Effects on School Enrollment and Performance of a Conditional Cash Transfers Program in Mexico

Pierre Dubois*, Alain de Janvry†, Elisabeth Sadoulet†

March 2008‡

Abstract

We study the effects of a conditional cash transfers program on school enrollment and performance in Mexico. We provide a theoretical framework for analyzing the dynamic educational process including the endogeneity and uncertainty of performance at school (passing grades) and the effect of a cash transfer program conditional on school attendance. This framework is developed to study the Mexican social program Progresa (called now Oportunidades) in which a randomized experiment has been implemented and allows us to identify the effect of the program on enrollment and performance at school. Using the rules of the conditional program, we can explain the different incentive effects provided. We also derive the formal identifying assumptions needed to estimate consistently the average treatment effects on enrollment and performance at school. We find empirically that this program had always a positive impact on school continuation whereas for performance it had a positive impact at primary school but a negative one at secondary school (a possible consequence of disincentives due to the program termination after the third year of secondary school).

Key words: education demand, schooling decisions, school performance, dynamic decisions, treatment effects, transfer program, randomized experiment, Mexico.

JEL Classification: C14, C25, D91, H52, H53, I21, I28, J24.

---

*Corresponding author:
Toulouse School of Economics (GREMAQ, INRA, IDEI) and CEPR
Manufacture des Tabacs
21 allée de Brienne
31000 Toulouse France
dubois@toulouse.inra.fr

†University of California, Berkeley.

‡We are grateful to Progresa for providing us the data, and Progresa staff of Hidalgo State for guiding in visiting communities. We thank Jérôme Adda, Orazio Attanasio, Antoine Bommier, Richard Blundell, Luis Braido, Pierre-André Chiappori, Carlos da Costa, Bruno Crépon, Marcelo Fernandes, Denis Fougère, Andrew Foster, Gustavo Gonzaga, Sylvie Lambert, Guy Laroque, Pascal Lavergne, Ethan Ligon, Thierry Magnac, Torsten Persson, Bas Van der Klaauw, Ed. Vytlacil for their comments as well as the seminar participants at CREST Paris, the University of Toulouse, Canadian Economic Association Conference in Montreal, Society for Economic Dynamics Conference in Stockholm, Econometric Society European Meeting and European Economic Association Annual Conference in Lausanne, Brazilian Econometric Society Meeting in Porto Seguro, PUC Rio de Janeiro, Getulio Vargas Foundation in Rio de Janeiro, ECARES, Université Libre de Bruxelles, INRA LEA Paris, CIRPEE Université Laval, Québec, Tinbergen Institute, the Institute of International Economic Studies at Stockholm University and University College London.

1 Introduction

In 1998, the Education, Health, and Nutrition Program, known by its Spanish acronym as Progresa[1] and renamed later Oportunidades, was introduced in rural Mexico. The purpose of the program is to provide resources and incentives to increase the human capital of the children of poor rural households, thus attempting to break the inter-generational inheritance of poverty. The program provides cash transfers as well as in kind health benefits and nutritional supplements to poor households, conditional on the child's school attendance and on regular visits to health centers. On average, these cash transfers represent 22% of the income of beneficiary families. The program has grown rapidly and was covering 2.6 million rural families in extreme poverty in 2000, corresponding to about 40 percent of all rural families in Mexico. Progresa operated in 50,000 localities in 31 states, with a budget of approximately one billion dollars for 2000. In 2008, Oportunidades covers more than 5 million families in both rural and urban areas.

In Mexican rural communities, children tend to begin their labor force participation at early ages in order to contribute to family income. One of the main objectives of Progresa is to reduce this early labor force participation of children and thereby increase their enrollment and attendance at school. The program is made up of three closely linked components, education, health, and nutrition based on the idea that positive interactions between these three components enhance the effectiveness of an integrated program over and above the separate benefits from each of these components. The educational component of Progresa provides monetary grants conditional upon attendance at school and constitutes the main part of benefits.

The purpose of this paper is to evaluate the impact of Progresa on the educational behavior of children. We develop a dynamic model of education demand incorporating incentive effects of the educational system on the behavior of students. The model incorporates the grants system introduced by Progresa and shows that such a program does not only affect enrollment decisions but also behavior at school in terms of incentives to pass to higher grades, a crucial point which is not addressed in most education demand models. The most recent developments of education demand models embody the dynamics and uncertainty associated with wages and returns to schooling as well as liquidity constraints (De Vreyer, Lambert, Magnac, 1999; Magnac and Thesmar 2002a, 2002b; Cameron and Heckman, 1998, 2001; Cameron and Taber, 2004; Rosenzweig and Wolpin 1996; Eckstein and Wolpin, 1999, Keane and Wolpin, 1997, 2001). But, most models assume that schooling decisions allow households to choose with certainty the level of school attainment reached by each child or at least that the decision to continue revised each year does not involve

[1] Programa de Educación, Salud y Alimentación.

any uncertainty in grade progression. Once the decision to enroll at school has been taken, the previous literature assumes that the child will pass the grade and benefit from the expected return of one additional year of schooling. This issue is addressed in Magnac and Thesmar (2002a) where grade completion is stochastic. They show in an application to France that one of the reasons for the rise in educational levels observed in France between 1980 and 1993 is the decreasing selectivity of the education system. Attanasio, Meghir and Santiago (2005) study the effect of Progresa on enrolment using a structural model where schooling costs can be stochastic and class repetition is also allowed. However, they 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. Cameron and Heckman (1998) model the transitions from one grade to the next as random processes depending on a number of characteristics without distinguishing whether non progression comes from school drop out or repetition. To our knowledge, there is no theoretical model where both the endogeneity and uncertainty in successfully passing grades are explicitly modelled. Here, we take explicitly these features into account because both school enrollment and school performance determine educational attainment. Understanding the determinants of school performance is particularly important in our case because the repetition of classes is quite frequent in Mexico.

We use data from the Progresa program to empirically estimate its effects on the discrete choices of school continuation and successfully completing grades. This estimation faces the usual identification problems in estimating discrete choice models with unobserved heterogeneity that generate a dynamic selection bias (Cameron and Heckman, 1998). However, Progresa implemented a randomized experiment which helps solve the evaluation problem, that individuals participating in a program cannot be simultaneously observed in the alternative state of non participation. Under the corresponding realistic assumptions, we study the identification of the average parameters of interest and show why we can identify the average program impact in the first year of the program without dynamic selection bias. With the available panel data, we observe continuation decisions and the students' cognitive achievements through their performance (success or failure of a grade). The theoretical model shows that, the transfer program can have either positive or negative effects on performance. Moreover, the impact of the Progresa program on enrollment depends on transfers compared to the opportunity cost of time spent at school by children. In addition, the program can also increase the learning effort of children going to school just because they want to receive future Progresa transfers that increase with grade. Therefore, empirical evaluation is needed in order to sort out the positive or negative impacts of the program. The results show that students actually internalize incentives in their educational behavior since the program affects not only enrollment decisions but also performance.

In section 2, we characterize the related problems of low enrollment and poor performance in secondary schools in rural Mexico, we describe the Progresa transfers with the incentive effects they create and present the data. In section 3, we then develop a life cycle model of education demand where the program impact is explicitly modeled in order to derive how the program design affects individual education decisions. Section 4 studies identification of the program impact in school continuation and performance probabilities. Identification of different average treatment effects is studied. Estimation results are presented in Section 5. Section 6 presents the results of a semi-structural estimation of the model based on stronger parametric identifying assumptions. Section 7 concludes.

2 Education in Rural Mexico and the Progresa Program

The Progresa program has three general components: health, nutrition, and education. The health component offers basic health care to all members of the family. The nutrition component includes a fixed monetary transfer for improved food consumption, as well as nutritional supplements targeted at all children under the age of two, to malnourished ones under five and to pregnant and breast-feeding women. Families must complete a schedule of visits to health care facilities in order to receive monetary support for improved nutrition. However, education is by far the most important component of the program in terms of cash transfers. It consists in payments to poor families with children attending school in grades 3 to 6 of primary school and 1 to 3 of secondary school. After three years in the program, families may renew their status as beneficiaries, subject to a reevaluation of their socio-economic condition.

2.1 School Attendance and Performance

Although educational levels are improving over time in Mexico, current levels in poor rural communities remain very low. In rural communities, there are still only 36% of 18 years old that have gone beyond primary school.[2] The major breaking point in school attendance occurs at entry in secondary school (Table 1). In primary school, continuation rates reach at least 95% in every grade, with the result that 85% of the children that start primary school complete the cycle. However, only 72.4% of the children that successfully complete primary school enroll in the first year of secondary school. The gender difference is very pronounced at this decisive step, with 75.1% of the boys going on to secondary school and only 69.4% of the girls.[3]

[2] These values include a downward bias in the educational achievement of the population of rural communities if the more educated are more likely to leave the poor communities. A careful examination of exit behavior would be necessary to properly assess the trend in education.

[3] The descriptive statistics in Table 1 are for the whole population ("poor" and "non poor") of children in control villages (i.e. villages where the Progresa program is not implemented).

Table 1 illustrates the key role of school performance in the decision to continue. There is here again a striking discontinuity at entry into secondary school. The continuation rate at the end of primary school is much lower than the one at any other grade level (except after third year of secondary school). The performance rate is the lowest at the first year of secondary school and there are very high dropping rates after a first year of trying secondary school without success. Table 1 also shows that dropping out at the fourth and fifth year of primary school is much more important if the child did not complete his grade. Continuation among those that succeeded is more than 95%.

Table 1: Continuation and Performance Statistics[4]

(% in 1998 for children at school in 1997)	Overall Continuation		Performance (grade success)		Continuation among those that passed		failed	
Grade attended in 1997	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Primary school								
2	98.5	97.5	76.2	77.9	98.5	97.1	98.5	98.8
3	96.9	96.8	77.9	78.1	96.9	97.8	96.8	92.9
4	96.7	96.0	78.5	80.4	97.2	96.9	95.2	92.2
5	94.9	95.9	83.7	85.1	95.6	96.8	91.0	91.0
6	75.1	69.4	84.1	84.7	73.5	65.9	83.9	89.3
Secondary school								
1	82.3	77.5	74.2	67.8	94.3	94.9	47.5	41.1
2	95.5	95.1	84.4	82.1	96.2	96.1	91.8	90.2
3	57.3	59.4	80.7	75.3	36.9	49.0	91.0	91.8

Consequently, there is a clear problem of school continuation in the absence of Progresa, especially at entry into secondary school, and it is intertwined with a problem of performance. Hence, the challenge facing Progresa is to address both issues, i.e., to change the incentives to enroll children and to improve their performance in school. As we will see in the next section, the design of the program has the potential of effectively addressing both issues.

2.2 Incentive Scheme of the Program

The educational component of the program consists in conditional cash transfers to families (given directly to the mother of children) for each eligible child going to school between the third grade of primary school and the third grade of secondary school upon attendance at school. Eligibility for the educational transfers is at the individual level for students from poor households in randomly selected treated localities (the poverty status of the household defines a household level eligibility criterion for the whole Progresa program). The levels of the grants increase as children progress to higher grades, probably in order to match the rising income children would contribute to their families if they were working. Additionally, the grants are slightly higher for girls than for boys at secondary school. All monetary benefits are given directly to the female (mother) in the family.

[4] Statistics for all households (eligible and non eligible according to poverty index) in control villages only.

Specifically, the educational program consists in (Progresa, 2000):

- unconditional annual transfer (almost always in cash) for school materials.

- bimonthly cash transfers depending on gender and grade, conditional on presence at school (at least 85% of school days, i.e., not more than 3 missings a month) from third year of primary school to third year of secondary school. Amounts are reported in Table 2.

- an upper limit for household level cash transfers (details about this rule will be explained and used in section 4.3).

- students lose eligibility if they repeat a grade twice.

Table 2: Monthly Progresa Transfers in Pesos[5]

Educational Grant by Student	1997	1998	1999
Primary School (Boys and Girls)			
1^{st} and 2^{nd} year	0	0	0
3^{rd} year	60	70	80
4^{th} year	70	80	95
5^{th} year	90	100	125
6^{th} year	120	135	165
Boys in Secondary School			
1^{st} year	175	200	240
2^{nd} year	185	210	250
3^{rd} year	195	220	265
Girls in Secondary School			
1^{st} year	185	210	250
2^{nd} year	205	235	280
3^{rd} year	225	255	305
Further Schooling	0	0	0
Cash Transfer for Food	90	100	125
Household level maximum benefit	550	625	750

Given these program rules, several incentive mechanisms potentially affect the behavior of treated households. Actually, all transfers, conditional or unconditional, create an income effect[6]. The conditionality of educational transfers on school attendance creates both static and dynamic incentives to enroll. The static incentive is related to the current transfer payment, which reduces the foregone income in going to school. The option of receiving future transfers if one stays in school creates the dynamic effect. Rising transfers with grade level further enhance the dynamic incentives. Increasing school attendance may raise students' knowledge and reduce grade repetition[7]. Rising

[5] Nominal values corresponding to the second semester of the year (changes occur every semester). Approximately 10 Pesos = 1 US$.

[6] To the extent that preference for and performance in school increase with income, transfers will have a beneficial effect. This is all the more true if the household faces short term liquidity constraints and schooling entails monetary costs, such as transportation to school.

[7] At the individual level, more school attendance is expected to improve learning. However, there may be negative externalities on those students which, in any case, would have attended school regularly if the increased number of children in a classroom lowers the quality of the school. As the data do not indicate the exact attendance in class, we cannot test the presence of these kinds of externalities.

transfers with grade give an incentive to pass grades and thus to perform better at school, in addition to the threat to lose eligibility after two repetitions. At last a possible negative effect of program termination may appear since in the third year of secondary school, students could prefer to repeat their grade rather than losing the transfer.[8]

Denote $\tau(l, g, 1)$ as the transfer received if the child of gender g and completed grade l attends school. $\tau(l, g, 1)$ is increasing in l until the end of the third year of secondary school after which the program stops and program benefits drop to zero. Note that with the cap on total household transfer, direct incentives to attend school are a function of family structure, giving us a possible variation in the value of transfers across children (as will be used in section 4.3).

2.3 A Randomized Experiment

The Progresa program operates in all poor communities (defined by a national marginality index developed from the 1995 census) that have minimal access to primary school and primary care facilities, and all households characterized as poor in these communities are eligible (see Skoufias, Davis, and Behrman, 1999). Poverty status of the household was established at the household level prior to the start of the program on the basis of a household census run in October 1997.

Because of the large scale of Progresa, it was decided to implement the program progressively and to design the first years of implementation in order to facilitate evaluation by experimental methods. A subset of 506 of the 50,000 eligible communities was selected to participate in the evaluation. Each of these communities was randomly assigned either to the treatment group where Progresa was implemented starting in 1998, or to the control group where Progresa would be introduced three years later (Behrman and Todd, 1999). On average, 78% of the population of the selected communities was deemed in poverty and hence eligible for the program. All households (eligible and non-eligible) of both types of communities were then surveyed twice a year during the three years of the evaluation. These experimental communities are located in seven states (Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz). There are 320 treatment localities and 186 control localities in the experiment. Program benefits began in May 1998. The unbalanced sample, including all individuals present at some point in time between October 1997 and November 1999, is of 152,000 individuals from 26,000 households. Because transfers are generous, almost all eligible families chose to participate (97%).

Two factors make the study design especially rigorous. One is the random assignment of

[8] To get an order of comparison, in Rural Mexico, the average daily wage of a 16-18 years old boys with completed junior high school in the sample was 25 pesos in 1997. A full time work of 20 days per month would generate an income of 500 pesos per month, compared to a maximum of 255 pesos from Progresa transfers. Therefore, heterogeneity of individuals and of labor market expectations and opportunities implies that it is likely that some students will prefer to repeat their third year of secondary school.

communities into treatment and controls. The other is the panel dimension of the data collected before the intervention of the program and subsequently every 6 months. We use data from the first two years of evaluation, which include a baseline survey in October 1997 and the follow up surveys in October 1998 and November 1999. We thus have information on enrollment during three consecutive school years 1997-98, 1998-99, and 1999-2000, and on performance in school during academic years 1997-98 and 1998-99.

A first look at statistics contrasting treatment (Progresa) and control communities in Table 3 show that pre-program enrollment and passing rates are higher in the Progresa communities (as already reported in Schultz (2000, 2004) for enrollment and in Behrman Sengupta and Todd (2005)).

Table 3: Performance and Completion Statistics in Treatment and Control Groups

Grade attended in 1997	Overall Continuation		Performance (grade success)		Continuation among those that passed		Continuation among those that failed	
	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
Primary school								
2	98.0	98.6	77.1	77.6	97.8	98.8	98.6	98.1
3	96.8	98.1	78.1	83.1	97.4	98.3	94.9	97.3
4	96.4	97.6	79.4	82.7	97.1	98.1	93.9	95.3
5	95.4	97.3	84.4	85.9	96.3	97.8	91.0	95.0
6	72.4	79.9	84.4	85.8	69.8	78.1	86.5	91.2
Secondary school								
1	80.1	87.3	71.2	75.2	94.6	96.5	44.2	59.7
2	95.3	94.9	83.3	85.3	96.2	96.1	91.0	87.7
3	53.0	56.4	78.2	78.9	42.4	49.0	91.4	83.9

The population of interest is that of poor people who are designated as eligible for Progresa. Those in the treatment group can receive the Progresa benefits and those in the control group cannot. The average household size is around 7 people. A little less than one third of individuals in the sample are indigenous. 15% of household heads have an educational level less than primary school, 30% completed primary school, and 52% completed secondary school.

3 A Dynamic Educational Model with Schooling, Effort, and Performance

In order to study the effect of the program on education, we elaborate a dynamic schooling model able to show that it may affect both enrollment and learning. The modelling of schooling decisions is generally done by assuming that the household decision maker maximizes the net expected income of a child. In this calculus, earnings are increasing with education, but education has a cost charged against this income which includes the opportunity cost of the time spent studying instead of working. As we have seen in the descriptive statistics, failure to pass a grade is a serious problem in Mexico's poor rural communities. To face up to this problem, Progresa was purposefully designed to be conditional not only on grade level but also on school performance. Moreover, the role of class repetition in the decision to drop out of school can be very important and the analysis of schooling decisions can be very misleading if one does not account for this phenomenon. Therefore, we develop a decision model where the return to schooling and its opportunity cost depend on the completed school grade of the individual and where successfully completing grades is uncertain and endogenous (it depends on a learning effort variable). Then, we explicitly include in the model the effect of a conditional cash transfer program which allows us to analyze the implications of the Progresa program on educational decisions.

Assume that the decision maker (here it could be the mother who is the household beneficiary) maximizes the discounted lifetime expected utility of the child. A child who has completed grade l at the beginning of an academic year is assumed to be automatically accepted in grade $l+1$ if he enrolls at school. Then, according to Progresa rules, if the household is eligible, the household beneficiary (generally the mother) is entitled to an educational transfer of $\tau(l, g, p)$. For poor people, p is equal to 1 in randomly selected treatment villages and 0 otherwise (with $\tau(l, g, 0) \equiv 0$). Let s be a variable equal to one if the child is actually going to school and zero otherwise. Let π, the *educational performance* of the child, be a function of his school level l, and an individual learning effort e: $\pi(l, e)$. This effort variable is meant to represent individual actions of the student such as attention in classes, being late at school, and studying at home. As educational learning and skills are not perfectly observable by the teacher, we assume that the student will complete grade $l+1$ if and only if $s = 1$ and $\pi(l, e) \geq \varepsilon$, where ε is a random variable with c.d.f. F and p.d.f. f. π depends on l because the level of performance required to pass varies with grade level. This function will depend on the selectivity of the educational system settled by the government (that is how hard are exams to pass). The function π can also depend on characteristics x of the student (a vector including individual and other characteristics like for example distance to school) but we don't need to explicitly introduce them in the theoretical model as long as they are exogenous.

Grade progression from year t to year $t+1$ is then determined by the following rule:

$$\begin{aligned} l_{t+1} &= l_t + 1 \qquad \text{if } s_t = 1 \quad \text{and} \quad \pi(l_t, e_t) \geq \varepsilon_t \\ &= l_t \qquad \text{if } s_t = 0 \quad \text{or} \quad \pi(l_t, e_t) < \varepsilon_t \end{aligned} \tag{1}$$

with the following assumptions:

Assumption 1 The probability of success $P(l_{t+1} = l_t + 1 | e_t, s_t = 1) = F \circ \pi(l_t, e_t)$ is increasing and concave in effort e_t.

This assumption is satisfied when the performance function $\pi(l,e)$ is increasing and concave in e and the random terms ε are i.i.d. across individuals and periods and their c.d.f. F is concave. We assume that a person with gender g and completed grade l, is able to work (either on farm, or at home, or outside) and gets earnings $w(g,l)$ (again the model could be written with $w(x,g,l)$ where individual characteristics x affect earnings).

Assumption 2 The earnings function $w(g,l)$ is increasing in the acquired level of education l.

All these variables refer to year t when the index t is used. We assume that the cost for a child for going to school in year t, denoted $c(e_t)$ depends on the learning effort e_t (plus the cost of transportation, and other costs of enrollment).

Assumption 3 The cost function $c(e)$ is increasing and convex in e, the level of learning effort at school.

Then, sending a child to school in year t costs $c(e_t) - \tau(l_t, g, p)$, while not sending him generates earnings $w(g, l_t)$, the opportunity cost of enrolling the child in school. Assuming that the decision process in the household results in the maximization of the intertemporal expected benefits for the child $V(l_t, g, p, s_t)$, the value of enrolling a child at the beginning of year t ($s_t = 1$) or that of not enrolling him ($s_t = 0$) knowing his completed grade l_t, his gender g and eligibility p can be written recursively as follows:

$$\begin{aligned} V(l_t,g,p,1) &= \max_{e_t}\{\tau(l_t,g,p) - c(e_t) + \beta E[\max_{s_{t+1}\in\{0,1\}} V(l_{t+1},g,p,s_{t+1}) \mid s_t = 1]\} && (2) \\ V(l_t,g,p,0) &= w(g,l_t) + \beta E[\max_{s_{t+1}\in\{0,1\}} V(l_{t+1},g,p,s_{t+1}) \mid s_t = 0] && (3) \end{aligned}$$

with β the discount factor and l_{t+1} following the law (1).

Because of the uncertainty of grade progression, parents revise their expected optimal choice at the beginning of each schooling year.

The value function for a child of education l_t, gender g, and eligibility p can be written:

$$\phi(l_t,g,p) = \max_{s_t\in\{0,1\}} V(l_t,g,p,s_t) \tag{4}$$

Substituting in expressions (2) and (3) gives

$$\begin{aligned} V(l_t,g,p,1) &= \max_{e_t}\{\tau(l_t,g,p) - c(e_t) + \beta E[\phi(l_{t+1},g,p) \mid s_t = 1]\} \\ V(l_t,g,p,0) &= w(g,l_t) + \beta E[\phi(l_{t+1},g,p) \mid s_t = 0] = w(g,l_t) + \beta\phi(l_t,g,p) \end{aligned}$$

because $P(l_{t+1} = l_t + 1 \mid s_t = 0) = 0$. This implies

$$\phi(l_t,g,p) = \max\{\tau(l_t,g,p) + \max_{e_t}\{\beta E[\phi(l_{t+1},g,p) \mid s_t = 1] - c(e_t)\}, w(g,l_t) + \beta\phi(l_t,g,p)\} \tag{5}$$

Then, we can show the following proposition:

Proposition 1 *The function ϕ defined by the Bellman equation (5) exists and is unique.*

Proof. See Appendix C.1. ■

Intuitively, we expect the value function ϕ to be increasing with completed grade. Though one could imagine that at the end of secondary school, when transfers stop, the value function could drop, this is something that our empirical observation in Mexico does not support. Being more educated is always valuable despite the Progresa transfers, whatever their possible negative effects on enrollment and learning effort. Thus, in the rest of this paper, we will suppose that:

Assumption 4 The value function $\phi(l, g, p)$ is always increasing with completed grade l.

In Appendix B, we show which sufficient conditions on the primitives of the model ensure that the endogenous value function ϕ is always increasing with completed grade. However, rather than assuming these "reasonable" sufficient conditions we will simply impose the assumption that $\phi(l, g, p)$ is increasing in l.

3.1 Program Impact on Effort and Performance

We could consider, as often done, that learning activity depends on the exogenous characteristics of children and cannot be adjusted once presence at school is required. Then, effort would have to be considered exogenous in our model. However, it is clear that higher returns to education (in a very broad sense) constitute an incentive for students to study and learn more. Introducing an endogenous learning effort, the maximization of the value function implies that learning effort is chosen conditional on enrollment so as to maximize $\tau(l_t, g, p) - c(e) + \beta E[\phi(l_{t+1}, g, p) \mid s = 1]$.

Proposition 2 *The learning effort e^* is strictly positive if and only if $\phi(l+1, g, p) > \phi(l, g, p)$ and satisfies the first order condition*

$$\beta[\phi(l+1, g, p) - \phi(l, g, p)] f \circ \pi(l, e^*) \frac{\partial \pi}{\partial e}(l, e^*) = c'(e^*) \tag{6}$$

Proof. See Appendix C.2. ■

Since the performance technology depends on grade (difficulty of tests, selectivity system...), proposition 2 implies that performance at school can be highly non-linear and non-monotonic in the grade level. It can be either increasing or decreasing with grade and thus the expected performance at school can also be either increasing or decreasing with grade.

These results call for taking into account heterogeneity of treatment effects since the theoretical impact of the program on performance depends on the sign of $\frac{\partial}{\partial p}[\phi(l_t+1, g, p) - \phi(l_t, g, p)]$ as shown by the following proposition (for notational ease we use $\frac{\partial}{\partial p}$ as if p was continuous, that is for a function $H(p)$, $\frac{\partial}{\partial p} H(p) = H(1) - H(0)$).

Proposition 3 *Treatment raises effort and performance at school in terms of probability to succeed in a given grade* l_t *if* $\phi(l_t+1,g,1)-\phi(l_t,g,1)>\phi(l_t+1,g,0)-\phi(l_t,g,0)$ *and reduces effort and expected performance if* $\phi(l_t+1,g,1)-\phi(l_t,g,1)<\phi(l_t+1,g,0)-\phi(l_t,g,0)$.

Proof. As shown in Appendix C.2, if $\phi\left(l_t+1,g,p\right)-\phi\left(l_t,g,p\right)\leq 0$, then $e_t^*=0$. If $\phi\left(l_t+1,g,p\right)-\phi\left(l_t,g,p\right)>0$, then e_t^* satisfies (6). With assumptions 1 and 3, the implicit function theorem implies that e_t^* is increasing in p if $\phi\left(l_t+1,g,1\right)-\phi\left(l_t,g,1\right)\geq\phi\left(l_t+1,g,0\right)-\phi\left(l_t,g,0\right)$ and decreasing in p otherwise. As $\pi\left(l,e\right)$ is an increasing function of e, the same applies for performance. ■

This proposition indicates that the effect of the program on learning and performance at school will depend on whether the value of getting an additional year of education is higher for treated or untreated students. Its sign is ambiguous and depends on the effect of the program on the curvature of $\phi\left(l,g,p\right)$ with respect to l. It is to be noted that it can be that the program has a positive impact on effort at a grade l while a negative at another lower or higher grade l'.

Remark that the previous proposition gives us an empirical test of whether the learning effort is actually endogenous which is of great importance for education policies. Actually, under the assumption that random treatment does not affect the performance and evaluation technologies π and F, if treatment affects the probability to pass a given grade, it means that the learning effort is endogenous.

3.2 Program Impact on the Enrollment Decision

The decision of enrollment is derived from the comparison of the value of going to school and not going. Define the decision to enroll the child at school by $s_t=1_{\{v(l_t,g,p)\geq 0\}}$ (1 for school enrollment and 0 otherwise) where $v(l_t,g,p)=V(l_t,g,p,1)-V(l_t,g,p,0)$ is the difference between the two conditional value functions.

The following proposition is straightforward and shows the derivatives of $v(l_t,g,p)$ with respect to the program treatment p which represents the effect of treatment on the propensity to choose schooling over working (with the notation $\frac{\partial v(l_t,g,p)}{\partial p}=v(l_t,g,1)-v(l_t,g,0)$).

Proposition 4 *The program impact on the value of going to school compared to not going is:*

$$\begin{aligned}\frac{\partial v(l_t,g,p)}{\partial p} &= \tau(l_t,g,1)\\ &\quad+\ \beta\{\ \left[\phi\left(l_t+1,g,1\right)-\phi\left(l_t,g,1\right)\right]\frac{\partial}{\partial p}P(l_{t+1}=l_t+1\mid s_t=1)\\ &\qquad+\ P(l_{t+1}=l_t+1\mid s_t=1)\ \frac{\partial}{\partial p}\left[\phi\left(l_t+1,g,p\right)-\phi\left(l_t,g,p\right)\right]\ \}\end{aligned}$$

Proposition 4 shows that $\frac{\partial v(l_t,g,p)}{\partial p}$ is composed of several terms:

The first term, $\tau(l_t,g,1)\geq 0$, is the direct incentives to go to school provided by the educational transfer. The second term is the discounted expected marginal value of program eligibility composed of:

- The marginal increase of the probability of succeeding in grade progression times the marginal value of getting an additional year of education: $[\phi(l_t+1,g,p)-\phi(l_t,g,p)]\frac{\partial}{\partial p}P(l_{t+1}=l_t+1 \mid s_t=1)$. It comes from the fact that higher incentives to succeed are given to eligible students.

- The increase provided by treatment of the marginal value in getting one additional year of education, given the probability of successfully completing the current grade, $P(l_{t+1}=l_t+1 \mid s_t=1)\frac{\partial}{\partial p}[\phi(l_t+1,g,p)-\phi(l_t,g,p)]$.

Moreover, according to proposition 3, $\frac{\partial}{\partial p}P\left(l_{t+1}=l_t+1 \mid s_t=1\right)$ and $\frac{\partial}{\partial p}[\phi(l_t+1,g,p)-\phi(l_t,g,p)]$ are of the same sign. This model clearly shows the implications of the program on the value for children of going to school compared to that of not going. In particular, it helps explain that the incentives provided by the program do not only depend on the reduction of the opportunity cost of schooling by the conditional transfers but also on the additional value provided by the expectation to receive transfers the year after and the expected value of being more educated. Therefore, the program impact has no reason to be simply proportional to transfers received. According to this model, the treatment effect of the program should be heterogeneous across individuals with different grades, gender but also across all characteristics affecting future wages for example.

Another implication is that the incentives to go to school represented by $\frac{\partial v(l_t,g,p)}{\partial p}$ depend on the probability of grade progression, on the cash transfer corresponding to the current grade and the one for upper grades but not that of lower grades.

As the value function depend on the design of the program, effort will depend on the pattern of potential future transfers corresponding to higher grades. The impact of the program depends naturally on the year being evaluated. In the case of Progresa, the program being a fixed 3 year term, its impact in the first year will be different from that of the second year since the remaining years of transfers are different even if the transfer schedule by grades and gender is constant.

Note also that even if the transfer function for some grade l and gender g is zero, $\tau(l,g,1)=\tau(l,g,0)=0$, we still have $\frac{\partial v(l,g,p)}{\partial p}\neq 0$ if for some grade $l'>l$, $\tau((l',g,1)>0$. Because of the expected benefit from transfers in higher grades, the cash transfer program also generates incentives in favor of schooling even if the student is not entitled to receive any grant in his current grade. In the particular case of Progresa, this indicates a possible incentive to schooling even in the first and second year of primary school although students receive nothing in their current grade.

4 Identification and Econometric Evaluation

The theoretical model developed in the previous section gives testable implications regarding the impact of a cash transfer program on enrollment decisions and performance outcomes. Our model calls for two kinds of empirical estimation. First, one can study the reduced form equations of the two main endogenous choices that we modelled: the school continuation decision and the performance at school. This approach has the advantage of not relying on the model developed while the model allows to interpret the empirical result. Second, one could be willing to estimate structurally the model in order to estimate the different components of the theoretical effects of the program. One of the difficulty of performing a structural estimation comes form the identification of the value function of education. One possible solution consists in simulating this value function (Eckstein and Wolpin, 1999), which is however quite difficult if one wants to take into account the effect of heterogeneity in observables in the behavior of agents. In our case, we thus prefer to adopt a kind of semi-structural approach (explained in section 6) where the value function or its derivative which enter the structural model are estimated instead of simulated.

We exploit this by estimating passage probabilities from grade to grade $P(l_{t+1} = l_t + 1 \mid s_t = 1)$ and enrollment decision $P(s_{t+1} = 1 \mid s_t = 1)$[9] for the years 97-98 and 98-99. Figure 1 shows the time line of events and decisions.

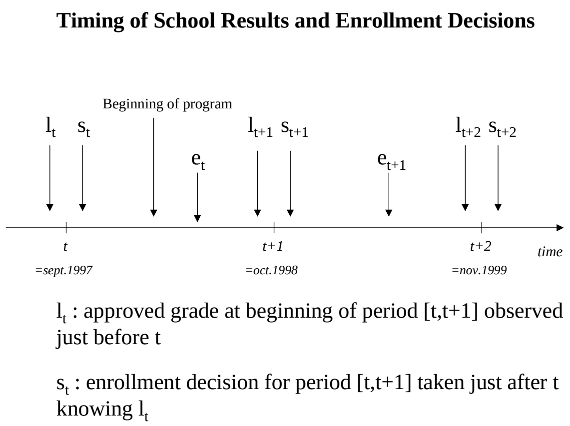

Figure 1: Timing of the Dynamic Decision Process

4.1 Econometric Specification

To simplify notations, the theoretical model presented is implicitly conditional on a set of exogenous characteristics that could affect the wage or costs of schooling that have to appear in the econometric

[9] Most estimation will be based on parametric maximum likelihood. The semiparametric identification of binary choice models is possible with some assumptions like location and scale normalization (Manski, 1985, 1988) but given the number of explanatory variables used in our regressions we will use simple parametric estimation methods.

specifications. Adding exogenous characteristics x_t to the grade progression model (1), we have: $P(l_{t+1} = l_t + 1 \mid s_t = 1) = F \circ \pi(x_t, l_t, e_t^*)$ where e_t^* is endogenously determined and depends on x_t, l_t, g, p. Therefore, we specify the grade progression as:

$$P(l_{t+1} = l_t + 1 \mid s_t = 1) = \varphi(X_t \gamma_{l_t,g} + \theta_{l_t,g} p) \tag{7}$$

where $\gamma_{l_t,g}$, $\theta_{l_t,g}$ are vectors of parameters specific to grade l_t and gender g, X_t is a vector of exogenous variables (including x_t, l_t, g), and φ is a c.d.f. (for example logistic or normal)[10]. Assuming that some unobserved component ξ of cost $c(e_{t+1})$ or wage $w(x_{t+1}, g, l_{t+1})$ is randomly distributed with logistic or normal c.d.f. φ, we get a structural form for of the probability of continuing school presented in section 6.

The reduced form of the model does not allow to decompose the effect of the program on each incentive component identified in the theoretical model. However, it allows to evaluate the total program impact on enrollment and performance at school and is more robust to mispecification than the structural model. We specify the grade progression probability in a reduced form as

$$P(s_{t+1} = 1 \mid s_t = 1) = \varphi(Z_{t+1}\delta_{l_{t+1},g} + \alpha.\tau(l_{t+1}, g, p)\, p + [Z_{t+1}p]\, \alpha^2 + [X_{t+1}p]\, \alpha^3) \tag{8}$$

where Z_{t+1} is a vector of exogenous variables (left implicit in the conditioning set). Of course, since the design of the program is such that transfers are gender and grade specific, coefficients $\alpha' s$ will be allowed to be gender and grade specific.

Our model thus leads us to estimate the impact of the program on two equations of interest: the probability of continuing school $P(s_{t+1} = 1 \mid s_t = 1)$ and the probability of passing grade $P(l_{t+1} = l_t + 1 \mid s_t = 1)$.

4.2 Identification of the Program Impact and the Dynamic Selection Problem

As shown by Cameron and Heckman (1998), the estimation of school transition models faces a problem of dynamic selection bias. Even if unobserved factors entering the school transition model are distributed independently of observable characteristics in the population enrolling in the first year of primary school (for example), the distribution of unobserved characteristics of students in the second year of primary school will be truncated and not independent of the distribution of observable characteristics because of the educational selection of students. This dynamic selection of the population of students introduces a bias in the estimation of transition models. Here, this difficulty certainly affects the estimation of probabilities to enroll at school and probabilities to successfully pass a grade. However, with randomization of treatment, the evaluation of the average

[10] This specification is consistent with the theoretical model for example (but not only) if the c.d.f. F of ε is normal or logistic and the performance function π is a linear index of its arguments.

program impact on these passage probabilities will not be biased by this dynamic selection problem. We explicitly formulate the necessary assumptions for identification and establish the relationship between randomization and the dynamic selection problem.

Grade transition probabilities conditional on the vector of observables $\omega_{t+1} = (Z_{t+1}, X_t, l_{t+1}, s_t)$, the treatment dummy $p \in \{0,1\}$ and unobserved characteristics $\widetilde{\theta}$ can be written[11]

$$E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) = \psi(\omega_{t+1}, p, \widetilde{\theta}) \tag{9}$$

where $\psi\left(.\right)$ a real valued function.[12] It is to be noted that with these notations, $\frac{\partial}{\partial p}\psi(\omega_{t+1}, p, \widetilde{\theta})$ can be seen as corresponding to the marginal treatment effect defined by Heckman and Vytlacil (1999, 2000a, 2000b) since the unobserved variable $\widetilde{\theta}$ is likely to affect participation of an individual into the corresponding grade where treatment is received.

As $\widetilde{\theta}$ is unobserved, we cannot identify $E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ but are rather interested in the average $E(s_{t+1} \mid \omega_{t+1}, p) = E_{\widetilde{\theta}}[E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})]$. The parameters of interest that we would like to identify are the average program impact

$$E_{\widetilde{\theta}}[\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] \tag{10}$$

and the average effect of some covariates ω_{t+1}

$$E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] \tag{11}$$

Cameron and Heckman (1998) showed clearly that even if the distribution of $\widetilde{\theta}$ is independent of ω_0 ($\widetilde{\theta} \perp\!\!\!\perp \omega_0$), this random effect assumption for the initial schooling stage will not be true for the subsequent ones because of the selection of students; that is, in general $\widetilde{\theta} \not\perp\!\!\!\perp \omega_{t+1}$. This dynamic selection bias implies that

$$\frac{\partial}{\partial \omega_{t+1}}E\left(s_{t+1} \mid \omega_{t+1}, p\right) = \frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}}\psi(\omega_{t+1}, p, \widetilde{\theta})] \neq E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1}, p, \widetilde{\theta})]$$

The value $\frac{\partial}{\partial \omega_{t+1}}E\left(s_{t+1} \mid \omega_{t+1}, p\right)$ is thus a biased estimator of $E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1}, p, \widetilde{\theta})]$ (the derivative of the average $E_{\widetilde{\theta}}\psi(\omega_{t+1}, p, \widetilde{\theta})$ is not equal to the average derivative). In the schooling transition probabilities, we will have biases equal to $B(\omega_{t+1}, 1) = \frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}}E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta})]$ for the treatment population and to $B(\omega_{t+1}, 0) = \frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}}E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta})]$ for the control population. Each bias being difficult to sign and quantify *a priori* (as in Cameron and Heckman, 1998), a solution is then to model the unobserved component $\widetilde{\theta}$, for example by using the Heckman and Singer (1984) technique introducing an arbitrary discrete non-parametric distribution for $\widetilde{\theta}$.

[11] Note that $E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) = P(s_{t+1} = 1 \mid \omega_{t+1}, p, \widetilde{\theta})$.

[12] The unobserved component $\widetilde{\theta}$ may be multidimensional without changing the following results.

However, as proposition 5 shows below, we do not encounter the same problem when evaluating the average program impact

$$\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p) = E(s_{t+1} \mid \omega_{t+1}, p=1) - E(s_{t+1} \mid \omega_{t+1}, p=0)$$

Actually, first note that randomization implies that treatment is orthogonal to observed and unobserved characteristics

$$p \perp\!\!\!\perp (\widetilde{\theta}, \omega_{t+1}) \tag{12}$$

which implies (13) that can be used in the following proposition.

Proposition 5 *If treatment $p \in \{0,1\}$ is orthogonal to the distribution of unobserved characteristics conditional on observables ω_{t+1} that is*

$$p \perp\!\!\!\perp \widetilde{\theta} \mid \omega_{t+1} \tag{13}$$

then

$$\frac{\partial}{\partial p}E\left(s_{t+1} \mid \omega_{t+1}, p\right) = \frac{\partial}{\partial p}[E_{\widetilde{\theta}}\psi(\omega_{t+1}, p, \widetilde{\theta})] = E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\psi(\omega_{t+1}, p, \widetilde{\theta})] \tag{14}$$

Proof. Proof in Appendix C.3. ■

With the randomization of treatment, property (13) is satisfied and Proposition 5 applies. Recall that $\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, p)$ is not identifiable because of the dynamic selection problem. However, we might still be interested in identifying how the program impact depends on ω_{t+1} which is $\frac{\partial}{\partial p}\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, p) = \frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, 1) - \frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, 0)$. The question is then to compare $B(\omega_{t+1}, 1)$ and $B(\omega_{t+1}, 0)$ because if both biases are the same in the treatment and control groups then $\frac{\partial}{\partial p}\frac{\partial}{\partial \omega_{t+1}}E(s_{t+1} \mid \omega_{t+1}, p) = \frac{\partial}{\partial p}E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1}, p, \widetilde{\theta})]$. We have the following Proposition:

Proposition 6 *The average treatment effect $E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\frac{\partial}{\partial \omega^k_{t+1}}\psi(\omega_{t+1}, p, \widetilde{\theta})]$ is identified and equal to $\frac{\partial}{\partial p}\frac{\partial}{\partial \omega^k_{t+1}}E\left(s_{t+1} \mid \omega_{t+1}, p\right)$ if one of the following condition is satisfied*[13]

$$\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0 \tag{15}$$

or

$$\int E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta})\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = \int E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta})\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \tag{16}$$

[13] The notation λ is used to designate cumulative distribution functions. For example $\lambda(\widetilde{\theta} \mid \omega_{t+1})$ is the c.d.f. of $\widetilde{\theta}$ conditional on ω_{t+1}.

Proof. See Appendix C.4. ■

Condition (15) means that the distribution of $\widetilde{\theta}$ does not depend on ω_{t+1}^k i.e. that there is no dynamic selection bias in the direction of ω_{t+1}^k. Condition (16) means that the marginal treatment effect $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ averages to zero when integrating with respect to $\frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1})$ which is always the case if $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ is constant across $\widetilde{\theta}$ because $\int \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$ since $\int d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \equiv 1$. Therefore, this is always true if the average treatment effect $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ does not depend on ω_{t+1}.

In the present case, neither assumption (15) nor (16) has to be valid given the randomization procedure. (15) will be wrong as soon as there is some dynamic selection which is likely in education transition models and (16) is unlikely to happen as soon as the treatment effect $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ depends on covariates ω_{t+1}. Therefore, the randomization process insures only the identification of the average impact $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p)$. The same argument can be applied to the performance probability $P\left(l_{t+1} = l_t + 1 \mid s_t = 1\right)$. The randomization condition (13) is sufficient to ensure that the dynamic selection bias present in the estimation of the conditional probabilities $P\left(s_{t+1} = 1 \mid s_t = 1\right)$ and $P\left(l_{t+1} = l_t + 1 \mid s_t = 1\right)$ will be the same for treated and untreated sample and will then cancel out in the estimation of the program impact.

The condition (12) that the joint distribution of observables and unobservables is independent of treatment (i.e. is the same across treated and control samples) is not testable, but an implication of it on the marginal distribution of observables can be checked (and empirically validated for data in 1997 by Behrman and Todd (1999), Schultz (2004)).While we can assume safely that the randomization of the program placement in the case of Progresa is such that the condition (12) is true at the beginning of the program in 1997, we can expect that this will not be true afterwards. Actually, the program impact being probably non zero (as the empirical results will confirm), the dynamic selection bias will not cancel out across treatment and control groups. Therefore, we expect that the conditional probabilities estimates of the program impact between 1998 and 1999 will be biased by a dynamic selection bias due to the impact of the program. For example, if the program has a positive effect on the propensity of continuing schooling, it can select individuals with (on average) lower unobserved factor also causing an increase in the individual propensity to go to school (like unobserved ability). This in turn would bias downward the probability to succeed in class estimated the following year.

4.3 Identifying the Elasticity of the Program Impact to Cash Transfers

Until now we have investigated the estimation of the average program impact. However, one may be interested in the elasticity of the program impact to the amount of cash transfers that is

$$\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T) \tag{17}$$

where T is the transfer received by the student (the previously defined treatment dummy is $p = 1_{(T>0)}$). To explain the identification method, we derive the following proposition:

Proposition 7 *Assume that there exists a random variable ω'_{t+1} such that the transfer T is $\widetilde{\tau}(g, l, p, \omega'_{t+1})$ and the following assumptions are satisfied:*
The average treatment effect $\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T)$ does not depend on ω'_{t+1} i.e.

$$\frac{\partial}{\partial \omega'_{t+1}}\left(\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T)\right) = 0 \qquad \text{(Exclusion Restriction)}$$

The program rule $\widetilde{\tau}(g, l, p, \omega'_{t+1})$ is such that

$$\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1}) \neq 0 \textit{ is known} \qquad \text{(Known Conditionality of Program Rule on Observables)}$$

and does not depend on unobservables $\widetilde{\theta}$

$$\frac{\partial}{\partial \widetilde{\theta}}\{\widetilde{\tau}(g, l, p, \omega'_{t+1})\} = 0 \qquad \text{(Program Rule Independent of Unobservables)}$$

The observed component ω'_{t+1} is independent of unobserved factors $\widetilde{\theta}$ conditionally on ω_{t+1}

$$\omega'_{t+1} \perp\!\!\!\perp \widetilde{\theta} \mid \omega_{t+1} \qquad \text{(IV assumption)}$$

Then $\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T)$ identifies[14] *$E_{\widetilde{\theta}}[\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta})]$.*

Proof. Proof in Appendix C.5. ■

In the Progresa program, this identification is provided by the maximum rule as follows[15]. The Progresa rules stipulate that household transfers cannot exceed some given maximum amount of money and impose a proportional adjustment rule for individual benefits. Table 4 gives examples of this proportional adjustment in terms of transfers to be received. The last column shows what is the transfer due for each child given this adjustment for family A and B which otherwise would get more than the maximum amount allowed while family C does not reach this amount. This monthly amount corresponds to what is lost if a child misses school without justification.

[14] Of course only within the range of variation of T in the data observed.

[15] Without this rule, the value of transfers $T = \tau(g, l, p)$ are conditional only on gender g and grade l which are very likely to be correlated with the individual unobservable components $\widetilde{\theta}$. Then, if transfers do not vary across individuals (conditionally on ω_{t+1}), $\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T)$ is not identifiable and only the average treatment effect $\frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1}, p)$ is identified.

Table 4: Example of the Maximum Rule

Example of the Maximum Rule in 1997	Progresa Grant	
	Without	With Proportional
Family A:	Adjustment	Adjustment
One Boy in Secondary School (1^{st} year)	175	$\frac{175}{605}\times 550 \quad = 159$
One Girl in Secondary School (2^{nd} year)	205	$\frac{205}{605}\times 550 \quad = 186$
One Girl in Secondary School (3^{rd} year)	225	$\frac{225}{605}\times 550 \quad = 205$
Total received by household	605	550
Family B:		
One Girl in Secondary School (1^{st} year)	185	$\frac{185}{585}\times 550 \quad = 173$
One Girl in Secondary School (2^{nd} year)	205	$\frac{205}{585}\times 550 \quad = 193$
One Boy in Secondary School (3^{rd} year)	195	$\frac{195}{585}\times 550 \quad = 184$
Total received by household	585	550
Family C:		
One Girl in Primary School (6^{th} year)	120	120
One Girl in Secondary School (2^{nd} year)	205	205
One Boy in Secondary School (3^{rd} year)	195	195
Total received by household	520	520

Noting T' the total transfer that the household would receive in absence of this maximum rule and M_{t+1} the maximum amount of money the household can receive at time $t+1$, the actual transfer received for a child is the known function

$$T = \tau(g, l, p) \min \left\{ \frac{M_{t+1}}{T'}, 1 \right\}$$

The assumption needed for identification is that the random variable $\min \left\{ \frac{M_{t+1}}{T'}, 1 \right\}$ does not affect the average treatment effect $\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T)$ i.e. that given observables ω_{t+1} the average effect of transfer T on schooling s_{t+1} is constant across values of T'. Concretely, it means that the effect of transfer T on individual schooling can depend on observable characteristics of a student but that conditionally on these characteristics ω_{t+1} there are other observable characteristics that affects T' but not the treatment effect. For example, the number of children of the household which generates variation in the individual transfer amount (because some reach the maximum and others not) may have a direct effect on the average treatment effect. However, conditionally on the number of children with for example equal number of boys and girls, it may be more reasonable to assume that the order of gender of children does not affect the average treatment effect directly while it provides some variation in the amount of transfers received. Table 4 shows examples of families with the same number of children, the same number of boys and girls but for which individual transfers of the girl in the second year of secondary school vary because of this rule. The presence at school of a second year secondary school girl will not bring the same transfer if she belongs to family A, B or C in the example of Table 4.

We therefore exploit this kind of variation and assume that conditionally of ω_{t+1} (which includes the number of children) the fact that the household reaches the maximum or not is random and uncorrelated with the unobserved characteristics $\widetilde{\theta}$ because it comes mainly from the random distribution of genders within the family. The conditions of identification given by Proposition 7 are then plausible even if not testable.

5 Empirical Results and Policy Implications

Given the identification issues shown in section 4, we first compute the average treatment effects on transition probabilities of the reduced form model i.e. the probability of continuing school $P(s_{t+1} = 1 \mid s_t = 1)$ and the probability of progressing by one grade $P(l_{t+1} = l_t + 1 \mid s_t = 1)$ by estimating a logit of the outcome discrete variable conditional on treatment and households characteristics. In this case, the average effect is the coefficient of the dummy variable for treatment (p). In a second step, using the Progresa rules and according to the identifying strategy described in section 4.3, we identify the effect of the transfer value on these outcome variables. In the data, the sample proportions of deviations from the pre-set amount because of the maximum benefit rule is of 14% in 97, 9% in 98, and 13% in 99.

The notations for grades is P1, P2, .. for primary school grades from 1, 2, .. and S1, S2 .. for secondary school grades. The tables of results by grade display the attended grade by the student in rows.

5.1 Performance

Results of the estimated probability of grade progression during school year 1997-98, are fully reported in Tables of Appendix D[16]. These probabilities are estimated using a logit model[17] with standard errors clustered at the village level. The means of marginal effects denoted $\overline{(\partial\varphi/\partial x)}$ are computed as means of the observation by observation marginal effects for each variable and presented together with coefficient estimates. Tables P98-1, P98-2 and P98-3 also present the full set of dummy variables included in the regressions and shows in particular that household size has a negative effect on performance, household head education and gender have no significant effect, and the student's age a negative one. In tables 5 and 6, the first two columns summarize the results of the average impact of transfers and the last two columns present the results when estimating

[16] Table P98-1 shows the estimates of the average treatment effects over all grades. Table P98-2 shows the average program impact by gender for primary and secondary school. Table P98-3 shows the average program impact by grade and gender. Moreover, all these Tables also present the estimates of the elasticity of program impact to transfers (17) which is identified according to Proposition 7.

[17] Results using a probit model are very similar. Though coefficients have different absolute values, as usual, means of marginal effects are very close when estimating with logit or probit.

the elasticity to transfers.

Table 5 presents the most important estimates. First, the average program impact on students is significantly positive. The average treatment effects present a 2 percentage point increase in the probability to successfully complete the grade. The average elasticity to transfers of performance (rows denoted by $\theta.\tau(l, g)$) is negative but Table P98-3 shows that when the program impact is estimated by grade and gender, there is no more negative elasticity[18]. The average treatment effect by gender for primary and secondary school are very different. The means of marginal effects for primary school are positive with a 6.1% increase in performance both for boys and girls while it is negative at secondary school with means of marginal effects of -21% for girls and -17% for boys. This result can be interpreted by the fact that the cash transfer program has a negative impact on learning effort because students want to remain as long as possible in the program. At primary school, they seem to increase their learning effort (willingness to benefit from higher transfers, better learning condition thanks to the program benefits in cash but also including better health care and nutrition components) but at secondary school the program has a negative effect because the probability of repetition increases.

The elasticity to transfers are still negative at primary school but positive at secondary school with means of marginal effects for girls of 7% and 10% for boys meaning that a 100 pesos increase in transfers of secondary school would increase the performance probability of 10% for boys and 7% for girls (in all Tables, the value of transfers are in hundreds of pesos). Finally, Table P98-3 in Appendix D shows the average treatment effect by gender and each grade level of primary and secondary school. It appears that a significant positive effect is found for the 3^{rd} year of primary school with means of marginal effects of 5.2% for girls and 4.3% for boys. Moreover, a negative significant effect is found on the 1^{st} and 3^{rd} year of secondary school with means of marginal effects of 37% for girls and 29% for boys in the 1^{st} year and 19% for girls and 26% for boys in the 3^{rd} year. When looking at the elasticity to transfers, we find however a positive effect for the 1^{st} and 2^{nd} year of secondary school meaning that the negative average effect is lower the larger are transfers. These negative effects on, performance at secondary school can be explained by the willingness to continue to benefit from Progresa transfers for more time at secondary school. For example, at the 3^{rd} year of secondary school this is explained by the fact that for higher levels the Progresa program is not active anymore.

These results on the performance probability also show that the learning effort is endogenous and

[18] When estimating the average effect over all grades, the value of the transfer has a negative impact, meaning that larger transfers have a lower impact. This is probably due to the fact that transfers increase with grade and can be explained if (for example) selectivity of the educational system increases such that the program impact is lower for higher grades.

affected by financial incentives given by the Program (but also through the effect of transfers on educational conditions of children with better food and nutrition and health care).

Table 5: Summary Results of Treatment Effects for Performance (1997-98)[19]

$P(l_{t+1}=l_t+1 \mid s_t=1)$		$\overline{(\partial\varphi/\partial x)}$	(t-stat)	$P(l_{t+1}=l_t+1 \mid s_t=1)$		$\overline{(\partial\varphi/\partial x)}$	(t-stat)
Average Treatment Effect				Including transfer amount $\tau(l,g)$			
$\theta.p$		0.020*	(2.63)	$\theta.p$		0.182*	(11.71)
				$\theta.\tau(l,g)$		-0.179*	(-11.27)
$\theta_{l,g}.p$	Primary, girls	0.061*	(7.36)	$\theta_{l,g}.p$	Primary, girls	0.194*	(21.26)
	Primary, boys	0.060*	(7.13)		Primary, boys	0.216*	(23.09)
	Secondary, girls	-0.219*	(-4.73)		Secondary, girls	-0.489*	(-4.09)
	Secondary, boys	-0.172*	(-3.76)		Secondary, boys	-0.528*	(-4.65)
				$\theta_{l,g}.\tau(l,g)$	Primary, girls	-0.287*	(-12.41)
					Primary, boys	-0.336*	(-13.86)
					Secondary, girls	0.071	(1.71)
					Secondary, boys	0.107*	(2.38)

For the year 98-99, it is important to recall that the conditions for identification (propositions 5 and 7) are likely to be invalid in these cases since Progresa was already active the year before which implies that the randomization assumption (13) cannot be ascertained in this case. However, it is interesting to remark that in this case the average program impact is not significantly different from zero[20]. By gender and grades in 98-99, we then find positive effect at primary school with means of marginal effects of 4.4% for girls and 4.6% for boys which is a bit lower than for the previous year 1997-98.

Note that the effects on performance for the 1^{st} and 2^{nd} year of primary school (l = 3, 4) are not identified (not in Table P98-3) given all the conditioning variables included in the regressions. However, when not conditioning on the full set of grade-gender dummies but only on primary school dummy and grade-gender dummies for secondary school, then the treatment effect on 1^{st} and 2^{nd} year of primary school is identified and found significantly positive. This is interesting because it shows an indirect effect of Progresa on the first two years of primary school which are not directly targeted by transfers.

5.2 School Continuation

Equation (8) is estimated. As for the performance probability, results of the estimated probability of school continuation for 1997-98 (in Fall 98) are reported in Tables E98-1, E98-2 and E98-3 in Appendix D. These probabilities are estimated using a logit model with standard errors clustered at the village level and the means of marginal effects are presented together with coefficient estimates.

[19] In all Tables, transfers τ are in hundreds of pesos, and * means estimate is significantly different form zero at 5% level.

[20] Results of the estimated probability of grade progression during school year 1998-99 are not detailed here but can be provided upon request.

All these Tables also present the estimates of the elasticity of program impact to transfers (17) in rows denoted $\alpha.\tau(l,g)$ that are identified according to proposition 7. Tables E98-1, E98-2 and E98-3 present the full set of right hand side variables used and they show in particular that the household's head education level has a positive effect, age and distance to nearest secondary school have a negative effect, male gender has a positive effect while household size has no significant effect.

Table 6: Summary Results of Treatment Effects for Continuation (1997-98)

$P(s_{t+1}=1 \mid s_t=1)$		$\overline{(\partial\varphi/\partial x)}$	(t-stat)	$P(s_{t+1}=1 \mid s_t=1)$		$\overline{(\partial\varphi/\partial x)}$	(t-stat)
Average Treatment Effect				Including the transfer amount $\tau(l,g)$			
$\alpha.p$		0.035*	(7.72)	$\alpha.p$		0.022*	(2.39)
				$\alpha.\tau(l,g)$		0.009	(1.68)
By school level and gender							
$\alpha(l,g).p$	Primary, girls	0.031*	(3.86)	$\alpha(l,g).p$	Primary, girls	0.034	(1.29)
	Primary, boys	0.031*	(3.78)		Primary, boys	0.046	(1.72)
	Secondary, girls	0.034*	(5.79)		Secondary, girls	-0.058	(-1.01)
	Secondary, boys	0.032*	(5.22)		Secondary, boys	-0.026	(-0.48)
				$\alpha(l,g).\tau(l,g)$	Primary, girls	-0.004	(-0.11)
					Primary, boys	-0.021	(-0.58)
					Secondary, girls	0.044*	(2.14)
					Secondary, boys	0.031	(1.30)

Table 6 shows that the average program impact on students is significantly positive which is consistent with Schultz (2004), Behrman et al. (2006). The average treatment effect presents a 3.5% increase in school continuation. The average elasticity to transfers of performance is not significantly different from zero. The average treatment effect for primary school is of 3.1% increase in school continuation for boys and girls and of 3.4% for girls and 3.2% for boys at secondary school. Elasticity to transfers is positive for girls at secondary school with a 100 pesos increase in transfers leading to a 4.4% increase in school continuation.

Table E98-3 in Appendix D shows that when significant there is a positive effect by grade and gender on enrollment at primary and secondary school. The means of marginal effects are around 3-4% for primary school and 3.5% in the 1^{st} year of secondary school but insignificant in the 2^{nd} and 3^{rd} years of secondary school. Concerning the elasticity of school continuation to a hundred pesos transfer, it is significantly positive for the 1^{st} year of secondary school with means of marginal effects of 4.7% for girls and 5.2% for boys.

As for 1998-99[21], it is found the average treatment effect presents a 3.1% increase in school continuation which is a little lower than in 1997-98. The average elasticity to transfers of performance is not significantly different from zero. The means of marginal effects for primary school are of

[21] Detailed results are not provided but available upon request.

3.2% increase in school continuation for boys, 3.6% for girls and of 3.7% for girls and 2.2% for boys at secondary school. Elasticity to transfers is positive but not significantly different from zero.

6 Semi-structural Estimation

Finally, we also estimate our model in a semi-structural way. It relies on stronger assumptions than the "reduced forms" studied until now because of the use of parametric distributions for the unobservables, but allows to estimate the different behavioral components of our model. Instead of doing a fully structural estimation of the value function ϕ, we treat the pointwise values of this function as parameters to be estimated. This method has the advantage of not having to rely on strong structural assumptions for the simulation of the value function (like the simplification of the role of heterogeneity due to the computational cost of the simulation). For this, we assume that conditionally on all exogenous variables included in our main equation, the role of the value of education is only specific to grade and gender (of that the effect of exogenous characteristics on this value is additively separable from grade and gender). Then, assuming that some unobserved component ξ of the cost $c\left(e_{t+1}\right)$ or wage $w(x_{t+1}, g, l_{t+1})$ is randomly distributed with logistic or normal c.d.f. φ. Using Proposition 4 and equation (7), we can write

$$\begin{aligned} v\left(l_{t+1}, g, p\right) &= p\left[\alpha^1_{l_{t+1},g}\tau\left(l_{t+1}, g, p\right) + \alpha^2_{l_{t+1},g}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p) + \alpha^3_{l_{t+1},g}\frac{\partial}{\partial p}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p)\right] \\ &\quad + \alpha^4_{l_{t+1},g}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p) + Z_{t+1}\delta_{l_{t+1},g} - \xi \end{aligned}$$

where $\alpha^2_{l_{t+1},g}$ corresponds to $\beta\frac{\partial}{\partial p}[\phi\left(l_t+1, g, p\right) - \phi\left(l_t, g, p\right)]$, $\alpha^3_{l_{t+1},g}$ to $\beta[\phi\left(l_t+1, g, 1\right) - \phi\left(l_t, g, 1\right)]$, Z_{t+1} is a vector of exogenous variables and $\delta_{l_{t+1},g}$ a vector of parameters specific to grade and gender. The term $\alpha^4_{l_{t+1},g}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p)$ is added to $Z_{t+1}\delta_{l_{t+1},g}$ in order to capture non linear effects of covariates on the base state value of education, but it can be interpreted also as the effect of the probability to succeed on the probability to continue schooling in the base state because the probability of continuing school is $P\left(s_{t+1}=1 \mid s_t=1\right) = E(1_{(v(l_{t+1},g,p)\geq 0)} \mid s_t=1)$. The probability to continue schooling is then

$$\begin{aligned} P(s_{t+1} &= 1 \mid s_t=1) = \varphi\Big(Z_{t+1}\delta_{l_{t+1},g} + \alpha^1_{l_{t+1},g}\tau(l_{t+1}, g, p)p + \alpha^2_{l_{t+1},g}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p)p \\ &\quad + \alpha^3_{l_{t+1},g}\frac{\partial}{\partial p}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p)p + \alpha^4_{l_{t+1},g}\varphi(X_{t+1}\gamma_{l_{t+1},g} + \theta_{l_{t+1},g}p)\Big) \end{aligned} \quad (18)$$

where the parameters α^k's are also grade and gender specific.

The theoretical model predicts that $\alpha^1 \geq 0$. The value function being increasing and concave, the model predicts that $\alpha^3 \geq 0$ and that α^2 is of the sign of the change in the performance probability implied by the program impact on this gender-grade category of individual $(\frac{\partial}{\partial p}P(l_{t+1} = l_t+1 \mid s_t=1))$.

Table 7 shows the estimation results of (7) and (18). Random terms are assumed normally distributed. The identification of the semi-structural form relies on the parametric functional form chosen; this is a stronger assumption than the identifying assumptions of the reduced forms presented before[22].

The results of Table 7 show that the direct effects of Progresa α^1's on school continuation due to transfers are positive with means of marginal effects between 4.7 and 7.8% for primary school. At secondary school, the means of marginal effects of these direct effects are significantly positive and around 7.3-8.8%. We also see that expected success in school induces enrollment ($\alpha^4 > 0$). The marginal impact of a one percentage point increase in the probability to successfully finish a grade is to induce an increase in enrollment by 1.5 to 3 percentage points in primary school and by 3.5 to 5.7 percentage points in secondary school.

Note that under the Progresa program, the enrollment rate is less sensitive to expected performance in secondary school (because $\alpha^2 < 0$). This result is consistent with the prediction of our model since it is also at secondary school that we observe a negative impact on performance. For example, a one percentage point increase in expected performance raises enrollment by 1.5 percentage points under Progresa rather than 3.8 percentage points in absence of Progresa. The fact that the estimates of coefficients α^2's are significantly negative at secondary school means that the change in the value of one additional year of education at secondary school generated by Progresa is negative; that is the value function of education is less increasing in grade level when the program is active than when it is not. This can be a consequence of the internalization by students of the full set of incentive effects on learning and of the program termination after the 3^{rd} year of secondary school on the value of education. Finally, the results are consistent with predictions of Proposition 4 because the coefficients α^3's when significantly different from zero are positive. This coefficient is significantly different from zero for girls in the 1^{st} year of secondary school with means of marginal effects of 1.45 which means that the value of one additional year of education is such that a one percent increase in the expected probability of passing the grade raises by 1.45% the enrollment probability.

In conclusion, Progresa has a strong positive direct effect on enrollment of 5 to 8 percentage points depending on the grade. This is however mitigated by two negative indirect effects. Actually, as Progresa has a negative effect on performance, it also indirectly decreases the value of going to school. The second effect is due to the fact that Progresa decreases the sensitivity of the enrollment decision to performance. The reduced form estimates, showed that the overall net effect of Progresa

[22] Note also, that coefficient estimates cannot be compared across estimations but only marginal effects are comparable.

is an increase in school continuation of around 3-4%. Although one has to be cautious given the parametric assumptions done, this semi-structural estimation shows that the direct positive incentive effect of Progresa would have been larger if one could avoid the negative effect of the brutal program termination after the third year of secondary school. A lesson to be remembered for the design of future conditional transfer programs in poor countries.

7 Conclusion

We analyzed the effects on school enrollment and performance of a conditional transfers program in Mexico. We provided a theoretical framework for analyzing the dynamic educational decision and process including the endogeneity and uncertainty of performance (passing grades) and the effect of a randomly placed conditional cash transfer program for children enrolled at school. This framework is used to study the Progresa program for which a randomized experiment has been implemented in rural Mexico and that allows us to identify the effect of the conditional cash transfer program on enrollment and performance at school. We found that Progresa had a positive impact on school continuation. However, the program seems to have a positive impact on performance at primary school but a negative one at secondary school. This empirical fact is a possible consequence of the disincentives provided by the program termination after the third year of secondary school. Interestingly, this paper showed the role of financial incentives on education in a developing country. Incentives concern both schooling decisions and learning. It seems substantively important to take into account the endogenous learning of students. In addition, the results also show that this kind of program also modifies the endogenous value function of education and thus the complete educational behavior of students directly targeted or not by the program. This cash transfer program conditional on school attendance proves some success in reducing school drop out but evidence of perverse negative effects of repetition of classes implies that the design of these conditional programs must be very careful.

A Data

The data used were provided by Progresa. To construct our variables and sample we used all the available relevant information from the following data sets: ENCASEH97 (Encuesta de Características Socioeconomicas de los Hogares), ENCEL (Encuesta de Evaluacion) of March 1998, October 1998, June 1999, and November 1999.

The schooling variable used in the analysis corresponds to the question on whether the child is currently going to school which means both enrollment and non permanent absenteeism.

The variables on grades correspond to the question on what is the last grade completed by the child. It is assumed that he or she is then entitled to enroll at the upper grade. Moreover, we use intermediate evaluation surveys like those of March 1998 or June 1999 to check consistency of the data across years, to correct the obviously erroneous answers and complete non response that sometimes happen in a given survey.

B Increasing Value function

Corollary 8 *The value function ϕ is increasing in l if individuals are sufficiently patient ($\beta \geq 1/2$) and transfers are increasing with grade (or not too much decreasing i.e. transfers not too large when the program ends) and not too large compared to the potential wage (conditions likely to be true in the case of Progresa).*

Proof. Simplifying notations by avoiding indices g and p when there is no ambiguity, we know that ϕ exists, is unique and is the fixed point solution of T where $\forall l : T(\phi)(l) = \max\{w(g,l) + \beta\phi(l), \max_{e\geq 0}\{\tau(l,g,p) - c(e) + \beta E[\phi(l'(l)) \mid s = 1]\}\}$ *s.t.* $l'(l) = l + 1$ if $s = 1$ and $\pi(l,e) \geq \varepsilon$ and $l' = l$ otherwise. T being a contraction mapping, the fixed point solution will be increasing in l if ϕ increasing in l implies $T(\phi)$ increasing in l.

$T(\phi)(l+1) - T(\phi)(l)$

$= \max\{w(g,l+1) + \beta\phi(l+1), \max_e\{\tau(l+1,g,p) - c(e) + \beta E[\phi(l'(l+1)) \mid s = 1]\}\}$

$- \max\{w(g,l) + \beta\phi(l), \max_e\{\tau(l,g,p) - c(e) + \beta E[\phi(l'(l)) \mid s = 1]\}\}$ where $l'(l) = l+1$ if $s = 1$ and $\pi(l,e) \geq \varepsilon$ and $l' = l$ otherwise.

So $T(\phi)(l+1) - T(\phi)(l)$ is the largest of the four following values:

a) $w(g,l+1) - w(g,l) + \beta[\phi(l+1) - \phi(l)] \geq 0$ because w and ϕ are increasing in l.

b) $\tau(l+1,g,p) - \tau(l,g,p) + \beta\max_e\{\{E[\phi(l'(l+1)) \mid s = 1] - c(e)\} - \beta\max_e\{E[\phi(l'(l)) \mid s = 1] - c(e)\}$ is likely to be positive if $\tau(l+1,g,p) - \tau(l,g,p)$ is positive or not too large compared to the discounted marginal value of a higher education degree.

c) $w(g,l+1) - \tau(l,g,p) + \beta\phi(l+1) - \max_e\{\beta E[\phi(l'(l)) \mid s = 1] - c(e)\} \geq 0$ if $\tau(l,g,p) \leq w(g,l+1)$ (the wage is sufficiently high compared to the transfer) because $E[\phi(l') \mid s = 1] < \phi(l+1)$ which implies that $\beta\phi(l+1) - \max_e\{\beta E[\phi(l') \mid s=1] - c(e)\} > 0$.

d) $\tau(l+1,g,p) + \max_e\{\beta E[\phi(l'(l+1)) \mid s=1] - c(e)\} + \beta\phi(l) - w(g,l) > 0$ because obviously $\phi(l) > \frac{w(g,l)}{1-\beta}$ implying that $\beta\phi(l) > \frac{\beta}{1-\beta}w(g,l) \geq w(g,l)$ if $\beta \geq 1/2$ and $\max_e\{\beta E[\phi(l'(l+1)) \mid s = 1] - c(e)\} \geq 0$. ■

C Proofs

C.1 Proof of Proposition 1

Noting $\phi(.,g,p) = \phi(.)$, let's first define an operator $T_{g,p}$ transforming $\phi(.)$ in $T_{g,p}(\phi(.))$ by

$$\begin{aligned} \forall l &: \quad T_{g,p}(\phi)(l) = \max\{w(g,l) + \beta\phi(l), \max_{e\geq 0}\{\tau(l,g,p) - c(e) + \beta E[\phi(l') \mid s=1]\}\} \\ s.t.\ l' &= \quad l+1 \text{ if } s=1 \text{ and } \pi(l,e) \geq \varepsilon \text{ and } l' = l \text{ otherwise.} \end{aligned}$$

$\phi(.)$ is the fixed point (if any) of the operator $T_{g,p}(.)$. If $T_{g,p}(.)$ is a contraction mapping, then its fixed point exists and is unique (see Stokey and Lucas, 1989). Using Blackwell sufficiency theorem, we just need to show that $T_{g,p}$ verifies the monotonicity and discounting properties. Let $\phi, \widetilde{\phi} \in C(\Re_+, \Re)$ such that $\forall l$, $\phi(l) \leq \widetilde{\phi}(l)$ then it is straightforward to check that $\forall l$, $T_{g,p}\phi(l) \leq T_{g,p}\widetilde{\phi}(l)$ (monotonicity property). Moreover $\forall l$, $T_{g,p}(\phi+\gamma)(l) \leq T_{g,p}(\phi)(l) + \beta\gamma$ because it is straightforward to check that $T_{g,p}(\phi+\gamma)(l) = T_{g,p}(\phi)(l) + \beta\gamma$. ϕ is the fixed point of T: $T_{g,p}(\phi) = \phi$.

C.2 Proof of Proposition 2

Conditional on schooling, the learning effort is chosen to maximize $\tau(l,g,p) - c(e) + \beta E[\phi_{g,p}(l') \mid s = 1, e]$. This program is always concave in e because $E[\phi_{g,p}(l') \mid s=1,e] = P(l' = l+1 \mid s = 1,e)\phi(l+1,g,p) + P(l' = l \mid s=1,e)\phi(l,g,p)$
$= \phi(l,g,p) + P(l' = l+1 \mid s=1,e)[\phi(l+1,g,p) - \phi(l,g,p)]$ and $P(l' = l+1 \mid s=1,e)$ is increasing concave in e. Then, the learning effort will satisfy the first order condition $\beta\frac{\partial}{\partial e}E[\phi_{g,p}(l) \mid s=1] = c'(e)$. Since $P(l' = l+1 \mid s=1,e) = F \circ \pi(l,e^*)$ and when $\phi(l+1,g,p) > \phi(l,g,p)$ the first order condition equation determining e^* is

$$\beta[\phi(l+1,g,p) - \phi(l,g,p)]f \circ \pi(l,e^*)\frac{\partial\pi}{\partial e}(l,e^*) = c'(e^*)$$

When $\phi(l+1,g,p) - \phi(l,g,p) < 0$ then $e^* = 0$. Since f is decreasing, $\pi(l,.)$ is increasing concave in e, and $c(.)$ increasing convex, we can use the implicit function theorem to find some properties of e^*. If $\frac{\partial\pi}{\partial l} = 0$, e^* has the same directions of variation in l than $\phi(l+1,g,p) - \phi(l,g,p)$. This implies that e^* is increasing in l if $\phi(l,g,p)$ is convex in l and decreasing in l if $\phi(l,g,p)$ is concave in l. If $\frac{\partial\pi}{\partial l} \neq 0$, then e^* can be either increasing or decreasing in l, according to the properties of $\frac{\partial^2\pi}{\partial e\partial l}$ and $\phi(l+1,g,p) - \phi(l,g,p)$.

C.3 Proof of Proposition 5

It comes from the fact that the conditional distribution of $\widetilde{\theta}$ on ω_{t+1} does not depend on p. Using the law of iterated expectations, a simple differentiation of the expectation proves it[23]:

$$\begin{aligned}
\frac{\partial}{\partial p}E(s_{t+1} \quad | \quad \omega_{t+1},p) &= \frac{\partial}{\partial p}\iint s_{t+1}d\lambda(s_{t+1},\widetilde{\theta} \mid \omega_{t+1},p) \\
&= \frac{\partial}{\partial p}\iint s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})d\lambda(\widetilde{\theta} \mid \omega_{t+1},p) \\
&= \int[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\psi(\omega_{t+1},p,\widetilde{\theta})]
\end{aligned}$$

because (13) implies that $d\lambda(\widetilde{\theta} \mid \omega_{t+1},p) = d\lambda(\widetilde{\theta} \mid \omega_{t+1},1) = d\lambda(\widetilde{\theta} \mid \omega_{t+1},0) = d\lambda(\widetilde{\theta} \mid \omega_{t+1})$.

C.4 Proof of Proposition 6

The difference of biases is

$$\begin{aligned}
\frac{\partial}{\partial p}B(\omega_{t+1},p) &= \frac{\partial}{\partial p}\{\frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}}\psi(\omega_{t+1},p,\widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1},p,\widetilde{\theta})]\} \\
&= \frac{\partial}{\partial \omega_{t+1}}E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\psi(\omega_{t+1},p,\widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1},p,\widetilde{\theta})]
\end{aligned}$$

because $d\lambda(\widetilde{\theta} \mid \omega_{t+1},p) = d\lambda(\widetilde{\theta} \mid \omega_{t+1})$. However

$$\frac{\partial}{\partial \omega_{t+1}}E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\psi(\omega_{t+1},p,\widetilde{\theta})] = \frac{\partial}{\partial \omega_{t+1}}\int[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

$$= \int[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]\frac{\partial}{\partial \omega_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) + \int \frac{\partial}{\partial \omega_{t+1}}[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

$$= \int[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]\frac{\partial}{\partial \omega_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) + E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\frac{\partial}{\partial \omega_{t+1}}\psi(\omega_{t+1},p,\widetilde{\theta})]$$

Then $\frac{\partial}{\partial p}B(\omega_{t+1},p) = 0$ if and only if $\int[\frac{\partial}{\partial p}\int s_{t+1}d\lambda(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})]\frac{\partial}{\partial \omega_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$. This is not always true since in general $\frac{\partial}{\partial \omega_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \neq 0$.

Noting ω^k_{t+1} some component of ω_{t+1}, the average change according to ω^k_{t+1} in the impact of p, $E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\frac{\partial}{\partial \omega^k_{t+1}}\psi(\omega_{t+1},p,\widetilde{\theta})]$, will be identified through the estimation of $\frac{\partial}{\partial \omega^k_{t+1}}E_{\widetilde{\theta}}[\frac{\partial}{\partial p}\psi(\omega_{t+1},p,\widetilde{\theta})]$ if and only if one of the following conditions is satisfied:

$$\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$$

or

$$\int \frac{\partial}{\partial p}E(s_{t+1} \mid \omega_{t+1},p,\widetilde{\theta})\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$$

$$\Leftrightarrow \int E(s_{t+1} \mid \omega_{t+1},1,\widetilde{\theta})\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = \int E(s_{t+1} \mid \omega_{t+1},0,\widetilde{\theta})\frac{\partial}{\partial \omega^k_{t+1}}d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

Then, we just need to use Proposition 5 to complete the proof.

[23] The notation $\lambda(\mu \mid \nu)$ always means the cumulative distribution of μ conditional on ν.

C.5 Proof of Proposition 7

We just need to prove that $\frac{\partial}{\partial T}E\left(s_{t+1} \mid \omega_{t+1}, T\right) = E_{\widetilde{\theta}}[\frac{\partial}{\partial T}E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta})]$. We have $\frac{\partial}{\partial T}E\left(s_{t+1} \mid \omega_{t+1}, T\right) = \frac{\partial}{\partial T}[E_{\widetilde{\theta}}\psi(\omega_{t+1}, T, \widetilde{\theta})]$ where $E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta}) = \psi(\omega_{t+1}, T, \widetilde{\theta})$. With $T = \widetilde{\tau}(g, l, p, \omega'_{t+1})$ and $\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})$ known (by the program rule),

$$\frac{\partial}{\partial T}\psi(\omega_{t+1}, T, \widetilde{\theta}) = \frac{\frac{\partial}{\partial \omega'_{t+1}}\psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})}{\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})}.$$

Thus, we can write

$$\begin{aligned}
E_{\widetilde{\theta}}[\frac{\partial}{\partial T}\psi(\omega_{t+1}, T, \widetilde{\theta})] &= \frac{E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega'_{t+1}}\psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})]}{\frac{\partial}{\partial \omega^k_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})} \text{ because } \frac{\partial}{\partial \widetilde{\theta}}\{\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})\} = 0 \\
&= \frac{\frac{\partial}{\partial \omega'_{t+1}}E_{\widetilde{\theta}}[\psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})]}{\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})} \text{ because } \widetilde{\theta} \perp\!\!\!\perp \ \omega'_{t+1} \mid \omega_{t+1} \\
&= \frac{\frac{\partial}{\partial \omega'_{t+1}}E\left(s_{t+1} \mid \omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1})\right)}{\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})}
\end{aligned}$$

which proves the proposition since

$$\frac{\partial}{\partial T}E\left(s_{t+1} \mid \omega_{t+1}, T\right) = \frac{\frac{\partial}{\partial \omega'_{t+1}}E(s_{t+1} \mid \omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}))}{\frac{\partial}{\partial \omega'_{t+1}}\widetilde{\tau}(g, l, p, \omega'_{t+1})}.$$

D Full Tables of Results

Table E98-1: Impact on Continuation Decision in 1998

$P(s_{t+1}=1 \mid s_t=1)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
Progresa Dummy $\alpha.p$	0.687*	(7.89)	0.035*	(7.72)	0.434*	(2.43)	0.022*	(2.39)
Transfer (100 Pesos) $\alpha.\tau(l,g)$					0.182	(1.69)	0.009	(1.68)
Covariates Z_{t+1}								
Gender (1: boy, 0:girl)	2.031*	(2.20)	0.101*	(2.21)	1.968*	(2.18)	0.098*	(2.19)
Household Head Education	0.189*	(4.13)	0.009*	(4.15)	0.190*	(4.15)	0.009*	(4.17)
Household Size	-0.000	(-0.03)	-0.000	(-0.03)	0.002	(0.13)	0.000	(0.13)
Age	-0.644*	(-20.92)	-0.032*	(-22.12)	-0.645*	(-20.97)	-0.032*	(-22.15)
Distance to Sec. School	-0.101*	(-4.58)	-0.005*	(-4.60)	-0.101*	(-4.57)	-0.005*	(-4.59)
Grade×Gender Dummies								
P2, girl	-0.016	(-0.03)	-0.000	(-0.03)	-0.064	(-0.12)	-0.003	(-0.12)
P3, girl	0.554	(1.74)	0.025*	(1.98)	0.444	(1.38)	0.020	(1.53)
P3, boy	-0.799	(-0.85)	-0.047	(-0.73)	-0.842	(-0.90)	-0.050	(-0.77)
P4, girl	1.672*	(5.13)	0.058*	(8.01)	1.559*	(4.70)	0.055*	(7.10)
P4, boy	-0.667	(-0.72)	-0.038	(-0.63)	-0.717	(-0.78)	-0.041	(-0.68)
P5, girl	1.505*	(5.61)	0.056*	(7.92)	1.377*	(4.95)	0.052*	(6.78)
P5, boy	-0.137	(-0.15)	-0.007	(-0.14)	-0.202	(-0.22)	-0.010	(-0.21)
P6, girl	2.046*	(7.64)	0.069*	(12.21)	1.887*	(6.65)	0.065*	(10.31)
P6, boy	-0.008	(-0.01)	-0.000	(-0.01)	-0.104	(-0.11)	-0.005	(-0.11)
S1, girl	-0.199	(-1.01)	-0.010	(-0.98)	-0.424	(-1.78)	-0.023	(-1.66)
S1, boy	-1.763	(-1.94)	-0.120	(-1.44)	-1.919*	(-2.11)	-0.134	(-1.53)
S2, girl	2.569*	(9.22)	0.075*	(18.31)	2.336*	(7.54)	0.071*	(14.15)
S2, boy	1.119	(1.19)	0.044	(1.57)	0.965	(1.03)	0.039	(1.30)
S3, girl	3.839*	(9.68)	0.084*	(28.56)	3.615*	(8.72)	0.082*	(24.97)
S3, boy	1.287	(1.37)	0.048	(1.89)	1.126	(1.20)	0.044	(1.58)
S4, boy	-1.891*	(-2.06)	-0.144	(-1.48)	-1.829*	(-2.04)	-0.138	(-1.47)
State Dummies (reference is Veracruz)								
Guerrero	1.409*	(5.65)	0.070*	(5.72)	1.403*	(5.62)	0.070*	(5.69)
Hidalgo	0.841*	(4.05)	0.042*	(4.09)	0.843*	(4.06)	0.042*	(4.10)
Michoacán	0.490*	(2.40)	0.024*	(2.42)	0.493*	(2.42)	0.025*	(2.43)
Puebla	0.672*	(3.33)	0.033*	(3.35)	0.671*	(3.33)	0.033*	(3.35)
Queretaro	0.784*	(3.73)	0.039*	(3.75)	0.786*	(3.74)	0.039*	(3.76)
San Luis	1.117*	(5.87)	0.056*	(5.93)	1.118*	(5.87)	0.055*	(5.93)
Intercept	8.117*	(15.08)			8.272*	(15.26)		
Observations	13894				13894			

Table E98-2 : Impact on Continuation Decision in 1998 by School Level

$P\left(s_{t+1}=1 \mid s_t=1\right)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
α(grade l,gender g).p								
Primary school, girl	0.704*	(3.37)	0.031*	(3.86)	0.774	(1.13)	0.034	(1.29)
Primary school, boy	0.694*	(3.30)	0.031*	(3.78)	1.111	(1.43)	0.046	(1.72)
Secondary school, girl	0.762*	(5.08)	0.034*	(5.79)	-0.981	(-1.19)	-0.058	(-1.01)
Secondary school, boy	0.712*	(4.65)	0.032*	(5.22)	-0.487	(-0.52)	-0.026	(-0.48)
α(grade l,gender g).$\tau(l,g)$								
Primary school, girl					-0.076	(-0.11)	-0.004	(-0.11)
Primary school, boy					-0.420	(-0.58)	-0.021	(-0.58)
Secondary school, girl					0.876*	(2.14)	0.044*	(2.14)
Secondary school, boy					0.625	(1.30)	0.031	(1.30)
Covariates Z_{t+1}								
Gender (1: boy, 0:girl)	1.597	(1.71)	0.079	(1.71)	1.456	(1.39)	0.072	(1.39)
Household Head Education	0.187*	(4.10)	0.009*	(4.12)	0.188*	(4.11)	0.009*	(4.13)
Household Size	0.000	(0.05)	0.000	(0.05)	0.010	(0.58)	0.000	(0.58)
Age	-0.642*	(-20.90)	-0.032*	(-22.09)	-0.643*	(-20.98)	-0.032*	(-22.19)
Distance to Sec. School	-0.101*	(-4.56)	-0.005*	(-4.58)	-0.099*	(-4.50)	-0.005*	(-4.52)
Grade×Gender Dummies								
P2, girl	-0.452	(-0.83)	-0.025	(-0.76)	-0.492	(-0.92)	-0.027	(-0.83)
P3, girl	0.116	(0.34)	0.006	(0.35)	0.098	(0.28)	0.005	(0.29)
P3, boy	-0.800	(-0.85)	-0.047	(-0.73)	-0.729	(-0.71)	-0.042	(-0.62)
P4, girl	1.233*	(3.58)	0.047*	(4.93)	1.219*	(3.54)	0.047*	(4.85)
P4, boy	-0.670	(-0.72)	-0.038	(-0.64)	-0.579	(-0.57)	-0.032	(-0.51)
P5, girl	1.063*	(3.73)	0.043*	(4.75)	1.055*	(3.70)	0.043*	(4.69)
P5, boy	-0.141	(-0.15)	-0.007	(-0.15)	-0.011	(-0.01)	-0.000	(-0.01)
P6, girl	1.603*	(5.75)	0.058*	(8.40)	1.606*	(5.35)	0.058*	(7.72)
P6, boy	-0.012	(-0.01)	-0.000	(-0.01)	0.187	(0.17)	0.009	(0.18)
S1, girl	-0.676*	(-3.16)	-0.038*	(-2.81)	-0.672*	(-3.14)	-0.037*	(-2.80)
S1, boy	-1.779	(-1.92)	-0.121	(-1.42)	-1.633	(-1.57)	-0.108	(-1.19)
S2, girl	2.093*	(7.41)	0.067*	(13.16)	2.035*	(7.19)	0.066*	(12.58)
S2, boy	1.103	(1.15)	0.043	(1.52)	1.229	(1.15)	0.047	(1.57)
S3, girl	3.372*	(8.50)	0.080*	(23.91)	3.301*	(8.32)	0.079*	(23.45)
S3, boy	1.271	(1.33)	0.048	(1.83)	1.368	(1.28)	0.051	(1.81)
S4, boy	-1.460	(-1.57)	-0.102	(-1.19)	-1.318	(-1.26)	-0.089	(-0.98)
State Dummies (reference is Veracruz)								
Guerrero	1.411*	(5.65)	0.070*	(5.72)	1.393*	(5.59)	0.069*	(5.66)
Hidalgo	0.859*	(4.13)	0.043*	(4.17)	0.864*	(4.16)	0.043*	(4.21)
Michoacan	0.495*	(2.42)	0.025*	(2.44)	0.508*	(2.48)	0.025*	(2.49)
Puebla	0.671*	(3.33)	0.033*	(3.35)	0.670*	(3.31)	0.033*	(3.34)
Queretaro	0.794*	(3.78)	0.040*	(3.80)	0.807*	(3.84)	0.040*	(3.86)
San Luis	1.118*	(5.87)	0.056*	(5.93)	1.119*	(5.88)	0.056*	(5.94)
Intercept	8.521*	(15.89)			8.459*	(15.80)		
Observations	13894				13894			

Table E98-3 : Impact on Continuation Decision in 1998 by Grade

$P\left(s_{t+1}=1 \mid s_t=1\right)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
α(grade l,gender g).p								
P2, girl	-0.322	(-0.35)	-0.017	(-0.33)	-0.319	(-0.35)	-0.017	(-0.33)
P3, girl	1.104*	(2.63)	0.043*	(3.54)	-1.773	(-0.40)	-0.129	(-0.29)
P3, boy	0.536	(1.09)	0.024	(1.25)	0.542	(0.08)	0.024	(0.09)
P4, girl	1.203*	(2.36)	0.045*	(3.35)	-3.860	(-1.01)	-0.410	(-0.66)
P4, boy	0.901*	(2.24)	0.037*	(2.84)	1.668	(0.36)	0.057	(0.56)
P5, girl	0.630	(1.77)	0.027*	(2.06)	-2.767	(-1.20)	-0.241	(-0.80)
P5, boy	0.837*	(2.15)	0.034*	(2.66)	3.699	(0.84)	0.088	(1.64)
P6, girl	0.390	(1.00)	0.018	(1.09)	5.158	(0.92)	0.105*	(2.31)
P6, boy	0.459	(1.32)	0.021	(1.47)	-2.903	(-0.89)	-0.258	(-0.58)
S1, girl	0.857*	(5.25)	0.037*	(6.18)	-1.004	(-1.14)	-0.060	(-0.95)
S1, boy	0.823*	(4.94)	0.035*	(5.77)	-1.159	(-1.16)	-0.071	(-0.94)
S2, girl	0.133	(0.29)	0.006	(0.29)	-3.686*	(-2.28)	-0.383	(-1.46)
S2, boy	0.629	(1.25)	0.027	(1.47)	-1.110	(-0.49)	-0.070	(-0.40)
S3, girl	0.280	(0.39)	0.013	(0.42)	2.792	(0.88)	0.072*	(2.25)
S3, boy	-0.174	(-0.33)	-0.009	(-0.31)	4.307	(1.02)	0.088*	(3.62)
α(grade l,gender g).$\tau(l,g)$								
P3, girl					4.236	(0.64)	0.210	(0.64)
P4, girl					6.550	(1.30)	0.325	(1.30)
P4, boy					-0.970	(-0.16)	-0.048	(-0.16)
P5, girl					3.512	(1.48)	0.174	(1.48)
P5, boy					-2.905	(-0.66)	-0.144	(-0.66)
P6, girl					-3.619	(-0.86)	-0.179	(-0.86)
P6, boy					2.577	(1.04)	0.128	(1.03)
S1, girl					0.947*	(2.13)	0.047*	(2.13)
S1, boy					1.048*	(2.01)	0.052*	(2.01)
S2, girl					1.840*	(2.36)	0.091*	(2.36)
S2, boy					0.894	(0.77)	0.044	(0.77)
S3, girl					-1.081	(-0.84)	-0.054	(-0.84)
S3, boy					-2.119	(-1.08)	-0.105	(-1.08)
Covariates Z_{t+1}								
Gender (1: boy, 0:girl)	0.137	(0.63)	0.007	(0.63)	-1.382	(-1.21)	-0.069	(-1.20)
Household Head Education	0.188*	(4.12)	0.009*	(4.14)	0.188*	(4.12)	0.009*	(4.14)
Household Size	0.000	(0.03)	0.000	(0.03)	0.015	(0.83)	0.000	(0.83)
Age	-0.643*	(-20.78)	-0.032*	(-22.06)	-0.650*	(-21.03)	-0.032*	(-22.11)
Distance to Sec. School	-0.101*	(-4.59)	-0.005*	(-4.60)	-0.098*	(-4.46)	-0.005*	(-4.48)
Grade×Gender Dummies								
P2, girl	0.109	(0.14)	0.005	(0.15)	-2.377*	(-2.89)	-0.196*	(-1.97)
P2, boy	0.948	(0.89)	0.038	(1.16)				
P3, girl	-0.045	(-0.12)	-0.002	(-0.12)	-2.525*	(-5.30)	-0.207*	(-3.65)
P3, boy	0.732	(1.91)	0.031*	(2.28)	-0.229	(-0.20)	-0.012	(-0.19)
P4, girl	1.026*	(2.78)	0.041*	(3.59)	-1.447*	(-2.96)	-0.097*	(-2.32)
P4, boy	0.692*	(2.24)	0.030*	(2.62)	-0.262	(-0.23)	-0.014	(-0.22)
P5, girl	1.098*	(3.40)	0.044*	(4.33)	-1.372*	(-3.01)	-0.089*	(-2.40)
P5, boy	1.248*	(3.90)	0.048*	(5.22)	0.300	(0.27)	0.014	(0.29)
P6, girl	1.751*	(5.14)	0.062*	(7.57)	-0.717	(-1.59)	-0.041	(-1.40)
P6, boy	1.558*	(5.18)	0.058*	(7.17)	0.618	(0.55)	0.027	(0.63)
S1, girl	-0.731*	(-3.37)	-0.041*	(-2.97)	-3.194*	(-8.16)	-0.293*	(-5.10)
S1, boy	-0.386	(-1.89)	-0.020	(-1.77)	-1.321	(-1.21)	-0.082	(-0.97)
S2, girl	2.454*	(6.14)	0.073*	(11.57)				
S2, boy	2.607*	(6.49)	0.075*	(12.49)	1.675	(1.46)	0.058*	(2.22)
S3, girl	3.591*	(6.77)	0.082*	(17.54)	1.145	(1.83)	0.044*	(2.50)
S3, boy	3.255*	(7.14)	0.085*	(15.62)	2.337*	(2.00)	0.071*	(3.53)
S4, girl					-2.441*	(-6.11)	-0.208*	(-4.10)
S4, boy					-0.920	(-0.84)	-0.057	(-0.70)
State Dummies (reference is Veracruz)								
Guerrero	1.423*	(5.70)	0.071*	(5.77)	1.398*	(5.57)	0.069*	(5.65)
Hidalgo	0.866*	(4.15)	0.043*	(4.19)	0.872*	(4.18)	0.043*	(4.22)
Michoacan	0.506*	(2.47)	0.025*	(2.49)	0.516*	(2.51)	0.026*	(2.53)
Puebla	0.672*	(3.32)	0.033*	(3.34)	0.666*	(3.28)	0.033*	(3.30)
Queretaro	0.800*	(3.79)	0.040*	(3.81)	0.815*	(3.86)	0.040*	(3.88)
San Luis	1.127*	(5.90)	0.056*	(5.96)	1.126*	(5.88)	0.056*	(5.93)
Intercept	8.532*	(15.82)			10.954*	(17.88)		
Observations	13894				13894			

Table P98-1 : Impact on Performance in 1998

$P\left(l_{t+1}=l_t+1 \mid s_t=1\right)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
Progresa Dummy $\theta.p$	0.163*	(2.67)	0.020*	(2.63)	1.376*	(12.52)	0.182*	(11.71)
Transfer (100 Pesos) $\theta.\tau(l,g)$					-1.542*	(-10.80)	-0.179*	(-11.27)
Covariates Z_{t+1}								
Gender (1: boy, 0:girl)	0.074	(0.31)	0.009	(0.31)	-0.072	(-0.23)	-0.008	(-0.23)
Household Head Education	-0.007	(-0.23)	-0.000	(-0.23)	-0.014	(-0.45)	-0.002	(-0.45)
Household Size	-0.033*	(-2.69)	-0.004*	(-2.67)	-0.049*	(-3.87)	-0.006*	(-3.84)
Age	-0.317*	(-17.25)	-0.038*	(-17.67)	-0.304*	(-16.63)	-0.035*	(-16.78)
Distance to Sec. School	0.020	(1.36)	0.002	(1.36)	0.019	(1.28)	0.002	(1.28)
Grade×Gender Dummies								
P3, girl	-2.206*	(-9.08)	-0.401*	(-8.22)	-3.866*	(-9.48)	-0.637*	(-16.51)
P3, boy	-2.146*	(-10.86)	-0.385*	(-9.91)	-3.673*	(-13.47)	-0.609*	(-22.08)
P4, girl	-1.810*	(-7.79)	-0.319*	(-6.57)	-3.152*	(-8.25)	-0.547*	(-10.10)
P4, boy	-1.821*	(-9.95)	-0.321*	(-8.50)	-3.027*	(-12.48)	-0.527*	(-15.02)
P5, girl	-1.161*	(-5.10)	-0.186*	(-4.19)	-2.426*	(-6.48)	-0.423*	(-6.26)
P5, boy	-1.242*	(-6.80)	-0.201*	(-5.61)	-2.356*	(-9.95)	-0.410*	(-9.63)
P6, girl	-0.931*	(-4.14)	-0.142*	(-3.45)	-2.026*	(-5.60)	-0.347*	(-4.94)
P6, boy	-0.801*	(-4.62)	-0.119*	(-3.93)	-1.742*	(-7.97)	-0.289*	(-6.91)
S1, girl	-0.755*	(-3.49)	-0.111*	(-2.97)	-1.497*	(-4.56)	-0.241*	(-3.80)
S1, boy	-0.591*	(-3.44)	-0.083*	(-3.01)	-1.214*	(-6.18)	-0.187*	(-5.18)
S2, girl	-0.137	(-0.56)	-0.017	(-0.54)	-0.287	(-0.91)	-0.036	(-0.84)
S2, boy	-0.119	(-0.61)	-0.015	(-0.59)	-0.213	(-1.03)	-0.026	(-0.98)
S3, girl	-0.222	(-0.90)	-0.029	(-0.84)	-0.229	(-0.73)	-0.029	(-0.69)
S4, boy	0.091	(0.38)	0.011	(0.39)	-0.039	(-0.14)	-0.005	(-0.14)
State Dummies (reference is Veracruz)								
Guerrero	-0.387*	(-2.16)	-0.047*	(-2.16)	-0.362*	(-2.04)	-0.042*	(-2.04)
Hidalgo	-0.286	(-1.66)	-0.034	(-1.66)	-0.271	(-1.58)	-0.032	(-1.57)
Michoacan	-0.313	(-1.80)	-0.038	(-1.79)	-0.316	(-1.81)	-0.037	(-1.81)
Puebla	-0.314	(-1.86)	-0.038	(-1.86)	-0.298	(-1.76)	-0.035	(-1.76)
Queretaro	0.020	(0.12)	0.002	(0.12)	0.032	(0.18)	0.004	(0.18)
San Luis	-0.401*	(-2.49)	-0.048*	(-2.49)	-0.372*	(-2.31)	-0.043*	(-2.31)
Intercept	6.943*	(17.67)			8.101*	(16.59)		
Observations	13911				13911			

Table P98-2 : Impact on Performance in 1998 by School Level

$P(l_{t+1} = l_t + 1 \mid s_t = 1)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
θ(grade l,gender g).p								
Primary school, girl	0.552*	(6.76)	0.061*	(7.36)	2.165*	(16.80)	0.194*	(21.26)
Primary school, boy	0.541*	(6.63)	0.060*	(7.13)	2.395*	(18.02)	0.216*	(23.09)
Secondary school, girl	-1.321*	(-5.80)	-0.219*	(-4.73)	-2.753*	(-4.21)	-0.489*	(-4.09)
Secondary school, boy	-1.096*	(-4.56)	-0.172*	(-3.76)	-3.017*	(-4.40)	-0.528*	(-4.65)
θ(grade l,gender g).$\tau(l, g)$								
Primary school, girl					-2.470*	(-12.45)	-0.287*	(-12.41)
Primary school, boy					-2.897*	(-13.89)	-0.336*	(-13.86)
Secondary school, girl					0.608	(1.71)	0.071	(1.71)
Secondary school, boy					0.925*	(2.37)	0.107*	(2.38)
Covariates Z_{t+1}								
Gender (1: boy, 0:girl)	-3.158*	(-8.86)	-0.380*	(-9.03)	-3.856*	(-9.25)	-0.448*	(-9.48)
Household Head Education	-0.009	(-0.30)	-0.001	(-0.30)	-0.019	(-0.62)	-0.002	(-0.62)
Household Size	-0.030*	(-2.46)	-0.004*	(-2.45)	-0.038*	(-2.97)	-0.004*	(-2.95)
Age	-0.308*	(-16.73)	-0.037*	(-16.93)	-0.300*	(-16.10)	-0.035*	(-16.39)
Distance to Sec. School	0.022	(1.51)	0.003	(1.51)	0.021	(1.44)	0.002	(1.44)
Grade$\times$Gender Dummies								
P3, girl	-3.294*	(-9.30)	-0.588*	(-12.11)	-3.931*	(-9.30)	-0.645*	(-15.28)
P4, girl	-2.919*	(-8.41)	-0.531*	(-9.34)	-3.117*	(-7.69)	-0.534*	(-8.66)
P4, boy	0.305*	(2.78)	0.034*	(3.01)	0.801*	(6.61)	0.077*	(8.16)
P5, girl	-2.280*	(-6.57)	-0.414*	(-6.03)	-2.345*	(-5.84)	-0.399*	(-5.37)
P5, boy	0.879*	(7.19)	0.084*	(9.27)	1.555*	(11.42)	0.124*	(17.14)
P6, girl	-2.041*	(-5.93)	-0.366*	(-5.16)	-1.839*	(-4.65)	-0.302*	(-3.96)
P6, boy	1.315*	(9.72)	0.112*	(14.28)	2.346*	(14.30)	0.157*	(25.78)
S1, girl	-1.599*	(-5.14)	-0.274*	(-4.23)	-1.109*	(-3.03)	-0.165*	(-2.54)
S1, boy	1.731*	(12.54)	0.131*	(20.15)	3.118*	(14.06)	0.173*	(32.84)
S2, girl	0.030	(0.11)	0.004	(0.11)	0.232	(0.72)	0.025	(0.77)
S2, boy	3.111*	(10.45)	0.160*	(30.76)	4.021*	(10.77)	0.169*	(39.88)
S3, girl	-0.147	(-0.53)	-0.019	(-0.51)	-0.078	(-0.26)	-0.009	(-0.26)
S3, boy	3.199*	(11.10)	0.159*	(33.45)	4.006*	(11.17)	0.166*	(41.03)
S4, boy	3.179*	(9.12)	0.151*	(31.47)	3.932*	(9.37)	0.156*	(38.41)
State Dummies (reference is Veracruz)								
Guerrero	-0.419*	(-2.35)	-0.050*	(-2.35)	-0.416*	(-2.32)	-0.048*	(-2.32)
Hidalgo	-0.251	(-1.46)	-0.030	(-1.46)	-0.277	(-1.59)	-0.032	(-1.59)
Michoacan	-0.285	(-1.64)	-0.034	(-1.64)	-0.307	(-1.73)	-0.036	(-1.73)
Puebla	-0.290	(-1.72)	-0.035	(-1.72)	-0.287	(-1.67)	-0.033	(-1.67)
Queretaro	0.046	(0.26)	0.006	(0.26)	0.038	(0.21)	0.004	(0.21)
San Luis	-0.365*	(-2.27)	-0.044*	(-2.27)	-0.369*	(-2.26)	-0.043*	(-2.26)
Intercept	7.684*	(16.52)			7.816*	(15.13)		
Observations	13911				13911			

Table P98-3 : Impact on Performance in 1998 by Grade

$P\left(l_{t+1}=l_t+1 \mid s_t=1\right)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
θ(grade l,gender g).p								
P3, girl	0.418*	(2.99)	0.052*	(3.31)	0.071	(0.08)	0.009	(0.09)
P3, boy	0.336*	(2.46)	0.043*	(2.66)	-0.498	(-0.64)	-0.074	(-0.58)
P4, girl	-0.320	(-1.77)	-0.048	(-1.66)	0.020	(0.02)	0.003	(0.02)
P4, boy	-0.234	(-1.28)	-0.034	(-1.22)	-1.342	(-1.50)	-0.228	(-1.29)
P5, girl	-0.056	(-0.25)	-0.008	(-0.25)	-1.649	(-1.48)	-0.290	(-1.29)
P5, boy	-0.180	(-0.83)	-0.026	(-0.80)	-1.527	(-1.74)	-0.265	(-1.51)
P6, girl	-0.177	(-0.69)	-0.025	(-0.67)	-2.357*	(-2.82)	-0.431*	(-2.75)
P6, boy	-0.312	(-1.23)	-0.046	(-1.15)	0.036	(0.03)	0.005	(0.04)
S1, girl	-1.989*	(-5.50)	-0.378*	(-5.08)	-2.467*	(-3.46)	-0.460*	(-3.45)
S1, boy	-1.617*	(-4.41)	-0.297*	(-3.88)	-3.226*	(-4.25)	-0.582*	(-5.52)
S1, girl	-0.333	(-0.66)	-0.050	(-0.62)	-2.682	(-1.71)	-0.501	(-1.81)
S2, boy	-0.296	(-0.62)	-0.044	(-0.58)	-3.282*	(-2.38)	-0.591*	(-3.16)
S3, girl	-1.091*	(-2.21)	-0.190	(-1.89)	-1.258	(-0.94)	-0.216	(-0.79)
S3, boy	-1.470*	(-2.72)	-0.269*	(-2.34)	-1.513	(-1.26)	-0.268	(-1.07)
θ(grade l,gender g).$\tau(l,g)$								
P3, girl					0.602	(0.42)	0.080	(0.42)
P3, boy					1.451	(1.09)	0.193	(1.09)
P4, girl					-0.517	(-0.37)	-0.069	(-0.37)
P4, boy					1.646	(1.23)	0.219	(1.23)
P5, girl					1.859	(1.43)	0.247	(1.43)
P5, boy					1.583	(1.55)	0.210	(1.55)
P6, girl					1.983*	(2.63)	0.263*	(2.63)
P6, boy					-0.316	(-0.36)	-0.042	(-0.36)
S1, girl					0.284	(0.74)	0.038	(0.74)
S1, boy					1.002*	(2.22)	0.133*	(2.22)
S2, girl					1.290	(1.50)	0.171	(1.51)
S2, boy					1.778*	(2.15)	0.236*	(2.15)
S3, girl					0.085	(0.14)	0.011	(0.14)
S3, boy					0.033	(0.05)	0.004	(0.05)
Gender (1: boy, 0:girl)	-3.718*	(-7.13)	-0.514*	(-7.19)	0.482	(0.74)	0.064	(0.74)
Household Head Education	-0.021	(-0.66)	-0.003	(-0.66)	-0.022	(-0.71)	-0.003	(-0.71)
Household Size	-0.031*	(-2.49)	-0.004*	(-2.49)	-0.018	(-1.36)	-0.002	(-1.36)
Age	-0.308*	(-15.99)	-0.043*	(-16.52)	-0.311*	(-16.01)	-0.041*	(-16.49)
Distance to Sec. School	0.017	(1.14)	0.002	(1.14)	0.018	(1.19)	0.002	(1.19)
Grade$\times$Gender Dummies								
P3, girl	-3.843*	(-7.35)	-0.659*	(-13.53)	-3.864*	(-7.45)	-0.650*	(-13.08)
P3, boy					-4.219*	(-9.61)	-0.676*	(-21.23)
P4, girl	-2.575*	(-4.93)	-0.469*	(-5.33)	-2.592*	(-5.01)	-0.458*	(-5.32)
P4, boy	1.210*	(8.97)	0.128*	(11.91)	-3.008*	(-6.82)	-0.524*	(-8.57)
P5, girl	-1.597*	(-2.98)	-0.284*	(-2.65)	-1.606*	(-3.02)	-0.276*	(-2.67)
P5, boy	2.072*	(11.21)	0.179*	(19.08)	-2.143*	(-4.78)	-0.379*	(-4.63)
P6, girl	-1.519*	(-2.84)	-0.269*	(-2.51)	-1.522*	(-2.87)	-0.260*	(-2.52)
P6, boy	2.484*	(11.99)	0.196*	(22.47)	-1.726*	(-3.90)	-0.299*	(-3.53)
S1, girl	-0.987	(-1.80)	-0.165	(-1.57)	-1.013	(-1.87)	-0.164	(-1.63)
S1, boy	2.962*	(12.14)	0.206*	(27.05)	-1.236*	(-3.52)	-0.206*	(-3.08)
S2, girl	0.696	(1.00)	0.080	(1.23)	0.690	(1.01)	0.076	(1.25)
S2, boy	4.207*	(9.51)	0.209*	(38.29)				
S3, girl	-0.579	(-1.21)	-0.091	(-1.08)	-0.603	(-1.27)	-0.092	(-1.14)
S3, boy	3.494*	(8.92)	0.199*	(29.70)	-0.739	(-1.28)	-0.115	(-1.13)
S4, boy	4.183*	(6.94)	0.195*	(35.52)	-0.031	(-0.04)	-0.004	(-0.04)
State Dummies (reference is Veracruz)								
Guerrero	-0.430*	(-2.40)	-0.060*	(-2.40)	-0.456*	(-2.53)	-0.061*	(-2.52)
Hidalgo	-0.290	(-1.67)	-0.040	(-1.67)	-0.279	(-1.60)	-0.037	(-1.60)
Michoacan	-0.302	(-1.71)	-0.042	(-1.71)	-0.287	(-1.63)	-0.038	(-1.63)
Puebla	-0.289	(-1.69)	-0.040	(-1.69)	-0.291	(-1.70)	-0.039	(-1.70)
Queretaro	0.025	(0.14)	0.003	(0.14)	0.033	(0.18)	0.004	(0.18)
San Luis	-0.385*	(-2.37)	-0.053*	(-2.36)	-0.393*	(-2.40)	-0.052*	(-2.40)
Intercept	7.562*	(12.45)			7.503*	(12.44)		
Observations	13911				13911			

Table 7 : Semi-structural estimation

Semi-structural Estimation of $P\left(s_{t+1}=1 \mid s_t=1\right)$ (t-stat), (*: 5% significance)	Coeff.	(t-stat)	$\overline{(\partial\varphi/\partial x)}$	(t-stat)
α^1(grade l,gender g)$.p$				
P3, girl	-0.019	(-0.02)	-0.001	(-0.02)
P3, boy	-0.258	(-0.31)	-0.017	(-0.28)
P4, girl	1.290	(1.59)	0.049*	(2.20)
P4, boy	1.968*	(2.80)	0.068*	(3.93)
P5, girl	1.273	(1.38)	0.052	(1.74)
P5, boy	1.675*	(2.42)	0.060*	(3.33)
P6, girl	2.225*	(2.67)	0.078*	(3.78)
P6, boy	1.130	(1.86)	0.047*	(2.35)
S1, girl	4.892*	(4.09)	0.087*	(50.32)
S1, $l_{t+1}=9$, boy	3.737*	(4.43)	0.088*	(24.04)
S2, girl	4.982*	(3.90)	0.082*	(50.73)
S2, boy	4.485*	(4.68)	0.086*	(22.70)
S3, girl	6.864*	(3.67)	0.073*	(48.02)
S3, boy	4.693*	(3.44)	0.080*	(50.07)
$\alpha^2(l_{t+1},g).\widehat{\varphi}_{t+2}.p$				
P3, girl	-2.064	(-1.27)	-0.119	(-1.27)
P3, boy	-6.353*	(-2.88)	-0.367*	(-2.89)
P4, girl	0.068	(0.06)	0.004	(0.06)
P4, boy	-0.229	(-0.26)	-0.013	(-0.26)
P5, girl	-0.813	(-0.79)	-0.047	(-0.79)
P5, boy	-1.360	(-1.76)	-0.078	(-1.76)
P6, girl	-1.653	(-1.89)	-0.095	(-1.89)
P6, boy	-1.373*	(-2.04)	-0.079*	(-2.04)
S1, girl	-2.621*	(-2.68)	-0.151*	(-2.68)
S1, $l_{t+1}=9$, boy	-0.801	(-1.10)	-0.046	(-1.10)
S2, girl	-4.032*	(-3.41)	-0.233*	(-3.41)
S2, boy	-3.268*	(-3.91)	-0.189*	(-3.90)
S3, girl	-5.089*	(-3.71)	-0.294*	(-3.71)
S3, boy	-3.942*	(-3.71)	-0.228*	(-3.71)
S4, girl	-6.282*	(-3.47)	-0.363*	(-3.46)
S4, boy	-4.966*	(-3.47)	-0.287*	(-3.46)
$\alpha^3(l_{t+1},g).\frac{\partial\widehat{\varphi}_{t+2}}{\partial p}.p$				
P3, girl	47.253	(0.61)	2.728	(0.61)
P3, boy	135.565*	(1.99)	7.827*	(1.99)
P4, girl	106.962	(1.79)	6.175	(1.79)
P4, boy	100.445*	(2.43)	5.799*	(2.43)
P5, girl	0.725	(0.05)	0.042	(0.05)
P5, boy	8.417	(0.28)	0.486	(0.28)
P6, girl	-32.660	(-1.44)	-1.886	(-1.44)
S1, girl	25.225*	(2.21)	1.456*	(2.21)
S1, $l_{t+1}=9$, boy	10.690	(1.51)	0.617	(1.51)
S2, girl	4.364	(0.50)	0.252	(0.50)
S2, boy	0.234	(0.05)	0.013	(0.05)
S3, girl	4.418	(0.39)	0.255	(0.39)
S3, boy	-1.384	(-0.14)	-0.080	(-0.14)
S4, girl	-0.587	(-0.10)	-0.034	(-0.10)
S4, boy	-7.907	(-0.92)	-0.456	(-0.92)

Table 7 *(continued)*

(t-stat), (*: 5% significance)	Coeff.	(t-stat)	$\overline{(\partial\varphi/\partial x)}$	(t-stat)
$\alpha^4(l_{t+1}, g).\widehat{\varphi}_{t+2}$				
P3, girl	3.761*	(3.46)	0.217*	(3.45)
P3, boy	5.775*	(2.92)	0.333*	(2.93)
P4, girl	2.900*	(4.75)	0.167*	(4.76)
P4, boy	2.668*	(4.21)	0.154*	(4.23)
P5, girl	2.755*	(4.74)	0.159*	(4.76)
P5, boy	2.793*	(5.23)	0.161*	(5.23)
P6, girl	3.210*	(5.15)	0.185*	(5.16)
P6, boy	3.801*	(6.86)	0.219*	(6.80)
S1, girl	5.065*	(6.32)	0.292*	(6.35)
S1, $l_{t+1} = 9$, boy	4.496*	(7.39)	0.260*	(7.32)
S2, girl	6.616*	(8.61)	0.382*	(8.70)
S2, boy	6.123*	(10.32)	0.354*	(10.23)
S3, girl	8.771*	(9.22)	0.506*	(9.26)
S3, boy	8.576*	(11.09)	0.495*	(10.88)
S4, girl	9.833*	(9.38)	0.568*	(9.37)
S4, boy	9.857*	(10.58)	0.569*	(10.38)
Covariates Z_{t+1}:				
Gender (1: boy, 0:girl)	-0.506	(-0.73)	-0.029	(-0.73)
Household Head Education	0.033	(1.16)	0.002	(1.16)
Age	-0.107*	(-2.75)	-0.006*	(-2.75)
Household Size	-0.004	(-0.41)	-0.000	(-0.41)
Distance to Sec. School	-0.010	(-0.72)	-0.000	(-0.72)
Grade×Gender Dummies				
P3, girl	-0.796	(-1.67)	-0.068	(-1.17)
P4, boy	-0.378	(-1.32)	-0.026	(-1.12)
P4, girl	-0.788	(-1.34)	-0.067	(-0.94)
P5, boy	-1.074*	(-2.84)	-0.105	(-1.81)
P5, girl	-1.587*	(-2.38)	-0.196	(-1.36)
P6, boy	-2.061*	(-4.42)	-0.312*	(-2.49)
P6, girl	-2.752*	(-3.63)	-0.512*	(-2.40)
S1, $l_{t+1} = 9$, boy	-4.226*	(-8.33)	-0.838*	(-15.88)
S1, girl	-5.347*	(-6.49)	-0.900*	(-68.59)
S2, girl	-6.242*	(-6.84)	-0.907*	(-411.13)
S2, boy	-5.633*	(-8.73)	-0.896*	(-160.94)
S3, girl	-7.767*	(-7.19)	-0.918*	(-600.54)
S3, boy	-7.336*	(-8.46)	-0.910*	(-594.18)
S4, girl	-10.487*	(-9.21)	-0.939*	(-612.99)
S4, boy	-10.212*	(-10.89)	-0.940*	(-613.69)
State Dummies (reference is Veracruz)				
Guerrero	0.872*	(5.46)	0.050*	(5.50)
Hidalgo	0.672*	(4.84)	0.039*	(4.91)
Michoacan	0.101	(0.81)	0.006	(0.82)
Puebla	0.418*	(3.36)	0.024*	(3.37)
Queretaro	0.384*	(3.05)	0.022*	(3.05)
San Luis	0.414*	(3.50)	0.024*	(3.51)
Intercept	1.343	(1.57)		
Observations		12 546		

References

Attanasio O., C. Meghir and A. Santiago (2005) “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresa” *mimeo*, UCL.

Behrman J. and Todd P. E. (1999) “Randomness in the experimental samples of Progresa” *Report to Progresa, IFPRI*

Behrman J. Sengupta P. and Todd P. E. (2005) “Progressing through Progresa: An Impact Assessment of a School Subsidy Experiment in Mexico” *Economic Development and Cultural Change* 54:1, 237-275

Bommier A. and Lambert S. (2000) “Education Demand and Age at School Enrollment in Tanzania”, *Journal of Human Resources,* 35, 1, 177-203

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

Cameron S. and J. Heckman (2001) “The Dynamics of Educational Attainment for Black, Hispanic, and White Males”, *Journal of Political Economy*, 109, 3, 455-499

Cameron S. and Taber C. (2004) “Borrowing Constraints and the Returns to Schooling” *Journal of Political Economy*, Vol 112

De Vreyer P., Lambert S. and Magnac T. (1999) “Educating Children: A Look at Household Behavior in Côte d’Ivoire”, *Working Paper INRA LEA n°9905*

Eckstein Z. and K. Wolpin (1999) “Why Youths Drop out of High School: The Impact of Preferences, Opportunities and Abilities”, *Econometrica,* 65, 1295-1340

Ham J. and Lalonde R. (1996) “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training”, *Econometrica*, 64, 1, 175-206

Hanushek Eric A. (1995) “Interpreting Recent Research on Schooling in Developing Countries”, *World Bank Research Observer*, 10, 2, 227-246

Heckman J. (1997) “Instrumental Variables: a study of implicit behavioral assumptions used in making program evaluations” *Journal of Human Resources*, 32, 441-462

Heckman J. (2001) “Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture”, *Journal of Political Economy*, 109, 4, 673-748

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998) “Characterizing Selection Bias Using Experimental Data”, *Econometrica*, 66, 1017-1098.

Heckman, J., Lalonde, R., and Smith, J. (1999) “The Economics and Econometrics of Active Labor Market Programs”, *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and D. Card, Eds., Amsterdam: Elsevier Science.

Heckman, J. and Vytlacil E. (1999) “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” *Proceedings of the National Academy of Sciences*, 96:4730-4734

Heckman, J. and Vytlacil E. (2000a) "Local Instrumental Variables" NBER Working paper T0252

Heckman, J. and Vytlacil E. (2000b) "The Relationship Between Treatment Parameters within a Latent Variable Framework", *Economics Letters*, 66(1): 33-39

Keane M. P. and Wolpin K. I. (1997) "Career Decisions of Young Men", *Journal of Political Economy*, 105, 473-522

Keane M. P. and Wolpin K. I. (2001) "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", *International Economic Review*, 42, 4, 1051-1102

Magnac T. and D. Thesmar (2002a) "Analyse économique des politiques éducatives: l'augmentation de la scolarisation en France de 1982 à 1993", *Annales d'Economie et de Statistiques*, 66:1-35

Magnac T. and D. Thesmar (2002b) "Identifying Dynamic Discrete Decision Processes", *Econometrica*, 70, 2, 801-816

Manski C. (1985) "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27, 313-334

Manski C. (1988) "Identification of Binary Response Models", *Journal of the American Statistical Association*, 83, 729-738

Newman J., Rawlings L., Gertler P. (1994) "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries", *The World Bank Research Observer*, 9, 181-202

Progresa (1999) "Evaluacion de Resultados del Programa de Educacion, Salud y Alimentacion - Primeros Avances" Report Edited by Progresa, Mexico City

Progresa (2000) "Reglas Generales para la Operacion del Programa de Educacion, Salud y Alimentacion".

Schultz T. P. (2000) "The Impact of Progresa on School Enrollments" *IFPRI Final Report*

Schultz T. P. (2004) "School subsidies for the poor: evaluating the Mexican Progresa poverty program, Pages 199-250 " *Journal of Development Economics,* Vol. 74, 1 , 199-250

Skoufias E., Davis B. and Behrman J. R. (1999) "An evaluation of the selection of beneficiary households in the Education, Health and Nutrition Program (Progresa) of Mexico" *Report to Progresa, IFPRI*

Stokey N. and Lucas R. (1989) *Recursive Methods in Economic Dynamics*, Harvard University

Taber C. (2001) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *Review of Economic Studies*, 68, 665-691.