$$

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

---

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

6 Estimation results

In this section, we report the results we obtain estimating different versions of the dynamic programming model we discussed above. In particular we will be discussing three different versions of the model. The first constitutes our basic model. In the second, we control for the pre-program difference in enrolment rates among non eligible individuals in treatment and control villages with a dummy in the specification for schooling costs that identifies the group of non eligible boys in treatment villages. Finally, we present the estimates obtained fitting a version of our model where we impose the separability assumption used by Todd and Wolpin (2006) to the boys in control towns only.

In Tables 3 to 5, we present, in Columns A and B, estimates of the two 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, while in the second they are accounted for by a dummy for non-eligible households in treatment localities. This dummy does not have a significant effect in the initial conditions equation (Table 4) but is significant in the structural model of educational participation (Table 5). The two degree of freedom likelihood ratio test has a p-value of 0.8%. However, the parameters hardly change when we move between the two specifications and the substantive implications of the two models are the same.

The third column in the Tables presents estimates of the model obtained from the control sample only. In this case, the experiment is not used to estimate the model and all incentive effects are captured by the wage, which acts as the opportunity cost of education. This model could be estimated on any standard observational data set including the type of information we use. The purpose of

estimating this model is to compare the predictions of a model estimated using the experiment to one that does not. We return to this issue later.

For all specifications the discount factor was estimated to be 0.89. The standard errors we report are conditional on this value. This value was obtained from a (rather coarse) grid search over several values, for our favourite version of the model.[17]

We estimate all the versions of the model on the sample of boys older than 9 and younger than 17. 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 (poor) for programme eligibility (potential if in a control village). 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

---

[17]It turns out that approximately the same value of the discount factor maximizes the Likelihood Function both in Column 1 and Column 2 of our tables.

start reporting, in Table 3, the estimates of the points of support 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 heterogeneity in our sample.

We first notice that the results do not vary much across the first two specifications, while they estimates in Column C are a bit different. The estimates in Table 3 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 above 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 locations of the points of support for the model that assumes separability is a bit different, but we can still identifies the three groups we have just discussed.

The loading factor for the first two models of the unobserved heterogeneity term is negative as expected. 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. Perhaps surprisingly, the loading factor for the third model is positive.

The initial condition ed_{it} is modelled, conditional on the unobserved heterogeneity, 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

	A	B	C
Point of Support 1	-16.064	-15.622	-4.29
	0.991	0.975	2.46
Point of Support 2	-19.915	-19.457	-17.62
	1.236	1.218	3.144
Point of Support 3	-12.003	-11.616	-0.267
	0.768	0.758	2.45
probability of 1	0.519	0.521	0.49
	0.024	0.024	0.032
probability of 2	0.405	0.403	0.27
	0.025	0.025	0.017
probability of 3	0.075	0.076	0.023
	-	-	-
load factor for initial condition	-0.119	-0.124	0.068
	0.023	0.023	0.013

Notes: Column A: Eligible dummy only; B: Eligible dummy and non- Eligible in treatment village dummy. C: Model estimated on control sample only. Asymptotic standard errors in italics

Table 3: The distribution of unobserved heterogeneity

than a certain grade. To save space, we do not report the estimates of the cut-off points. 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.

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, which enter the initial conditions equation but not the equation for current schooling utility, 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 indi-

cates that 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, which do not vary much across the three columns, 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 5. 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 given the positive coefficient, an increase in wages decreases school attendance) and 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.[18]

---

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

	A	B	C
poor	-0.281	-0.246	-0.280
	0.027	0.040	0.051
Ineligible individual in a PROGRESA village	-	0.060	-
	-	0.040	-
Father's education			
Primary	0.188	0.189	0.218
	0.024	0.024	0.04262
Secondary	0.298	0.298	0.281
	0.027	0.028	0.05302
Preparatoria	0.596	0.596	0.499
	0.052	0.052	0.09107
Mother's education			
Primary	0.168	0.169	0.231
	0.024	0.024	0.04446
Secondary	0.328	0.329	0.398
	0.028	0.028	0.05139
Preparatoria	0.296	0.286	0.334
	0.058	0.056	0.09740
indigenous	0.002	0.001	0.133
	0.024	0.024	0.04611
Availability of Primary 1997	0.365	0.367	0.691
	0.070	0.070	0.19003
Availability of Secondary 1997	0.744	0.741	-0.568
	0.173	0.173	0.34909
Km to closest secondary school 97	0.001	0.001	-0.0002
	0.003	0.003	0.00007
Availability of Primary 1998	-0.134	-0.144	-0.449
	0.113	0.117	0.23524
Availability of Secondary 1998	-0.789	-0.783	0.516
	0.172	0.172	0.34776
Km to closest secondary school 98	-0.007	-0.006	0.00015
	0.003	0.003	0.00007
Cost of attending secondary	0.005	0.004	-0.00019
	0.022	0.022	0.00037

Notes: as in Table 3. State dummies included
Availability means school in the village.

Table 4: Equation for Initial conditions

	A	B	C
wage	0.103	0.110	0.357
	0.033	0.033	0.100
PROGRESA Grant	4.154	3.869	-
	1.386	1.036	-
parameter in terminal function $\ln(\alpha_1)$	5.798	5.797	6.59
	0.099	0.099	0.175
parameter in terminal function $\ln(\alpha_2)$	-1.104	-1.103	-1.62
	0.020	0.020	0.089
Poor	0.159	-0.128	0.431
	0.118	0.164	0.274
Ineligible individual in a PROGRESA village		-0.492	
		0.198	
Father's Education - Default is less than primary			
Primary	-0.131	-0.133	-0.486
	0.093	0.092	0.217
Secondary	-0.271	-0.267	-0.959
	0.113	0.112	0.261
Preparatoria	-0.751	-0.735	-2.176
	0.253	0.250	0.558
Mother's Education - Default is less than primary			
Primary	-0.136	-0.131	-0.870
	0.096	0.096	0.233
Secondary	-0.137	-0.129	-1.119
	0.113	0.113	0.254
Preparatoria	-0.920	-0.893	-2.158
	0.284	0.283	0.645
indigenous	-0.605	-0.591	-1.018
	0.107	0.107	0.241
Availability of Primary 1998	2.734	2.744	3.092
	0.220	0.219	0.499
Availability of Secondary 1998	-0.160	-0.189	0.789
	0.152	0.152	0.425
Km to closest secondary school 98	0.001	0.001	0.00078
	0.004	0.004	0.00014
Cost of attending secondary	0.005	0.005	0.013
	0.001	0.001	0.0033
Age	2.772	2.758	2.903
	0.180	0.180	0.354
Prior Years of education	-2.995	-2.996	-3.621
	0.214	0.214	0.621
Discount rate	0.89	0.89	0.975
log-Likelihood	-26,676.676	-26,672.064	-8862.34

β = State dummies included
Notes as in Table 3.

Table 5: Parameter estimates for the Education choice model

From a policy perspective, the key parameters of the model are the wage coefficient and the coefficient on 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 wage effect is higher when we estimate the model on the control villages alone. The difference between the wage coefficient in column A and C is significant and has a t-value of 3.25 based on a Durbin-Wu-Hausman type test.

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.[19] As mentioned above, the coefficient of the grant is expressed as a fraction of the wage coefficient. The values of about 4 for this coefficient reported in Table 5 indicates that the effect of the grant on school attendance is considerably larger than the effect of the wage.[20] This contrasts with 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). Such an assumption is easily rejected by a likelihood ratio test.[21] As discussed earlier this test rejects the hypothesis that education and consumption can be taken to be separable

---

[19] 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.

[20] 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.

[21] 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).

and implies the importance of having experimental information to assess the impact of policy in this case.

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) 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 (the likelihood ratio statistic is 300). 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.[22]

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.

7 Simulations

The model is complex and non linear. To check how the various specifications of the model are able to predict the impact of the grant and to quantify the effect of the main variables of interest from a policy point of view we present the

---

[22]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 given in their logs in column 3 of Table 5.

results of simulations of various versions of the model. We start. by comparing the impacts predicted by the different versions of the model with the impacts of PROGRESA as estimated by comparing enrolment rates in treatment and control localities, as we have done in Section 3. We then study changes to the program and alternative policies.

7.1 Predicting the impact of PROGRESA

To check how our basic model predicts the impact of the program, we simulate the behaviour of the children in our sample under two different scenarios. The first correspond to the actual data and includes, in the treatment communities, the grant. It should be stressed that, in these communities, the grant is assumed to be permanent. The second is a scenario in which, counterfactually, we set the grant equal to zero. We then compute the probability of school enrolment of eligible boys in treatment communities under the two scenarios, average these probabilities for different age groups and interpret the different in probabilities as the impact of the program.

When removing (counterfactually) the grant from treatment areas, we compute two predictions. In one case we keep children wages constant across comparisons, assuming that is that the program did not affect wages. This is what one would calculate based on a partial equilibrium model estimated on the controls only. We label this "effect of Grant - No GE". In a second calculation, when we remove the grant we also adjust wages downwards based on our estimates of the impact of the grant on wages. We label this "effect of Grant - GE". To these results, plotted in Figure 1. ,we juxtapose the experimental impacts discussed in Section 3. , which is the dotted line.

In column C of Tables 3 to 5, we presented estimates of a model estimated

47

on the sample of control villages only. We have already demonstrated that the effect of the grant is not equivalent to a corresponding reduction of the wage, which is the opportunity cost of education. We interpret this as a rejection of separability between consumption and education (see Browning and Meghir, 1991). This in itself implies that a model estimated without the experimental variation cannot accurately predict the effects of the experiment on educational participation. In addition to show how our model predicts the impact of the program, we also want to show the extent to which the rejection of the simpler model that assumes separability between consumption and child labour, is a quantitatively important issue.

The left hand side panel in Figure 1 relates to our basic model estimated both on the treatment and control samples. As we plot the difference in probabilities only for eligible children, and as the estimates of most coefficients do not change much between the two versions of the model (with and without the non-eligible treatment village dummies), not surprisingly the age profile of the impacts look very similar. Therefore, in the Figure, we only report one of them.

The right hand side panel in Figure 1. reports the impacts predicted by the model estimated only on control villages with the separability assumption imposed. In this case, the impact of the grant is derived from the coefficient on the wage.

Not surprisingly, the average effect of the treatment (estimated experimentally at 0.05) is predicted quite accurately by the model at 0.047. Moreover, the model predicts reasonably well the inverted U-shape of the impact. Estimating the model based on the controls only we obtain an average effect of 0.031. However, what is more interesting is to consider the age pattern of the effect.

48

Consider first the comparisons that keep wages constant. The age pattern is very similar to the experimental one, except that the model misses the 14% peak effect at age 14 and compensates with higher effects around these ages. The failure of the model to replicate the peak is unsurprising given the simple and parsimonious functional form we have chosen. On the other hand, when we consider the predictions from the model estimated based on the sample of control villages only, we find that it consistently underestimates the effects of the program by large amounts. This is because the wage effect is much smaller than that of the grant, despite the fact that when we estimate the model on the controls alone we obtain a wage coefficient which is more than twice what we obtain in the sample including the treatment villages.

The experimental evidence, as shown in the wage regression we discussed above, showed that the program pushed up children wages by 6%. Thus, comparing treatment and control villages does not measure the partial effect of the subsidy on school participation but includes also the effect that the program has through its impact on the labour market in the treatment villages. Thus the appropriate comparison between experimental effects and model predictions is one that allows for the effect on wages, which is given by the dashed lines in Figure 1. The impact of correcting the effect for wages is minimal in the model estimated on the whole sample. As a result the model still compares well to the experimental effects. However, when we adjust wages in the model estimated on controls alone the effect decreases substantially. Noting that this is the most appropriate comparison with the experimental effects, which include the impact on wages, this result shows that the model estimated on the controls alone performs even worse in predicting the experimental results.
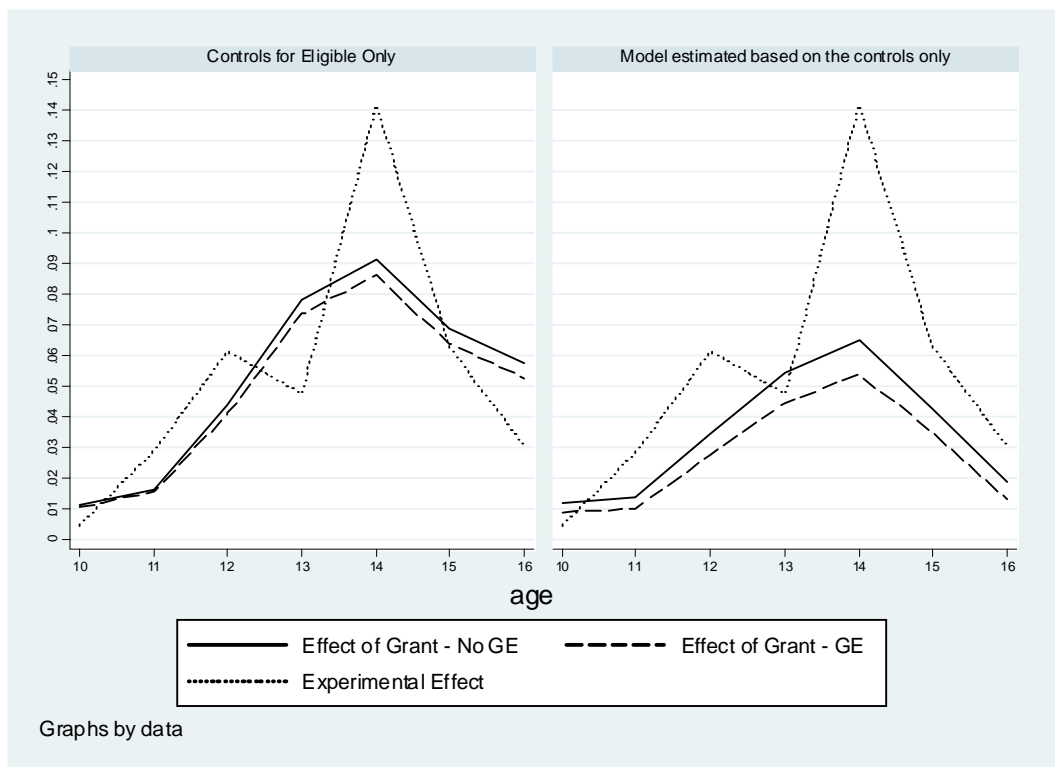
Figure 1: Comparing the model estimated on the whole sample and only on the control sample

The model estimated on control villages only is similar, if a bit simpler, to that estimated by Todd and Wolpin (2006). The version we estimated is simpler in that it does not consider all the details they included in their model. We do not model, for instance, fertility choices and the like. It should be stressed, however, that these simplifications do not explain the worse predictive performance of the model in column C. Indeed, the model we proposed in Column C does not perform worse, for the age interval 12-15 than the model Todd and Wolpin (2006) presented for boys.[23]

We conclude by stressing that our basic model (or its modified version) seem to be able to predict the impact of the program better than the Todd and Wolpin (2006) model. The interpretation of this difference lies in the specification of the model and in particular in the role that non-separability of preferences can play. The fact that Todd and Wolpin (2006) impose separable preferences might explain why their model fails to predict well the impact of the program on boys, which constitutes our sample. It also emphasises the potential use of experimental evidence in estimating structural models.

7.2 Anatomy of PROGRESA and some alternatives.

One of the main advantages of having a structural model, is the possibility of performing policy experiments. We now use the model to simulate school participation under different scenarios. First, we compare the effect of the current program with that of a similar program that differs in the way in which the grant varies with the grade attended.

The aim of this exercise is to compare the impact of the current grant, as

---

[23] Todd and Wolpin (2006) estimate the model for boys and girls. For boys they present predicted impacts for boys 12 to 15.

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 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 by targeting the grant to the older children we can almost double the impact relative to the predicted effect from the model shown in the left panel of Figure 1. This occurs 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 because almost all children go to school below grade 6, making it an unconditional transfer for that age group. The overall resources targeted to families with children do not change with this reform, but the incentive structure does.

This change to the grant structure seems to suggest a modification to the program that would much improve its ability to increase enrolment rates. This is particularly important because the modified program costs, in the steady state, the same amount as the current one. From the point of view of the

52

households, notice that they receive the same amount of resources over time: what changes is when they receive them. If households can borrow against the future grant, then the only effect of this reform is to improve incentives for school participation at later ages. If on the other hand families are liquidity constrained the tradeoff may be more serious, particularly if the grant at a younger age affects nutrition or other child inputs.[24] Attanasio and Rubio-Codina (2008) show that the impact of the PROGRESA grant on a variety of nutritional outcomes for very young children does not depend on whether they have primary school age siblings. This might be an indication that a change in the grant structure as the one describe might not have large negative effects. More recently, the Mexican government is piloting two versions of the program in which primary school grants are eliminated and secondary school grants are increased.

We next consider a number of alternative experiments. In particular we consider the effect of decreasing the wage by an amount equivalent to the grant; the effect of distance to school; and the impact of the grant on children with low cost of education. All three experiments are summarized in figure **??**. In all cases we use the model A in the tables. No grant is our baseline.

First we decrease the wage by an amount equivalent to the grant.[25] We see that the effect of the wage is estimated to be much lower than the grant; for example at age 15 the incentive effect is less than half the one in Figure **??**. This evidence re-emphasises the point already made, that the experimental data provides information on behaviour that may not be available through observed

---

[24] Attanasio and Kaufman (2009) show that for the Oportunidades/PROGRESA population, liquidity constraints can be important.

[25] 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.
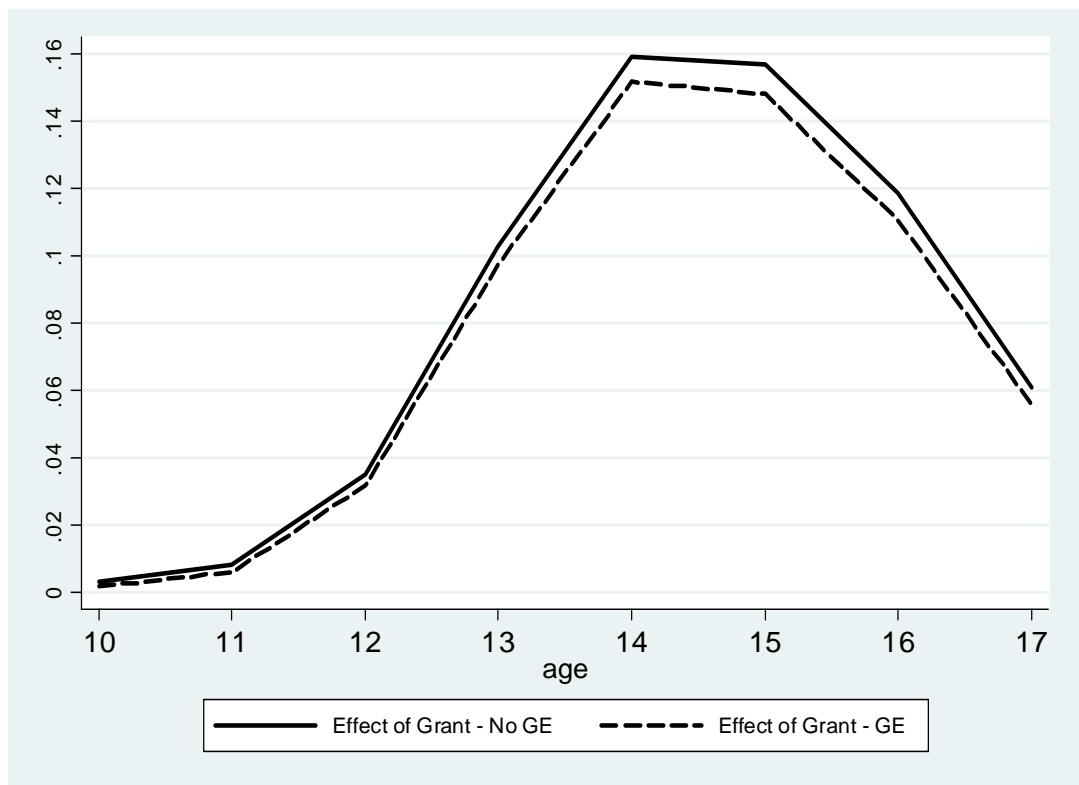
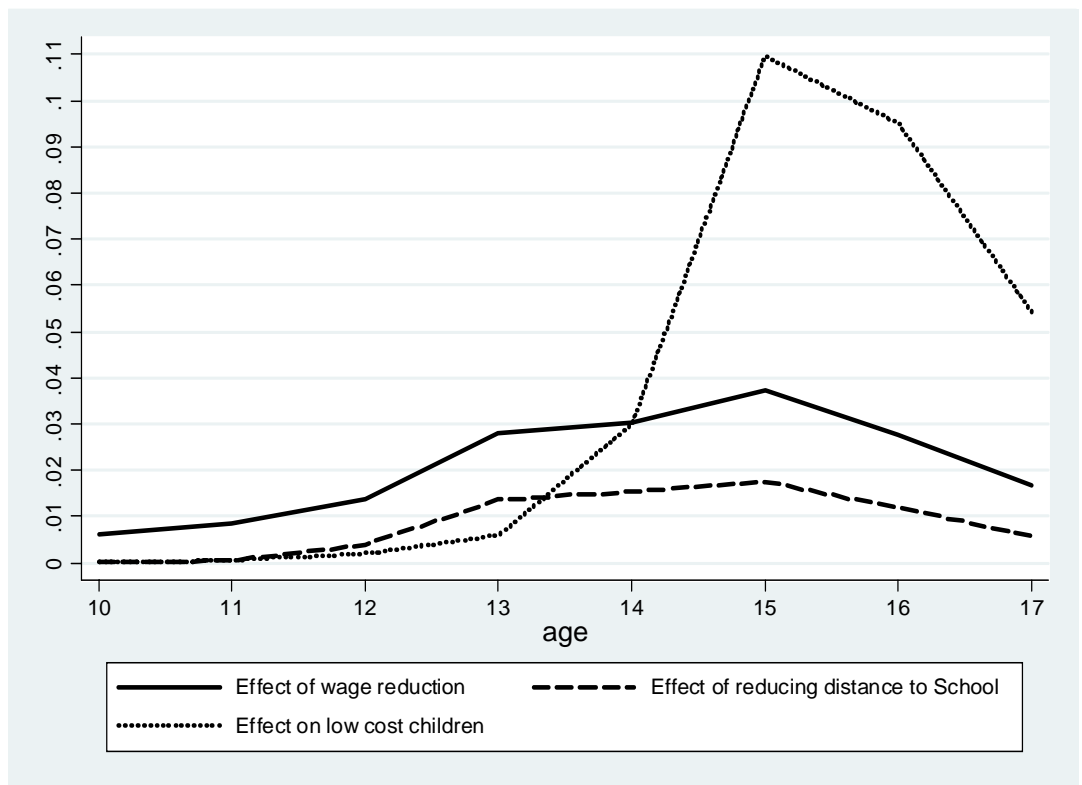Figure 2: The impact of an alternative grant structure.

Figure 3:

data.

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 could 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). According to our parameter estimates 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 GE 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

Angelucci, M and G. DeGiorgi (2009) "Indirect Effects of an Aid Program: The Case of Progresa and Consumption"

American Economic Review, 2009, 99(1), 486-508.

Attanasio, O.P. and K.M. Kaufmann (2009) "Educational Choices, Subjective Expectations, and Credit Constraints" NBER Working Papers 15087, National Bureau of Economic Research.

Attanasio, O. and M. Rubio-Codina (2008): "Revaluatig Conditional Cash Transfers: Is the Oportunidades Primary School Stipend Necessary?", IFS Mimeo

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

Browning, Martin & Meghir, Costas (1991) "The Effects of Male and Female Labor Supply on Commodity Demands," Econometrica, vol. 59(4), pages 925-51, July.

Burtless, Gary and Jerry A. Hausman (1978) "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment" The Journal of Political Economy, Vol. 86, No. 6 (Dec., 1978), pp. 1103-1130

Published by: The University of Chicago Press

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.

Moffitt Robert A. (1979)"The Labor Supply Response in the Gary Experiment" The Journal of Human Resources, Vol. 14, No. 4 (Autumn, 1979), pp. 477-487

Meghir, C. and L. Pistaferri (2004) "Income Variance Dynamics and Heterogeneity" Econometrica, January 2004 pp 1-32

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/Sko

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.

| | The impact of PROGRESA | | |
| | on boys school enrolment | | |
Age Group	Impact on Poor 97	Impact on Poor 97-98	Impact on non-elig.
10	0.0291	-0.0007	0.0723
	(0.027)	(0.026)	(0.060)
11	0.0240	0.0176	-0.0111
	(0.016)	(0.014)	(0.019)
12	0.0478	0.0420	0.0194
	(0.021)	(0.020)	(0.043)
13	0.0391	0.0396	0.0411
	(0.028)	(0.025)	(0.049)
14	0.0838	0.0731	-0.0460
	(0.032)	(0.027)	(0.051)
15	0.0963	0.0816	0.0617
	(0.035)	(0.033)	(0.062)
16	0.0350	0.0472	0.0517
	(0.036)	(0.030)	(0.059)
12-15	0.056	0.043	0.015
	(0.019)	(0.012)	(0.023)
10-16	0.048	0.053	0.006
	(0.013)	(0.012)	(0.032)
Standard errors in parentheses are clustered at the locality level.			

Table 6: Difference in difference estimates - August 1997 - October 1998

10 Appendix: difference in difference estimates of PROGRESA impacts.

In Table 6 we report the estimated impacts of PROGRESA on boy's enrollment for each age from 10 to 16, as well as the average impact for boys aged 10-16 and for the age group 12-15. The sample we use in the estimation of the structural model and for Table 2 is slightly different because of attrition. To compute the estimates in Table 6 we use a slightly different sample because of the necessity to use baseline data.