

EDUCATION CHOICES IN MEXICO: USING A STRUCTURAL MODEL AND A RANDOMIZED EXPERIMENT TO EVALUATE PROGRESA

Orazio Attanasio
Costas Meghir
Ana Santiago



EDePo

**Centre for the Evaluation
of Development Policies**

THE INSTITUTE FOR FISCAL STUDIES
EWP04/04

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

Orazio Attanasio Costas Meghir Ana Santiago

November 2001*

Abstract

In this paper we evaluate the effect of a large welfare program in rural Mexico. For such a purpose we use an evaluation sample that includes a number of villages where the program was not implemented 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. We also find that an approximately 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.

*Preliminary and Incomplete

1 Introduction

The scope of this paper is to evaluate the effect of one component of a large welfare program in rural Mexico that started in 1997. Progresa is a program whose main aim is to improve the process of human capital accumulation in the poorest communities in rural Mexico. The program is made of several components, targeting nutrition, health and education. However, by far the most important one is the education one. Mothers in the poorest households in a set of targeted villages are given grants to keep their kids in school. The grants start in third grade and increase until the ninth and they are conditional on school enrolment and attendance. In this paper we lay down a framework for the evaluation of this particular component of the program in raising enrolment at various ages and grades. Moreover, we also want to evaluate whether the design of the program, for instance the amount by which the grant increases with the grade, has achieved stated objectives.

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 officials in charge of the program 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 were randomized out of the program with the purpose of providing a control group that would enhance the evaluation. However, rather than being excluded from the program all together, in the control villages the program was postponed for about two years, during which period, four waves of the panel were collected. Within each community in the evaluation sample, all households, both beneficiaries and non-beneficiaries, were covered by the survey. In the control villages, it is possible to identify the would-be beneficiaries were the program implemented.

If one believes that the randomization is well designed and one is willing to ignore the fact that the control villages were only temporarily excluded from the program (that is that announcement effects were irrelevant), the effect of the program on enrollment is easily computed as the difference in mean enrollment between treatment and control

villages. If one is worried about pre-program differences in enrollment due to possible problems with the randomization scheme, as a pre-programme survey is available, one could use difference-in-difference type of techniques to measure such an effect. However, if one is interested in slightly more complicated issues, such as how the impact of the program would change if one were to change the amount by which the grant increases with grade, simple differences are not enough. Moreover, if announcement effects in control villages are important, as is likely to be the case for any investment activity, the comparison between treatment and control samples would yield misleading results. In what follows, we argue that without the guidance of an economic model, the data originating from a randomised experiment cannot be utilised to the full. On the other hand without the randomised experiment, neither the simple evaluation nor the estimation of the structural model would be as credible. We show that a combination of the two approaches can make the best use of the data for understanding the impact of the policy.

The contributions of the paper, therefore, are two. On the methodological side, we show how one cannot avoid using a structural model even in the presence of a randomized experiment to fully evaluate and fine tune a specific policy. We also show how the randomized experiment induces extremely useful exogenous variation that helps enormously in the identification of a richer and more flexible structural model. On the practical side, we study the effectiveness of a program based on specific monetary incentives in fostering the accumulation of human capital in rural communities in Mexico. This is important as lags in the accumulation of human capital have been identified by several commentators as one of the main reasons for the relative modest growth performance of Latin American economies in comparison, for instance, with some of the South East Asian countries. For this reason, the programme we study and similar ones have received considerable attention in Latin America.

The rest of the paper is organized as follows. In Section 2, we present the main features of the programme and of the evaluation survey we use. In section 3, we present some simple results on the effectiveness of the programme based on the comparison of treatment

and control villages and discuss the limitations of this evidence. In section 4, we present a structural dynamic model of education choices that we estimate by Maximum Likelihood. Section 5 presents the results we obtain from the estimation and uses the model to perform a number of policy simulations that could help to fine-tune the programme. Finally, Section 6 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 programme 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 programme's benefits. When the programme was started, the administration decided to collect a large longitudinal survey with the scope of evaluating the effectiveness of the progra, In this section, we describe the main features of the programme 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. The health component consists of a number of initiatives aimed at improving information about vaccination, nutrition, contraception and hygiene and of a program of visits for children and women to health centres. Participation into the health component is a pre-condition for participating into the nutrition component that, in addition to a basic monetary subsidy received by all beneficiary households, gives some in kind transfers to households with very young infants and pregnant women. 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 addition to the (bi) monthly payments, beneficiaries with children in school age receive a small annual grant for school supplies. Finally, all the transfers are received by the mother in the household. Before giving additional details on the education component of

the program, we discuss how the program targets communities and households.

The Program first targeted the poorest communities in rural Mexico. 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 basic 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)

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 Progreso 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 households are called 'densificados'.

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\$777m 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. The program has received a considerable amount of attention and publicity and similar programs are currently being implemented in Honduras, Nicaragua and Argentina. (See IFPRI (2000) for additional details on the program and its evaluation).

The program represents a substantial help for the beneficiaries. The nutritional com-

ponent was 100 pesos per month (or 10 US dollars) in the second semester of 1998, which corresponds to 8% of the beneficiaries' income in the evaluation sample.

We report the details of the educational grant in Table 1. All the figures are in current pesos, and can be converted in US dollars at approximately an exchange rate of 10 pesos per dollar. As mentioned above, the 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 loses eligibility.

Table 1: 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

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 were 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 were interviewed, for a total of roughly 25,000 households. Using the information of the 1997 survey and that in the March 1998 survey, each household can be classified as poor or non-poor, that is, each household can be identified as being entitled or not to the program.

One of the most interesting aspects of the evaluation sample is the fact that it contains a randomisation 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 issues. 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.¹

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

As mentioned above, if one ignores the issue of announcement effects in control villages and believes that the randomization properly balances treatment and control villages, one can measure the impact of the program on a given outcome (such as enrolment of children of a given age or completed grade) by comparing the average enrolment (at a given age or grade) in treatment and control villages. Such a comparison, given the maintained assumptions is simple and does not require any behavioural assumption. In this section, we discuss some of the evidence that can be (and has been) obtained using this method. Some of the results we present are taken from Santiago (2001) and are similar to those presented by Schultz (2001).

The appropriatedness of the randomization can be tested checking for presence of significant difference in measured variables in the baseline surveys in treatment and control villages. An extensive study of this type has been performed by Behrman and Todd (2000). They find that, for the large majority of a very wide set of variables, there are no statistical differences between treatment and control villages.

In Table 2, we reproduce some of the exercises performed by Behrman and Todd (2000) and report some summary statistics for control and treatment villages. We focus on our sample of children of ages between 6 and 16 and report statistics from the October 1997-March 98 and the November 98 surveys, that is before and after the start of the programme. As it is evident from the Table, by and large there are no apparent differences in pre-programme variables. The only worrying feature of their study is that one of the few variables for which they do find a statistically significant difference between treatment and control villages, is the pre-program level of enrollment. Especially for some grades,

¹To construct the children wage variable we use in our empirical work, we combine the information on men agricultural wage from the locality questionnaire, with information on children wages (by age and completed grade) for children working in each village. Details of the construction of this and other variables are given in the data Appendix.

enrolment rates are slightly larger in treatment than in control villages. While obviously one would expect differences 'statistically significant' at the 5% level in about 5% of the variables, it is a bit worrying that one of these turns out to be one that is crucial for the evaluation of the program. We come back to this issue below. At this stage, however, we proceed to compare enrolment in treatment and control villages to measure the impact of the program.

Our first set of results is in Table 3, where we compare both eligible and non-eligible households in treatment and control villages. The average treatment effect is of 0.036 points on the eligible households and 0.041 points on the non-eligible ones; both effects are statistically different from zero. Surprisingly, the effect seems larger for the non-eligible families than for the eligible ones! Table 4 reports similar computations performed by completed grade. According to this table, among the eligible households, the strongest effect is for children with completed primary education (almost 9 percentage points). There are also positive effects for children with incompleting primary (3%) and completed secondary education (3.7%). The effects on children from non-eligible families is implausibly high for children with completed secondary (11%) and sizeable for completed primary (6.8%).

Table 5 redoes the same exercise by age: among eligible households the strongest effects are observed for children aged between 12 and 14: the effect ranges between 5% and 13%. Among the non-eligible children, the effect is never statistically significant and quite small for children younger than 15 (with the exception of 12 year old, for whom the point estimate is 6% but with a standard error of 4%). For children aged 15 and 16, once again, the estimated effect is implausibly high at 9% and 11%.

Obviously the Progresa grant is not the only determinant of enrolment decisions. While a proper randomization allows us the possibility of ignoring other determinants and still get an unbiased estimate of the effect of the program, we may gain efficiency if we control on other important determinants of the enrolment decision. In Table 6, we estimate a simple Logit model for school enrollment on the October 1998 wave. In the

next sections, we will show that this model can actually be obtained imposing some strong restrictions on the dynamic model we discuss below: to make these results comparable with the results we present below on our structural model, unlike in the previous tables, we focus on boys only.

In addition to the treatment and control dummies interacted with eligibility status, our logit includes a set of controls, such as state dummies, parental education and ethnicity, age of the children . Moreover, to capture the economic opportunity cost of attending school, we add a child wage variable, which is obtained multiplying the agricultural male wage in the village of residence for a correction that takes into account the age and education of the child. A detailed description of the procedure to obtain such a variable is contained in the Appendix. Each of the two specifications we consider (one with and one without the last grade completed) is estimated on two samples: the first is constituted of all boys, while the second of boys older than 9. In two of the four specifications, we also include the last grade passed by the child. As we discuss in the next section, this variable is obviously endogenous, being the outcome of past school decisions. Below we show that introducing a model for the initial school level (where the additional variables that allows us to identify the model is the lag availability of school in the village), does not make much difference to the results of the enrolment decision.

The first thing to notice is that we obtain positive program effects, that are of similar magnitude for both eligible and non eligible households. Similar results are obtained if we estimate two separate models for eligibles and non eligibles. These results confirm our findings in Table 3. Second, the results are roughly similar in the two subsamples. Third, the specifications that include the last grade passed, which is strongly significant, present a strong and significant wage effect. Such effect, however, is much reduced when the last grade is omitted from the specification. Fourth, there are large negative age effects. Fifth, we find negative effects of variables that proxy for the cost of attending secondary school. Sixth, all specifications contain a set of variables, such as parental background, whose sign and significance is as expected.

To summarize the results based on post-programme evidence, under the assumption that the randomization balances the two samples and that there are no announcement effects, we find that the program has some positive effect among eligible households, especially for children with completed primary education and of ages between 12 and 14. For younger children the effect is much smaller. Somewhat surprisingly, we also find some effects on non-eligible children. However, Tables 4 and 5 show that these arise mainly for older children and that, for that group, are somewhat implausible.

As we mentioned above, the study of Behrman and Todd (2000) found some difference in pre-programme enrollment rates. In Tables 7 to 9, we focus on these differences in enrolment rates between treatment and control villages. Table 7 computes enrolment rates for all children and for eligible and non-eligible children, while Table 8 and 9 redo the exercise looking at different completed grades and different ages. Notice that the differences are stronger among non-eligible children and that the only significant effects are found among children with relatively higher education attainment (complete primary and higher). In Table 9, the same differences are analysed by age. Table 7 indicates that pre-programme enrollment rates are higher in treatment than in control towns. However, these differences are statistically significant only among non-eligible children. These differences might explain the effect estimated on the non eligibles in Table 3 and in the logit reported in Table 6. The differences in Table 8 and Table 9 are almost never significant: the exception are the enrollment rates of non eligible children aged 14 to 16. Once again, these differences can explain the effects estimated on the non eligibles.

It is difficult to explain these pre-programme differences, given that for most other variables the randomization seems to have worked. Given that the focus is the effect of the program on enrollment rates, we try to exploit the presence of a baseline (pre-programme) survey to correct for these differences. Table 10 is the first attempt in this direction. Here we consider the transitions between the enrolment and non-enrolment status between the pre-program and post program surveys in both control and surveys. Notice that the treatment villages have more transitions from non- participation to participation and less

transitions in the opposite directions than the control villages.

Our next step consists in using difference in difference estimators to measure the effect of the program using the 1997 survey as well as the October 1998 one, that is the pre-programme wave of the evaluation panel and the first post-programme one. Table 11 presents the diff-in-diff estimator for the whole sample and for the subsamples of eligible and non-eligible children. The overall average effect of the programme is estimated at 2.2 points overall, and to 2.5 points for the sample of eligible. For non eligible children, the effect is only 0.0047 and is not statistically different from zero.

Table 12 and 13 redo the computations presented in Table 11 conditioning on age and last grade passed, respectively. Among the eligible children, we find some strong and significant effects among 12 and 14 year old, while nothing particularly significant on the results that condition on grade passed. Among the non eligible children, we do not find any significant result, with the possible exception of the 12 year olds, where the effect is estimated at 8% with a standard error of 4.8%.

As we mentioned above, the results presented in this section are difficult to interpret if one thinks of the possibility of announcement effects. And even if one ignores such effects or thinks that they are unimportant (maybe because liquidity constraints), it is difficult to use these results to extrapolate what effects different programs would induce on enrollment. An interesting question to ask is what would happens if we change the program structure. For these reasons we construct a dynamic model that is capable of addressing these issues. It is to the description of such a model that we now turn.

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. We assume that children have the possibility of going to school up to age 17. All formal schooling end 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

viceversa. 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. 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. Second, we allow for state dependence: The number of years of schooling affects the utility of attending in this period. We discuss this issue at length below.

4.1 The basic framework

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

As we are currently 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

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

where z_{it} relates to a number of taste shifter variables, including parental background

and age. 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} represents a logistic 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 that we discuss below. Finally, the term μ_i represents unobservables which we assume have a constant impact over time. As we discuss below, we will be assuming that μ_i is a discrete random variable whose points of support and probability distribution we estimate.

The utility of not attending school is denoted by

$$u_{it}^w = (\delta + \vartheta_i)w_{it}$$

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. ϑ_i is a zero-mean random variable, representing heterogeneity in the sensitivity of child i 's decision to the wage. When we consider this additional form of heterogeneity, we assume that ϑ_i is a discrete random variable whose point of support and probability distribution we estimate along with those of μ_i .

As mentioned above, after age 17, we assume individuals work and earn wages depending on their level of education. 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 simply as a quadratic function of years of schooling, with the parameters to be estimated alongside the other parameters of the model.²

Since the problem is not separable over time, schooling choices involves comparing

²We 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.

the costs of schooling now to its future and current benefits. The latter are intangible preferences for attending school including the potential childcare benefits that parents may enjoy.

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, 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 effort or 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.³ 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. 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

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

³Since we estimate this probability from the data we could also allow for dependence on other characteristics.

where the expectation is taken over the possible outcomes of the random shock ε_{it} . The value of working is similarly written as

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

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. Finally 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.

4.2 Estimation

In terms of estimation, the problem in the absence of unobserved heterogeneity ($\mu_i \equiv \mu$, $\forall i$) other than through the iid shock ε_{it} , is relatively simple. The likelihood function is based on the probability of attending school that takes the form:

$$\begin{aligned} P(Attend_{it} = 1|z_{it}, x_{it}^p, x_{it}^s, ed_{it}, wage) = & F\{u_{it}^s - u_{it}^w \\ & \beta[E \max \{V_{it+1}^s(ed_{it} + I), V_{it+1}^w(ed_{it} + I)\} \\ & - E \max \{V_{it+1}^s(ed_{it}), V_{it+1}^w(ed_{it})\}]\} \end{aligned}$$

where the expectation is taken over both ε and I where relevant.

The difference between the (current) values of going to school and working will reflect both the pecuniary tradeoffs (the effect of the wage and the cost of going to school) and other relevant factors, such as the disutility of work and (possibly) the utility of going to school. Notice that the most general version of our model allows these effects to be heterogeneous across individuals through the terms μ_i and ϑ_i . The difference in square brackets reflects the difference between the future value function implied by the current choice.

Assuming the unobserved preference shock ε_{it} is logistic, when the discount factor (β) is zero our model collapses to simple logit model. With a positive discount factor, instead, the model needs to be solved at each iteration to compute the future value functions

V_{it+1}^s and V_{it+1}^w . In our case these computations are relatively simple since the expected value of the value functions can be computed analytically, because of the distributional assumption we make. Given assumptions on the terminal value function for each final grade, the expressions in equation (1) can be computed by backward recursion..

As mentioned above, in the presence of unobserved heterogeneity, we assume that the constant μ_i (and possibly ϑ_i) is a discrete random variable, distributed independently of all characteristics $z_{it}, x_{it}^p, x_{it}^s$, and the $wage_{it}$.⁴ However, given the structure of our model and the fact that we use a single cross section, we have an important initial conditions problem because we do not observe the entire history of schooling for the children in the sample. That is, we cannot assume that the random variable μ_i (and ϑ_i) 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 assume that conditional on unobserved heterogeneity κ_i the level of schooling achieved up to now follows a Poisson distribution with mean $\exp(h'_i\zeta + \kappa_i)$ where h_i includes variables reflecting past schooling costs such as the availability of secondary schools in pre-experimental years. The probability of the stock of schooling and of attending school in this period are conditionally independent (given $z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it}$, and the unobservables $\mu_i, \vartheta_i, \kappa_i$). Hence we can write the probability of $ed_{it} = e$ and of child i attending school as

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

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

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

The endogeneity of the stock of schooling is captured by the potential dependence of ϑ_i, μ_i and κ_i . Thus we assume that we can model this joint distribution as

$$F(\mu_i = m, \vartheta_i = s, \kappa_i = k) = p_{msk}$$

for $m \in M, s \in S$, and $k \in K$ where M, S and K are the set of points of support for μ, ϑ and κ . Hence for an individual with observable characteristics $z_{it}, x_{it}^p, x_{it}^s, h_i, wage$

⁴In practice dependence with the wage rate can be allowed for. However, the wage data is not rich enough to estimate a joint model of school participation and wages.

the observed probability of attending and having reached a level of schooling e is

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

$$\sum_{m \in M} \sum_{s \in S} \sum_{k \in K} p_{msk} \{ P(Attend_{it} = 1 | z_{it}, x_{it}^p, x_{it}^s, wage_{it}, ed_{it}, \vartheta_i = s, \mu_i = m)$$

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

The number of points of support as well as the values that m , s and k can take and the probabilities at these points can be estimated as suggested in Heckman and Singer (1983).

4.3 Extension to Family decisions

The framework above considers each child a single decision unit and ignores the possible interrelationship of schooling decisions across siblings. It may well be that schooling decisions are made to maximise an overall welfare function for the household rather than the lifetime welfare of the individual child. And even if parents behaved to maximize each child's welfare, decisions across siblings are not independent if unobserved cost components or heterogeneity are correlated across siblings. Finally, the costs of schooling may depend in some way on the number of children. None of these issues have been addressed up to this point in our model.

If one is willing to assume that parents maximize the welfare of each sibling, the dependence between siblings as well as the dependence on family size can easily be dealt with. For the first we can allow the heterogeneity to be correlated across siblings. One simple way to do this is to assume that the costs for each sibling depend on a family factor with weights that vary across siblings in a way that is possibly a function of age and order of birth. In particular, we could model the unobserved component for child i belonging to household h as: $\mu_{ih} = \lambda(\text{age}_{i,h}, \text{order}_{i,h}, x_h) \nu_h$ (where x_h is a vector of household specific variables) and assume that ν_h is given by a discrete random variable, as discussed above.

For the second, we can allow quite easily dependence on the number of siblings and on birth order, simply by including these variables as elements in the cost function. However two difficulties arise here. First, we would have to assume that household composition and unobserved heterogeneity are independent. Second we would need a model predicting

the evolution of household composition, when predicting the future costs of schooling.

The whole exercise becomes much more complicated if one wants to allow for the possibility that the household maximizes a function that reflects the welfare of the parents that might be, in turn, a non additive function of all the children expected welfare. In such a situation, the relative ability of children and the redistributive motives that parents might have become important and, in a dynamic framework, difficult to model. Even though this aspect might be extremely important for the evaluation of Progresa, as the program might distort incentives to the allocation of resources across children, we have left it for future research

4.4 Modelling the impact of Progresa

As we have discussed in detail in Section 2, the program we want to evaluate has two main aspects. One is a straight income supplement scheme that is received by all beneficiary households regardless of children enrolment. The other is a grant (for each child who can enroll in grades 3 to 9), which is contingent on the child attending school. This second aspect of the grant, which progresses according to the figures in Table 1, affects the relative price of education.

One obvious way to model the impact of the 'conditional' part of the programme is to treat the grant as a "negative" wage and the income supplement as an addition to the income for eligible households. Such a strategy would be unavoidable in the absence of the evaluation sample, and in particular, of the randomized experiment. However, it is possible that the funds obtained under the program have different effects than income or wages. This might be because of disutility of labour effects are more complex than those captured by our model or some stigma that may be attached to the grants, or because of the way the program is administered and the person receiving it within the household. All these are important issues, which we have not modelled here. In general, however, it seems likely that the marginal effect of an increase in labour income on school participation choices might be different from the marginal effect of an equivalent decrease in the grant. In our model we can allow for these effects without modelling them explicitly letting the

coefficient on the grant to be different from the (negative of) the coefficient on the wage. Such a coefficient is identified by the presence of the randomised experiment. By testing whether the coefficient on the grant is the same as that on the wage we can investigate whether: (a). The impact of the program can be captured based on estimates from the control sample only and (b). the extent to which the schooling contingent grant has a different effect from the wage.

As we mentioned above, one of the issues that the use of a structural model allows us to address is whether there are important anticipation effects that make the comparison between treatment and control villages enrolment rates uninformative about the impact of the program. In particular, we can estimate the model under different assumptions on when it is perceived that it will be implemented in the control villages. In this way, we can still exploit the exogenous variability induced by the experiment while taking into account announcement effects. In what follows we estimate the program under the assumption that the program is implemented in the control villages in one year, two years or never.

5 Estimation results

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

5.1 The basic model

In Table 14, we present estimates of the basic model. The model is estimated on the sample of boys older than 9. The first column in Table 14 reports the result obtained under the hypothesis that the discount factor is zero. As we discussed above, except for the unobserved heterogeneity term that makes the distribution of the residuals a bit more flexible than the standard extreme value assumption, these results are related to

the simple logit presented in Table 6. In Column 2, we present the results obtained with a discount rate of 0.9 under the assumption that the program will never be implemented in the control villages. In columns 3 and 4, we report the results obtained under the assumption that the program is implemented in the control villages in 1 and 2 years respectively. It turns out that 0.9 was, among the ten equally spaced values between 0 and 1 of the discount factor that we tried, the one that yielded the highest value of the likelihood function for all three specifications.⁵

In the basic version of the model, there is no heterogeneity in the slope of the wage or grant variables. As in Table 6, among the cost variables which are deemed to be relevant for the enrolment decision, we consider parental background variables (such as education and ethnicity of the parents and a dummy for 'poor' households), cost variables (such as the distance to schools and cost of schools) and the wage available to a child of a given age and education in the village of residence. The Progresa grant is immediately relevant for beneficiary households living in Progresa villages and (for the specifications in column 3 and 4) relevant in the future for beneficiary households living in control villages. To capture the income effect induced by the non-conditional part of the grant, we introduce a beneficiary-treatment dummy. Finally, to capture the possibility of spill over effects (or possibly pre-programme differences between treatment and control villages) we also consider a dummy for non beneficiary household living in Progresa villages.

In this version of the model, we ignore two of the issues we discussed above. First, we ignore the endogeneity of current year of schooling. Second, we ignore the fact that some children live in the same household and the same village. Moreover, the only form of unobserved heterogeneity is in the cost level, for which we fit a two points of support distribution.

The coefficients on the grant and on the interaction of grant and age we report are expressed as a fraction of the coefficient on the wage: that is if we indicate with α_w the coefficient on the wage and with α_g the coefficient on the grant, the value reported in the

⁵The discount factor turns out to be reasonably well determined in our estimates.

tables is α_g/α_w (and analogously for the coefficient on the interaction of grant and age). The interesting null hypothesis is that the reported coefficient on the grant is one, so that the effect of the wage and of the grant are the same.

The terminal value function is constrained to be non negative and increasing in the value of education. In particular, if we indicate with β_1 and β_2 the coefficients that determine the shape of the value function, the latter is given by $V(i) = \exp(\beta_1)/(1 + \exp(-i * \exp(\beta_2)))$.

Several interesting features emerge from the table. First, as in the results reported in Table 6, age and completed grade turn out to be particularly important, with opposite signs. Second, as in some of the specifications of Table 6, the wage is also important. Third, the grant seems to have a much larger impact on schooling than the wage. Notice that we can identify separately the coefficient on the wage and that on the grant because of the presence of a randomized experiment. Moreover, except for the specification where it is assumed that the program is implemented in the control villages in one year, the effect of the grant increases with the age of the child. Fourth, as in the simple logit, the sign and significance of the controls we introduce (parental background, ethnicity and state dummies, cost variables) are consistent with what one would expect. Fifth, when we assume that the program is never implemented in the control villages (an assumption that is implicit in the diff in diff procedure), we obtain a slightly negative impact of the program at age 10, which then becomes more and more positive. However, when we assume that the program will be implemented in 1 year in the control villages, we find a positive (and significant) effect even at age 10. Overall the effect is much stronger than with the assumption that the program is never implemented in the control villages. The likelihood function is higher in this case than in the case in which the program is never implemented in the control villages. Sixth, when we assume that the program is implemented in 2 years, we obtain results that are in between those obtained with the other two assumptions. The likelihood function declines relative to the case in which the program is implemented in one year. Seventh, the two points of support indicate

the presence of two types that are very different, in that one type has much stronger preferences for going to school. This group constitutes about 20% of the sample. Eighth, the results change quantitatively across specifications. The specification that achieves the highest value of the likelihood function is the one that assumes that the program is perceived to be implemented in 1 year. Ninth, as in Table 3 and 6 that use only post-programme data, we find a positive and significant effect on non beneficiary programs living in Progresas villages. However, once we factor in the grant, the effect for the beneficiary households is much larger. Tenth, we can convincingly reject the constancy of the terminal value function. The function is concave in years of education and the coefficient that determines its slope is strongly significant.

5.2 Robustness of the results, limitations and extensions

In this sub-section, we investigate how robust our results are to various issues. First, we experimented with a large variety of controls which might capture the effects of school costs and include those that turned up to be most important. In particular, we investigated various ways of introducing parental education and other parental background variables. The results we obtain are robust to all these changes. The specification we chose was the most parsimonious and robust of the various variables that we considered.

We also experimented with introducing heterogeneity in the coefficient of the wage and the grant. As we discussed above, this term is introduced as a discrete random variable with an arbitrary correlation with the other heterogeneity terms. We failed to identify any significant heterogeneity in this dimension, at least with two points of support.

As we mentioned above, probably the most serious issue with the simplest version of the model is the potential endogeneity of the last grade passed. While it is an undeniably important variable, one would expect that such term is correlated with unobserved heterogeneity, as it is the outcome of past education choices. To get around this problem we introduced a reduced form equation for this variable, which we modeled as a function of various controls and the distance of the village from primary and secondary schools in

previous years. While the current distance from secondary schools is an obvious determinant of the current school cost and therefore of enrolment decisions, lagged values of the same variables are likely to be important determinant of the last grade completed. To the extent that lagged school distances vary across villages in addition to the current school distances, they provide variation that could identify our two-equation model. Our evidence indicates that this is the case: there seem to have been some differential changes in school availability in the villages over time and this is reflected in our completed grade data.

We therefore proceeded to estimate our two equation model. In particular, we model the unobserved heterogeneity term in the completed grade equation as a discrete value random variable with an arbitrary pattern of correlation with the unobserved heterogeneity in our school choice equations.

Estimation of the two equation system yielded the surprising result that the past grade does not seem to be correlated with unobserved heterogeneity in school choices. As a consequence, the results we obtained with the two-equation system hardly changed at all relative to those in Table 14.⁶

As we discussed above and as we show in the following section, the structural model we estimate has several advantages. However, it is not without problems. In our opinion, there are two important drawbacks in our analysis. The first is, as we mentioned at the beginning, that we do not model family decisions. Households make decisions about several children simultaneously and it is possible that the grant introduces some distortions in the allocation of investment in human capital across siblings. Indeed, as we pointed out in Section 3, there is some evidence of this, in the result that the grant seems to have a negative effect on children with completed secondary (who are not entitled to any grant).

The second is that, as we only considered a post programme wave, we cannot dis-

⁶To check where this result comes from, we estimated the reduced form on its own fitting a flexible functional form. We did not find evidence of significant unobserved heterogeneity even in the single reduced form equation, even when we controlled for relatively few observables.

tinguish between spill-over effects and pre-programme differences between control and treatment villages. The evidence we presented from the diff-in-diff estimates indicate that most of the post-programme difference between non-beneficiary households in treatment and control villages could be explained by pre-programme differences rather than spill-over effects. This issue, however, needs to be investigated further. The most natural way would be to estimate the model we presented using both pre-programme and post-programme waves.

6 Simulations

Using the estimation results, we perform two simulation exercises. First, we use the model to compute what is the effect of the grant under different scenarios, that is under different assumptions about when the program is implemented. Second, we use the estimates to assess what is the effect of a change in the structure of the program.

We perform the first exercise by first simulating the model with the existing grant and then again setting the grant equal to zero. This simulation therefore yields the effect of the grant, which can be computed at different ages. Provided one believes that the program is never implemented in the control villages, the comparison of these results to the diff-in-diff results presented in section 3, , should give one an indication of how well the model is fitting the data. An additional word of caution, however, is granted as the diff-in-diff results neglect all dynamic effects.

In Figure 1, we plot, for each age, the increase in the proportion of enrolled children induced by the programme. This is computed from the difference in the simulation of the model without the grant and the model with the grant. In the Figure we plot the results obtained with the assumption that the programme is never implemented in the control villages and with the assumption that the programme is implemented in one year.

Considering first the effect obtained under the assumption that the programme is never implemented in the control villages, we notice that, as in the diff-in-diff case, the effect varies with age and peaks between the ages of 12 and 14. Having stressed the similarity

in patterns, however, we should stress that our model slightly under-predicts the changes observed in the data. If we move to the evidence obtained under the assumption that the program is implemented in the control villages in 1 year, we find a large increase in the estimated effect of the program, as one would expect. This indicates that neglecting the announcement effects that might affect the control communities, one underestimate the effect of the programme substantially.

The second exercise we perform is aimed at evaluating the effect of a change in the structure of the program. In particular, we focus on a 'balanced budget' reform of the programme. 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). This exercise is performed using the model that assumes that the program is implemented in the control villages in 1 year. In Figure 2, we plot the effect of the old grant, the effect of the reformed grant and the difference between the two. Notice that, because of the reduction of the grant to zero for early grades, for early ages there is a slight reduction of the effect of the program on enrollment. However, this is more than compensated by the substantial increase in the enrolment rates of older children. If the objective of the programme is to increase enrollment rates, it seems that backloading the program is much more effective than the current design.

This result is not entirely surprising, as the enrolment rates of very young children are already relatively high and presumably difficult to increase with the type of monetary incentives offered by the program. It seems that at the end of primary school, when enrolment drops dramatically, there is much more scope for action. What the exercise does is to quantify how to a feasible re-allocation of resources would affect enrolment rates. Notice that this type of exercise is not feasible without a behavioural model that allows us to extrapolate the effect of the change in the program in a consistent way.

7 Conclusions

In this paper we have evaluated the effect of a large welfare program in rural Mexico aimed, among other things, at increasing enrolment rates among poor children. The evaluation survey we use has an important randomization component that induces truly exogenous variation in the enrolment into the program. After presenting some evidence based on simple differences and differences in differences, we show that many questions are left unanswered by this type of techniques and propose the use of a structural model. The exogenous variation induced by the randomization is useful in estimating a relatively flexible version of the program. The use of a coherent dynamic optimization model allows us to address a number of important issues, such as the announcement effects induced in control villages by the fact that the program is known to be implemented there after a lag. Our results indicate that these effects are important and do make a difference.

The estimates we obtain are sensible and indicate that the program is quite effective in increasing the enrolment of children at the end of their primary education. Our simulations also indicate, however, that the performance of the program could be improved by back-loading the program, that is offering more resources to older children and less to relatively younger one.