Preliminary and Incomplete

Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility:

Assessing the Impact of a School Subsidy Program in Mexico

Petra Todd

and

Kenneth I. Wolpin

April, 2002

I. Introduction:

This paper has two goals. The first is to assess the validity of a dynamic behavioral model of parental decision-making about child schooling and fertility by exploiting data from a controlled social experiment. The experiment is designed to augment completed schooling levels of children in rural Mexico by providing subsidies to parents conditional on school attendance. The validity of the model will be assessed according to how well structural estimates of the model, based on data from the randomized-out control group and from the treatment group prior to the intervention, predict the experimental impact of the program. The second goal is to use the structural estimates of the behavioral model to perform an evaluation of policy interventions that are not part of the original experimental design, such as variations in the subsidy schedule, and to assess the longer-term impact of the program on behaviors related to child schooling decisions that extend beyond the life of the program, such as completed family size.

It is well known that the structural estimation of dynamic behavioral models requires auxiliary assumptions about the functional forms of structural relationships, i.e., preferences, technology and other constraints, and the distributions of unobservable random elements. Assessing the validity of such models by relying on tests of model fit to sample elements of the data used in estimation provides useful, but far from compelling, evidence on the validity of the model. Such models are often subjected to a form of "pre-test" estimation in that the final formulation of the model is based on the fit of prior formulations to certain aspects of the data. This practice reduces the value of within-sample fit tests as a method of validating the model.

To mitigate the effect of pre-test estimation, there have been a number of attempts to assess model validity through out-of-sample forecasts. However, such applications are sparse and have been limited by the nature of the data. Keane and Wolpin (1997), for example, used the estimates of their model of occupational choice, based on a cohort of young men between the ages of 16 and 26 (over the years 1978 to1988), to forecast occupational choices for other nearby cohorts between the ages of 27 and 44 (over the years 1989-

1995).[1] Although clearly informative, because the data are highly age-trended and the model builds in such trends, tests based on this kind of out-of-sample data may not be able to discriminate finely among alternative models.[2]

Another type of test makes use of regime shifts. For example, Lumsdaine, Stock and Wise (1992) compared the ability of structurally and non-structurally estimated models to forecast the impact of a pension "window plan" on the departure rates of workers from a single Fortune 500 firm. The workers were subject to a defined benefit plan which provided a significant incentive to remain with the firm until age 55, but to leave before 65. However, in 1982, vested workers over the age of 55 were offered a bonus of 3 to 12 months salary to retire. Forecasts of the models' predictions about the impact of the bonus on retirements, based on pre-1982 data, were compared to actual retirements. The forecast is of a large change in the pension rules, and thus provides an arguably more convincing test of the validity of the model than do within-sample tests. However, the estimates do make use of within-sample pension benefit variation as a source of model identification, the same kind of variation represented by the window plan.[3]

In this paper, we similarly use out-of-sample forecasts to assess the validity of a structurally estimated model, but the comparison is to a completely new program rather than a change in an existing program. We study the Mexican school subsidy program PROGRESA which provides monetary transfers to parents for their

[1] The model is estimated on data from the 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience and the forecasts are compared to data from the March Current Population Surveys.

[2] Keane and Wolpin found that the structurally estimated model forecasts white-collar employment better than a non-structurally estimated model, but that the situation is reversed in forecasting blue-collar employment. However, they also found that a model in which individuals behave myopically provided incredible out-of-sample forecasts of occupational choices and wages.

[3] They found that the structural dynamic programming model forecasts the impact of the window plan better than a "reduced form" probit specification.

children's school attendance. PROGRESA was implemented as a social experiment beginning in 1997.[4] We obtain structural estimates of a model of household fertility and child schooling decisions using data on the randomly selected control group and on the treatment group prior to the experiment, for whom there are longitudinal data over three survey years. We assess the performance of the model by comparing the impact of the program predicted by the model to the experimental impact on the treatment group. By design, the control and treatment groups are identical (up to sampling error) so that the behavioral model relevant to the control group should be the same as the model relevant to the treatment group. This experiment therefore represents a unique opportunity to assess the validity of a structurally estimated model that is fit to observational data.[5]

The program evaluation literature also makes use of experimental data, where the goal is to evaluate the performance of estimators of program effects that are based on observational data of program participants and non-participants. Lalonde (1986) compared estimates of the impact of a job training program based on a variety of non-experimental estimators to an experimental benchmark. He concluded that non-experimental estimators are unreliable because they deviate substantially from experimental effects and because there is no criterion for choosing among competing estimators. Heckman and Hotz (1989) proposed a variety of specification tests aimed at narrowing the range of the non-experimental estimates. In the more recent literature, Heckman, Ichimura,Smith and Todd (1998) used experimental data from a job training social experiment to nonparametrically characterize the nature of selection bias in observed wages. Heckman, Ichimura and Todd (1997) used the same data to study the performance of a class of matching estimators. In both papers, data from a randomized-out control group are combined with observational data to assess the validity of different non-experimental estimators.

---

[4] PROGRESA stands for Programa de Eduacacion, Salud, y Alimentacion (Program of Education, Health and Nutrition).

[5] It is also possible, however, given the experimental design, to account for the possibility that treatment and control groups, or at least the subsamples we use in our analysis, are not identical given data on the treatment group prior to the selection process, i.e., data from the pre-program year (1997).

The estimators studied in the program evaluation literature would not be suitable for evaluating the effect of programs that have not been implemented, an important goal of this paper, because they require data on program participants. In this paper we forecast the effect of treatment (the school subsidy) using data only on non-participants, for whom direct schooling costs (e.g., tuition) are zero. We show that the existence of an active child labor market and variation in child wages, a component of the opportunity cost of school attendance, can be used to identify our schooling model, which enables us to forecast the effect of a school subsidy program without variation in the direct cost of schooling.

A structurally estimated model that is a valid representation of behavior allows the evaluation of the impact of a variety of counterfactual policies.[6] In contrast, social experiments provide information only about the impact of the program as it was implemented. They do not allow the evaluation of variations in the program or longer run effects that extend beyond the life of the experiment and cannot be used to evaluate radically different programs.

Using the model's estimates, we determine the impact of the program for alternative subsidy schedules. In particular, we estimate the (minimum) cost of achieving targeted gains in school completion levels. Second, we assess the impact of extending the program to ineligible families, i.e., those that did not meet the income and assets requirement. Third, we estimate longer-term impacts of the program. Because the program has been in effect for only two years, it is not possible by simply comparing the treated to the untreated households to evaluate the longer run impact of the program. For example, even if the program is viewed by the experimental group as permanent, the impact of the program on the existing families is conditioned on their current circumstances, e.g., the number of children they have and their grade completion levels at that time. However, the longer-run impact of the program may be to affect those circumstances. One obvious longer-run effect of a schooling subsidy may be to alter the number of children families have. In that case, directly evaluating the

[6] Wolpin (1996) reviews a number of examples in the context of discrete choice dynamic programming models.

long-term effect of the program would require that the experimental program be continued long enough to observe these changes, which is costly and sometimes politically infeasible. The model we adopt incorporates decision-making about both child schooling and fertility and, thus, allows an assessment of the longer-run impact of the program. Fourth, we compare the impact of the subsidy program on schooling to radically different policies, e.g., that of enforcing legal restrictions on the use of child labor and mandatory school attendance laws.

The model developed in this paper assumes that a married couple makes sequential decisions about the timing and spacing of births and about the time allocation of children, including their school attendance and market participation. Parents receive utility contemporaneously from their children's current schooling levels and from their leisure time (home production). Parental consumption, which also yields contemporaneous utility, is enhanced by their children's earnings (for children under the age of 16) and parents bear a cost of rearing children. The decision to bear a child (for a woman to become pregnant) is made over a finite horizon beginning at the woman's age at marriage and ending when the woman is no longer fecund (assumed to be age 43). Parent's income (market earnings plus home produced goods that are sold) is an exogenous function of the husband's age. Parental preferences and income and children's earnings are subject to time-varying stochastic shocks, and also differ permanently among households according their type. To reduce the complexity of the behavioral model, we focus on landless households for whom wage earnings account for almost all of household income.

II. A Description of the PROGRESA Program:

We begin with a description of the PROGRESA program and the evaluation research that has already been performed. PROGRESA is a new, large-scale anti-poverty and human resource program begun in Mexico in 1997 that provides aid to approximately 2.6 million poor families living in rural areas.[7] The program's major

[7] These households account for 40 percent of all rural households and 10 percent of all households in Mexico (See Gomez de Leon and Parker (2000))..

goal is to stimulate investments in children's human capital. The program attempts to align household incentives with program goals by providing transfer payments that are contingent on children's regular attendance at school.[8] Programs with features very similar to those of PROGRESA have been initiated in many other Latin American and Asian countries.[9]

In recognition of the fact that older children are more likely to engage in family or outside work, the transfer amount provided under the PROGRESA program varies with the child's grade level. It is greatest for children in junior secondary grades (grades 7 through 9) and is also slightly higher for female children, who traditionally have lower secondary school enrollment levels (see table A.1).[10] In addition to the educational subsidies, the program also provides some monetary aid and nutritional supplements for infants and small children that are not contingent on schooling.[11] In total, the benefit levels that families receive are substantial relative to their income levels. The monthly average total cash transfer is US $55 (more than 75% is due to the educational subsidy), which represents about one-fourth of average family income (Gomez de Leon and Parker, 2000).

For purposes of evaluation, the second phase of the PROGRESA program was implemented as a randomized social experiment, in which 506 rural villages (in 7 states) were randomly assigned to either participate in the program or serve as controls.[12] Randomization, under ideal conditions, allows mean program

---

[8] Children are required to attend at least 85% of days as verified by principals and teachers.

[9] For example, such programs exist in Bangladesh, Pakistan, Chile, Colombia, Brazil, Guatemala, and Nicaragua.

[10] Prior to 1992, Mexico had compulsory schooling that required that children complete at least 6 years of schooling. In 1992, the law was changed to require the completion of 9 years of schooling. However, as our data show, the law is not strictly enforced. Although a large proportion of children complete 6 years of schooling, the vast majority complete less than 9.

[11] Some of this aid is contingent on visiting a health clinic.

[12] The 506 localities were selected in a stratified random sampling procedure from localities identified by PROGRESA to be eligible to participate in the program, because of a "high degree of marginality" (determined mainly on the basis of analysis of data in the 1990 and 1995 population

impacts to be assessed in a simple way through comparisons of outcomes for treatments and controls. Behrman and Todd (2000) provide evidence that is consistent with the randomization having been carefully implemented. They document that the treatment and control groups are highly comparable prior to the initiation of the program. Over the time period covered by our data, the households living in the control villages did not receive program benefits.[13]

The data sets gathered as part of the PROGRESA experiment provide rich information at the individual and household levels, including information on school attendance and achievement, employment and wages of children and the income of the household. Data are available for all households located in 320 villages randomly assigned to the treatment group and for all households located in 186 villages assigned to the control group. The data that we analyze were gathered through two baseline surveys administered in October, 1997 and March, 1998 and through three follow-up surveys administered October, 1998, May, 1999, and November, 1999. Households residing in treatment localities began receiving subsidy checks in the fall of 1998. In addition to the household survey data sets, supplemental data gathered at the village level and at the school level are also available, most importantly for our purpose, the travel distance to the nearest secondary school.

Within treatment localities, only households that satisfy program eligibility criteria receive the school subsidies, where eligibility is determined on the basis of a marginality index designed to identify the poorest families within each community.[14] Because program benefits are generous relative to families' incomes, most families deemed eligible for the program decide to participate in it, although not all families are induced by the

---

censuses (1990 Censo, 1995 Conteo)). There are 31 states in Mexico.

[13] However, the program has recently been expanded to include many of the control localities, so that it is possible that the behavior of the controls groups over the time period we observe them could have been influenced by their expectation of eventually receiving benefits. We will return to this important issue below.

[14] Program eligibility is based in part on discriminant analysis applied to the October 1997 household survey data. The discriminant analysis uses information on household income, assets, and child status in determining program eligibility.

transfers to send their children to school.[15] There are over 30,000 eligible children residing in the experimental (control plus treatment) villages. The data actually pertain to over 75,000 children, because data collection was exhaustive within each village and included children from ineligible families. There are 9,221 separate households in the control villages and 14,856 in the treatment villages.

Most of the existing research on the PROGRESA social experiment focuses on estimating the experimental impacts through mean comparisons of various outcome measures for treatment and control children. Gomez de Leon and Parker (2000) and Parker and Skoufias (2000) examine how children's time use, e.g., time spent working for pay, differs for children participating in the program. Shultz (2000a,b,c), Behrman and Todd (2000), and Behrman, Sengupta, and Todd (2000a,b) analyze the effect of the program on school enrollment rates, grade advancement rates, and student test scores. These papers document significant effects of the program on school enrollment and grade progression rates.

Behrman, Sengupta, and Todd (2000a) present empirical evidence that shows the existence of substantial experimental effects of the program on the process by which children complete their schooling. They estimate a Markov schooling transition model to illustrate the effects of the program along different dimensions, including (a) ages of school matriculation, (b) grade repetition rates, (c) dropout rates, and (d) school reentry rates among dropouts. These are the kind of effects that the behavioral model will forecast using data on the randomized-out control households combined with pre-program data on the treatment households.

Behrman, Sengupta and Todd find that for 12 year old children the grade progression rate is 11% higher for the treatment group than for the control group for the grade 6 to 7 transition, and 9% higher for the grade 5 to 6 transition. The differences are statistically significant as indicated by the p-values. The dropping out rate is lower for the treatment group across all grade levels and the school reentry rate (among dropouts) is about 18% higher. A substantial fraction of treatment group children reentering school reenter at grade 7, which may be

[15] A family could participate in the health component of the program, but not in the school subsidy component.

related to the fact that there is a large marginal increase in the school subsidy levels that occurs between grades 6 and 7.

For children age 6 to10, program participation is associated with earlier ages of school matriculation, less grade repetition and better grade progression. Notably, there are significant program effects for young children at ages when they do not yet receive program subsidies, which suggests forward-looking behavior on the part of the parents as assumed in our dynamic behavioral model. For children in the age range 11 to 14, the program significantly decreases the dropout rate, particularly during the transition from primary to secondary school, and encourages school reentry among those who dropped out prior to the initiation of the program. Overall tests of the equality of the treatment and control estimated transition matrices (Pearson chi-squared tests) reject equality at almost all ages, which indicates a significant effect of the program.

III. Variable Definitions and Descriptive Statistics:

Variable Definitions

Our estimation sample consists of landless households in which there was a woman under the age of 50 reported to be the spouse of the household head. This restriction reduced the sample to 1,531 households located in the control villages in 1997 and 2,162 households located in the treatment villages in 1997. Additional exclusions based on missing or otherwise inconsistent data reduced the sample to 1,362 households in the control villages (of which 1,355 are also observed in 1998) and 1,949 households in the treatment households. As of 1997, there were 4,501children born to the control households and 6,219 to the treatment villages, on average 3.3 and 3.2 children per household respectively. Of these, 2,096 children in the control village and 2,845 in the treatment villages are between the ages of 6 to 15 as of the October, 1997 survey. In contrast to the entire sample, landless households tend to be poorer and, therefore, have a higher proportion of eligible households. As of the 1997 survey, about 52 percent of the all households were eligible to participate in the program, while 66 percent of the landless households were eligible to participate.

Unfortunately, the data provide information concerning school attendance and work essentially only at the survey dates. Therefore, allocating children to the school-work-at home categories that pertain to an entire school year requires additional assumptions. In defining school attendance, we use the data on school enrollment in the week prior to the survey and data on highest grade completed at the time of the survey. Specifically, we followed the following rule in determining school attendance during each of the two school years, 1997-98 and 1998-99, covered by the surveys: (1) A child was considered as having attended school for the entire year if a child that was reported as enrolled in at least one of the two surveys during each school year and was reported as completing at least one grade level. (2) A child was considered as having not attended if the child was reported as not enrolled in both surveys during each school year and did not complete a grade level. (3) Essentially, all other cases were hand-edited to provide a consistent sequence of attendance and grade completion. A child who was determined to have attended school, but did not complete a grade level, was assumed to have failed that school year. School attendance information was obtained for children between the ages of 6 and 15. Highest grade completed was obtained for all children born to the woman.

A child was defined as working during the school year if the child did not attend school using the criteria above and had been reported as working for salary (for 1997, in the October 1997 survey and for 1998, in the October 1998 survey). The weekly wage was provided in the surveys for those who were reported working in the week previous to the survey. A child was defined as being at home if the child was neither attending school nor working.

Parents' weekly income was obtained from the October surveys and includes market earnings of both parents as well as their self-employment income.[16] Both the children's weekly wage and the parents' weekly income were multiplied by 52 to obtain an annual equivalent.

Descriptive Statistics

---

[16] It is extremely rare for children to have contributed to the self-employment income of the household.

Table 1 presents basic sample statistics. The mean age of the wives in the sample as of 1997 is 30.5, and that of their husbands 34.4. The mean age of marriage of the women is 18.1. On average, the families had 3 children as of 1997 and added another .2 children by 1998.The highest grade completed of children age 7 to 11 is 2.4 years of schooling, while children who were between the ages of 12 and 15 had completed 5.8 years, and those age 16 and over, 6.6 years. As also shown, there is almost no difference in the completed schooling of this latter group by sex. Parent's income in 1997 was, on average, about 12,000 pesos (approximately 1,100 U.S. dollars). Approximately 9 percent of children between the ages of 12 and 15 were working as of the 1997 interview. Among those that worked their average income ranged from about 6,000 pesos for those 12 and 13 years of age to over 8,000 pesos for those who were 15 years of age.

Data were also collected about the households' villages. Two "distance" variables are of particular relevance: the distance from the village to a secondary school and the distance from the village to the nearest city. As seen in table 1, about one-quarter of the villages have a secondary school located in the village, and among those villages that do not, the average distance to a village with a secondary school is approximately three kilometers. The villages are also generally quite distant from major cities. The average distance of the households from a city is 135 kilometers.[17]

Table 2 provides more detail concerning the time allocation of children. The first two columns contrast the reported school attendance rates (in percent) of children by sex from ages 6 through 15 based on the raw data, whether or not the child was enrolled as of the October 1997 interview date (in column one), and the revised rates based on the rules described above (in column 2). The third column shows the percentage of children working for pay and the last column the percentage at home (100 - (2) - (3)).

As is apparent from comparing the first two columns, the revised attendance rates are slightly higher than the raw attendance rates. Based on the revised rates, school attendance appears to be almost universal from ages 7 to 11 for both boys and girls. Attendance at age 6 is lower, particularly for boys, although over 90

---

[17] We thank T. Paul Schultz for making this data available to us.

percent. Attendance rates fall to 89 percent for males and to 90 percent for females at age 12, an age by which many children have completed primary school (grade 6). After age 12, attendance rates continue to decline for both sexes, but more rapidly for girls. By age 15, attendance rates are only 48 percent for boys and 40 percent for girls. The percentage of children estimated to be working for pay is under one percent for boys and zero for girls at each age under 12. At age 12, the percentage of children working for pay begins to rise and to diverge by sex. By age 13, 8.6 percent of boys are working for pay and by age 15, 28 percent are working. In contrast, only 16 percent of females are working for pay at age 15.[18]

Girls progress through the early grades somewhat faster than boys, but ultimately complete about the same amount of schooling. As of October 1997, girls who are 12 years of age have completed about .3 more years of schooling on average than have boys of the same age. At age 16, that difference has completely disappeared, with both sexes having completed, on average, 6.6 years of schooling. Girls are more likely to complete sixth grade, but are also more likely drop out of school after completing it. As seen in table 3, among the children in our sample who are age 15 or 16 in 1997, 22 percent of the boys and 17 percent of the girls have less than 6 years of schooling, 32 percent of the boys and 39 percent of the girls have exactly 6 years of schooling and 46 percent of the boys and 44 percent of the girls have more than 6 years of schooling. Failure rates are slightly higher for boys than for girls, 15.2 percent for boys and 14.5 percent for girls over all grades, but considerably higher at the primary grades, 15.7 percent vs. 13.9 percent.[19]

Table 4 provides information about fertility patterns. In particular, it shows the duration distribution from the date of marriage to the birth of each of the first three children. Fertility occurs rapidly after marriage.

---

[18] Child labor laws prohibit children under the age of 14 from working and also limit the kinds of employment and the length of the work day. Our model assumes that these restrictions are not binding, which is consistent with the fact that we observe children under the age of 14 who are working.

[19] National examinations are given at each primary grade level and adequate performance determines grade progression, although compliance is left to teachers. A certificate is awarded after the completion of primary school and junior secondary school.

A little more than 50 percent of the women had their first birth within a year of marriage.[20] First births occurred within two years for seventy percent of the women. As the second column shows, of the women who had at least two children, only 11 percent of the women had two births in two years, but 35 percent had their second birth within 3 years of marriage and two-thirds within 5 years. About 10 percent did not have their second birth until after 10 years of marriage. Less than 2 percent of the women, among those with at least three children, had three births in three years and about one-quarter had three births in 5 years. Over 20 percent of these women had their third birth after 10 years of marriage. Thus, most women have their births quickly after marriage, although some delay for a significant period.

Once children leave school, they only rarely return. As seen in Table 5, only 13 percent of males age 13 to 15 who worked in one year attended school in the next year. Similarly of those who were home in one year, only 15 percent attended school in the next year. Comparable figures for females are 20 percent (although the sample size is only 5) and zero percent. The school-to-school transition over these ages exhibits substantial permanence for both males and females, with 86.2 percent of the males and 76.9 percent of the females who attended school in one year also attending in the next year. The home-to- home transition for females and the work-to-work transition for males also exhibit such permanence. Among females in this age group 92.5 percent of those who were home in one year were also home in the next year, and among males 62.5 percent of those who worked in one year were also working the next year.

IV. The Model:

Model Description

[20] The duration to the first birth is calculated as the age of the woman in 1997 minus the age of the child in 1997 minus the age of the woman at marriage. Ten percent of first births were reported to have occurred at an age prior to the woman's age at marriage and 14 percent coincident with the woman's age at marriage. For the cases where the birth occurred at or before the age at marriage, the marriage was assumed to have occurred one year prior to the birth of the first child. An additional 26 percent of first births occurred at an age one year post-marriage. The sum of these is about equal to the 52 percent of first births occurring in the first year of marriage reported in the table.

In each discrete time period, a married couple makes fertility and child time allocation decisions. Specifically, a decision is made about whether or not to have the woman become pregnant, and have a child in the next period, and, for each child between the ages of 6 and 15, whether or not to send the child to school, have the child work in the labor market (after reaching age 12) or have the child remain at home.[21] At ages older than 15, children are assumed to make their own schooling and work decisions. Women can become pregnant beginning with marriage (at age $t = t_m$) and ending at some exogenous age (at t = T-1) when she becomes infecund. The contribution of the husband and wife to household income is exogenous (there are no parental labor supply decisions) and stochastic, and the household cannot save or borrow.[22] The contributions to household income from working children (under the age of 16) are pooled with parental income in determining household consumption.

Children who neither attend school nor work for pay are assumed to contribute to household production (parental utility) with the level of their contribution depending on their age and sex (and possibly on the age composition of other children in the household). Thus, the cost of sending a child to school consists of the opportunity loss in either home production or household income, each of which may differ by the child's age and sex. Parents are also assumed to derive utility in each period from the current average level of schooling that their children have completed, from the current number of children who have graduated from elementary school (grade 6) and from the current number who have graduated from junior secondary school (grade 9). Schooling is publicly provided and therefore parents bear no direct costs. All of the villages have their own primary schools (grades 1 through 6), but not all villages have junior secondary schools (grades 7 through 9). We allow for a psychic cost of attending a secondary school (grades beyond 6) that varies with the distance

[21] Although some information on contraceptive use is available, it is not detailed enough to allow modeling contraceptive decisions.

[22] Labor force participation rates of married women are quite low, between 10-20% at most ages.

from the village to the nearest village with a secondary school. We allow for an additional psychic cost if a child attends grade 10, which often involves living away from home.[23]

More formally, let p(t) = 1 if a woman becomes pregnant at age t (and 0 otherwise) in which case a child is born at t+1, n(t+1)=1. Further let b(t+1) = 1 if the child that is born is male (and 0 otherwise) and g(t+1) = 1 if the child is female. Also, let **p(t)** be the vector of pregnancies up to age t, and **n(t+1)** be the corresponding vector of births that occur up to age t+1, **b(t+1)** the corresponding vector of male births and **g(t+1)** the vector of female births.[24] The stock of children through t (the sum of pregnancies through t -1) is denoted by N(t) = N(t-1) + n(t), and analogously the stock of male children by B(t) and that of female children by G(t). A child born at the woman's age τ is zero years old at τ and, thus, t - τ years old at t. A child of birth order n is born at the woman's age τ_n.

Let s(t, τ) = 1 if a child of age $t - \tau$, between the ages of 6, the minimum age of school eligibility, and 15, the last age at which parents are assumed to make a schooling decision, attends school at t and zero otherwise. The corresponding vector of school attendance decisions for school age children at t is **s(t)**, where an element is zero when there does not exist a child of a given age. Cumulative schooling at t for a child born at τ is given by S(t, τ) = S(t-1, τ) + c(t-1, τ) s(t-1, τ), where c(t-1, τ) =1 if a year of schooling is successfully completed and zero otherwise and the corresponding vector over all children attending school in that period is **c(t-1)**. The completion of a grade level conditional on attendance is probabilistic. The probability of completion is given by $\pi^c(t, \tau, S(t-1), \mu_c | s(t-1) = 1)$, where μ_c is a permanent family-specific component of the success probability. The completion probability also may differ by the child's sex. The vector of cumulative schooling at t over all children is $\mathbf{S(t)}$ and the mean schooling level of those children at t, $\overline{S}(t)$. Sex-specific schooling variables can be similarly defined and denoted with b or g subscripts.

---

[23] It is more straightforward to treat these costs as utility losses rather than monetary costs, given that full monetary costs associated with school attendance are not observed in the data.

[24] Bold type is used to indicate a vector.

Finally, let $h(t, \tau) = 1$ if a child born at τ works at t, and zero otherwise, with **h(t)** the corresponding vector over all children at t. Children must be at least twelve years old to be eligible for work, i.e., $h(\tau+k, \tau)=0$ for k<8. The residual time allocation for a child who is neither in school nor at work is to be at home, which we denote as $l(t, \tau) =1 - h(t, \tau) - s(t, \tau)$. Sex-specific variables, as before, carry b and g subscripts.

Parental contemporaneous utility at any time t depends on household consumption, the stock of children and their timing and spacing, whether the woman is currently pregnant and her age at pregnancy, the number of children of various ages and sexes that are at home currently (neither in school nor at work) and their schooling levels. Preferences for being pregnant and for having children at home are assumed to be subject to stochastic shocks. In addition, households are assumed to differ permanently in their preferences for the number of children they have and for their children's schooling. The utility function is thus given by

$$\begin{aligned} U(t) = U(&C(t),\ p(t),\ \mathbf{n(t)},\ \mathbf{l_b(t)},\ \mathbf{l_g(t)},\ \mathbf{S(t)}; \\ &z_s,\ \epsilon_p(t),\ \epsilon_{lb}(t),\ \epsilon_{lg}(t);\ \mu_N,\ \mu_S,\ \mu_{lg},\ \mu_{lb}), \end{aligned} \tag{1}$$

where C(t) is household consumption, z_s is the distance to a secondary school, the ϵ's are stochastic shocks to being pregnant and to the value attached to having children of each sex at home and the μ's reflect permanent differences across households in their preferences for children, for schooling and for the home time by sex.

The parental utility function (1) is written generally enough to include the possibility, for example, that the value of household production is greater for older female children when there are also very young children in the household. It also allows for a quantity-quality fertility tradeoff if the quantity of children and their average schooling are substitutes in parental utility (Becker and Lewis (1973), Rosenzweig and Wolpin (1980)). The exact representation of the utility function is determined according to model fit criteria.[25]

[25] It is the absence of economic theory about the form of the utility function and our inability to directly elicit preferences that makes necessary this kind of pre-testing of the model.

Family consumption at t is equal to total family income.. Family income is the sum of parental income (y_p) and the earnings of children (y_o) who work in the market.[26] Thus, the family's budget constraint is given by

$$(2) \quad C(t) = y_p(t) + \sum_n y_o(t, \tau_n) h(t, \tau_n).$$

Income generating functions differ for parents and children. Parental income at t, which includes both earnings and self-employment income, depends on the age of the male parent, on the distance of the household's village from a city, a random shock at t and a permanent parent-specific unobservable component. Similarly, the earnings of a child depends on the child's age and sex, on the distance of the household's village from a city, on a time-varying (but not child-varying) shock and on a permanent unobservable component that is the same for all children (within the same household). The distance from a city affects wage offers due to differences in the skill price reflecting the extent of the labor market to which the household has access. Specifically,

$$(3) \quad \begin{aligned} y_p(t) &= y_p(a_p(t),\ z_c,\ \epsilon_{y_p}(t);\ \mu_{y_p}) \\ y_o(t,\ \tau_n) &= y_o(t-\tau_n,\ I(b(\tau_n) = 1),\ z_c,\ \epsilon_{y_o}(t);\ \mu_{y_o}). \end{aligned}$$

The five time-varying ϵ-shocks are assumed to be jointly serially uncorrelated. Their joint contemporaneous distribution is denoted by $f(\boldsymbol{\epsilon}(\mathbf{t}))$.[27] The permanent components of parental preferences and income, of child earnings and grade completion are also assumed to be jointly distributed according to $g(\boldsymbol{\mu})$. In the application, we assume g to be discrete with a fixed number of support points, which we denote as indicating family "type." These permanent components are known to parents from the beginning of the marriage.

[26] Child rearing costs are essentially indistinguishable from the psychic value of children of different ages included in parental utility.

[27] The implicit time-varying shock to grade completion is assumed to be independent of all other shocks in the model.

At any t, the couple is assumed to maximize the present discounted value of remaining lifetime utility. In any period, the family will face K(t) mutually exclusive alternatives, where K varies over time with the number of children eligible to attend school and work and the woman's age. Define $d_k(t) = 1$ if the kth alternative is chosen at t, and = 0 otherwise. (The ordering of the K(t) alternatives is irrelevant.) Further, define $\Omega(t)$ to be the state space at t, namely all of the relevant factors that affect current or future utility or that affect the distributions of future shocks, that is, $\mathbf{b(t)}$, $\mathbf{g(t)}$, $\mathbf{S_b(t)}$, $\mathbf{S_g(t)}$, $a_p(t)$, $\boldsymbol{\epsilon}\mathbf{(t)}$, $\boldsymbol{\mu}$, t_m, z_s, z_c .

The maximized present discounted value of lifetime utility at t, the value function, is given by

$$(4) \quad V(\Omega(t),t) = \max_{\{d_k(t)\}} E\{\sum_{\tau=t}^{\overline{T}} \delta^{\tau-t} U(t) \mid \Omega(t)\}$$

where $\overline{T}$ is the end of the couple's life and the expectation is taken over the distribution of parental preference and income shocks, the children's earnings shock and the implicit shocks to grade completion for choices that involve school attendance for some or all the children.[28] The solution to the optimization problem is a set of decision rules that relate the optimal choice at any t, from among the feasible set of alternatives, to the elements of the state space at t. Recasting the problem in a dynamic programming framework, the value function can be written as the maximum over alternative-specific value functions, $V^k(\Omega(t),t)$, i.e., the expected discounted value of alternative $k \in K(t)$, that satisfy the Bellman equation, namely

$$\begin{aligned} & V(\Omega(t),\ t) = \max_{k \in K(t)} [V^k(\Omega(t),\ t)] \\ (5) \quad & V^k(\Omega(t),\ t) = U^k(t,\ \Omega(t)) + \delta E(V(\Omega(t+1), t+1 | d_k(t)=1), \Omega(t)) \quad \text{for } t < \overline{T}, \\ & \qquad = U^k(\overline{T},\ \Omega(\overline{T})) \qquad \text{for } t = \overline{T}. \end{aligned}$$

Model Solution:

[28] The integration is also performed over whether a birth outcome is a boy or a girl. We assume the probability of each gender outcome to be .5.

The solution of the optimization problem is in general not analytic. In solving the model numerically, its solution consists of the values of $EV(\Omega(t+1), t+1 \mid d_k(t)=1, \Omega(t))$ for all k and elements of $\Omega(t)$. We refer to this function as Emax for convenience. As seen in (5), treating these functions as known scalars for each value of the state space transforms the dynamic optimization problem into the more familiar static multinomial choice structure. The solution method proceeds by backwards recursion beginning with the last decision period.

There are two complications in solving the model numerically. First, at any fecund period in which there are children of school and work age the choice set is of order $2 \cdot 3^{N_1(t)}$, where the first term represents the choice of whether or not to have a child and the second reflects the number of joint school attendance - work choices (of which there are 3). For example, if there are three children between the ages of 8 and 15, there are 54 possible choices.[29] One way to reduce the size of the choice set in a way that is (mostly) consistent with the data is to assume that for each sex, a child may attend school only if all younger children attend school and, independent of sex, a child may work for pay only if all older children work for pay.[30] In the case of three school/work age children, if they are of the same sex the number of alternatives is now reduced to 20. We do not impose these restrictions on 6 and 7 year old children to accommodate the fact that school entry is sometimes delayed.

Second, the size of the state space makes a full solution of the problem computationally intractable. The Emax functions must be calculated for all state values at each t. As long as the ages of children affect lifetime utility, as it does because of the age restrictions on children's eligibility for schooling and work, the state space will include the entire sequence of births (by sex given that the value of having a child remain at home may be sex-specific) and not simply the stock of children. With 30 fecund periods, there are 3^{30} such sequences. In addition, even though only the average schooling of children age 16 and over affects utility, at any t the

[29] For women under the age of 50 in the 1997 data, the modal number of children of school age is 3 (about 25%) and 98% of households have five or fewer.

[30] Violations of the assumption in the 1997 survey occur in about 5% of the households in the case of schooling and in about 1% of the households for working.

schooling level of each child affects expected lifetime utility at t because children are of different ages. To solve the dimensionality problem, we adopt an approximation method in which the Emax functions are expressed as a parametric function of the state variables or composites of the state variables, using methods developed in Keane and Wolpin (1994, 1997, 1999). In particular, the Emax functions are calculated at a subset of the state points and their values are used to fit a polynomial approximation in the state variables.[31] In addition, to limit the size of the state space, we also assume that women can have no more than eight children.[32] As in Keane and Wolpin, the multivariate integrations necessary to calculate the expected value of the maximum of the alternative-specific value functions at those state points are performed by Monte Carlo integration over the ϵ-shocks.

Model Estimation:

The solution to the agents' maximization problem serves as input into estimating the parameters of the model. The numerical solution method described above provides (polynomial approximations to) the Emax functions that appear on the right hand side of (5). The alternative-specific value functions, $V^k(t)$ for k=1,..,K(t), are known up to the parental random preference and income shocks and the earnings shock of the children. Thus, conditional on the deterministic part of the state space, the probability that an agent is observed to choose option k takes the form of an integral over the region of a subset of the random shocks such that k is the preferred option.

Specifically, in the decision model presented above the observed outcomes at each period include (i) the choice (from the feasible set) made by the couple, whether or not to initiate a pregnancy, which children to send to school, which to work in the market and which to remain at home, (ii) the wages received by the children who work in the market, (iii) the success or failure of those children who attend school to complete a

[31] Stinebrickner (2001) develops a local approximation to the Emax function.

[32] Only about 3 percent of women in our sample report having more that eight children. In the empirical implementation, we assume that children of birth order greater than eight were not born.

grade level and (iv) parental income. Let the outcome vector at t be denoted by $O(t)=\{d^k(t), \mathbf{y_o(t)}, \mathbf{c(t)}, y_p(t)\}$. Suppose we observe these outcomes for a sample of N households beginning at marriage, t = t_{mn}, and ending at some t=$\overline{t_n}$. Then, the likelihood for this sample is

$$(6) \quad \prod_{n=1}^{N} \Pr(O(\overline{t_n}),\ \ldots\ldots\ ,\ O(t_{m+1,n}),\ O(t_{mn})) \mid \overline{\Omega}(t_{mn}), \boldsymbol{\mu}),$$

where $\overline{\Omega}(t_{mn})$ is the observable components of the initial state space at the time of marriage, that is, the state space net of the family's type (the $\boldsymbol{\mu}$ vector) and stochastic shocks at $t=t_{mn}$. The observable part of the state space at marriage consists only of the age of the woman at t_{mn}, the age of the man at t_{mn} and distance from a secondary school. Because type is unobserved, it must be integrated out. Thus, the sample likelihood is

$$(7) \quad \sum_{j=1}^{J} \prod_{n=1}^{N} \Pr(O(\overline{t_n}),\ \ldots\ldots\ ,\ O(t_{m+1,n}),\ O(t_{mn})) \mid \overline{\Omega}(t_{mn}), \text{type}=j) \Pr(\text{type}=j \mid \overline{\Omega}(t_{mn}))\ ,$$

We assume that the initial conditions, the ages of marriage of both parents and the distance from a secondary school, are exogenous conditional on type.

There are two additional considerations in computing the likelihood. Because we assume that the child wage shock is family-specific, having an observation on the wage for two children in the same family working in the same period who have different wages (conditional on the relevant observable determinants of child earnings, child age and sex as in (3)) will lead to a degenerate likelihood. We therefore assume that the children's wages are measured with error, which seems like a reasonable assumption in any event.[33] Thus, assuming a multiplicative measurement error observed child earnings is given by $y_o^{obs}(t) = y_o(t)\exp(\eta(t))$.

Another difficulty arises because, for most of the families, we do not observe decisions for consecutive periods. In particular, although we have a complete fertility history for all women, we do not have a complete school attendance and work history for children who are above the school or work eligibility ages at the first

[33] We follow this strategy as opposed to allowing for child-specific wage shocks to avoid having to integrate over all of the child shocks in calculating the Emax functions. The problem of degeneracy exists more generally, namely that with family level shocks some choices may not be generated by the model. Restricting the choice set as above reduces the likelihood of this event, but does not eliminate it necessarily. Our estimation procedure smooths over zero likelihood events. After estimation, we will check to see whether any outcomes in the data have zero probability of occurrence.

survey. For example, consider a family with 3 children whose ages are 10, 13 and 16 as of the October 1997 survey date and whose marriage occurred in 1980 when the woman was age 19 (t_m). For this family, we observe fertility outcomes at every t between 1980 and 1997, the woman's age 19 through 36. However, the family began to make schooling decisions when the oldest child was 6, in 1987 when the woman was age 26. We are thus missing the full set of outcomes from the woman's age 26 through 35. Although it is conceptually straightforward to accommodate this feature of the data into the likelihood function (7), it is computationally infeasible to perform the integrations over all of the feasible unobserved choice paths as would be required to calculate the likelihood.

In order to avoid missing outcome data, it would be necessary to restrict the sample to marriages that occurred between 1989 and 1997 only. For the earliest marriages in this range, in 1989, the oldest child would turn 6 in 1997, the first age at which a schooling decision is made. Although those families provide useful identifying information, it is obviously not possible to identify all of the parameters of the model solely from those observations because we would not observe schooling decisions beyond the age of 7 and no children work before age 8.

For all families, we do observe the complete set of outcomes in the two survey years, 1997 and 1998. The difficulty in using that data is that the state variables at the time of the surveys, including for instance the birth history and the schooling levels of all children, are not exogenous.[34] The assumption of serial independence in the shocks, however, implies that the state variables at any time t are exogenous with respect to decisions at t conditional on type. Thus, the likelihood for the observations in 1997 and 1998 can be written, analogous to (7),

$$(8) \quad \sum_{j=1}^{J} \prod_{n=1}^{N} \Pr(O(t_n^{98}),\ O(t_n^{97}) \mid \overline{\Omega}(t_n^{97}), \text{type}=j) \Pr(\text{type}=j \mid \overline{\Omega}(t_n^{97})) \ ,$$

[34] This is exactly the initial conditions problem in discrete choice models as discussed in Heckman (1981).

where t_n^{97} and t_n^{98} are the ages of the woman in 1997 and 1998. The problem with (8) is that we must specify how the type distribution is related to the state variables. In actuality, the form of this conditional distribution function is given by the structure of the behavioral model together with the relationship between type and the initial state variables, i.e., the second term in (7). There is clearly a trade-off in how one specifies this conditional type distribution. The more flexible the functional form the better the approximation to its true functional form and the closer the exogeneity requirement is met. However, the more flexible the form, the more parameters there are to estimate. Furthermore, these parameters are themselves functions of the structural parameters, and are therefore extraneous.

To summarize, in estimating the model we use (7) for the families with complete decision histories as described above (Sample A) and we use (8) for the families with incomplete decision histories (Sample B), ignoring the information about pregnancy decisions made prior to 1997 . Now, given the assumption of joint serial independence of the vector of shocks (conditional on type), both (7) and (8) can be written as the product of within-period outcome probabilities conditional on the corresponding state space and type. Each of these conditional probabilities are of dimension equal to the number of contemporaneous shocks in $\epsilon(t)$.

To illustrate the calculation of the likelihood, it is sufficient to consider a specific outcome at some period. Suppose that the kth alternative, $d_k(t)$, that is chosen at period t is to send at least some children to work. The children who work are observed to have wages given by $y_{oj}(t)^{obs}$, where j signifies the jth working child and the superscript “obs” distinguishes the observed wage from the true wage, $y_{oj}(t)$. Then the likelihood contribution for such an observation is (for a given type)

$$\Pr(d^k(t)=1, \tilde{y}_o(t)^{obs} \mid \Omega(t), \text{type}) = \int_{\tilde{y}_o(t)} \Pr(d_t^k=1 \mid \tilde{y}_o(t), \Omega(t), \text{type}) \cdot \Pr(\tilde{y}_o^{obs}(t), \tilde{y}_o(t) \mid \Omega(t), \text{type}), \tag{9}$$

where "~" signifies the vector of child wages over j and the integration is of the same order as the number of children who work.[35] Notice that it is necessary to integrate over the vector of true wages in (9) because the choice probability depends on true wages , which we observe only with error. Probability statements for other alternative choices are calculated similarly. We calculate the right hand side of (9) by a smoothed frequency simulator.[36]

The entire set of model parameters enter the likelihood through the choice probabilities that are computed from the solution of the dynamic programming problem. Subsets of parameters enter through other structural relationships as well, e.g., child wage offer functions, the parents' income function and the school failure probability function. The estimation procedure, i.e., the maximization of the likelihood function, iterates between the solution of the dynamic program and the calculation of the likelihood.

Approximate Decision Rules

In this section, we present estimates of approximate decision rules order to investigate relationships among the variables in the state space of our model and the decisions that they are assumed to affect. These approximate decision rules can be used as a guide to the choice of specific state space elements, as discussed below. Approximate decision rules can be motivated in the following way.

---

[35] For convenience, we have ignored parents' income in the formulation of the likelihood function as well as whether the children that were sent to school failed to progress to the next grade level; the modifications of (9) to account for these additional observable variables are straightforward.

[36] The kernel smoothed frequency simulator we adopt was proposed in McFadden (1989). For each of K draws of the error vector, $\epsilon_p(t), \epsilon_{lb}(t), \epsilon_{lg}(t), \epsilon_{y_p}(t), \eta(t)$, noting that $\epsilon_{y_o}(t)$ is chosen to satisfy the observed wage for each child, that is, inclusive of the measurement error. The kernel of the integral is

$\exp[\frac{V^k(t) - \max(\mathbf{V^j(t)})}{\tau(t)}] / \sum_i \exp[\frac{V^i(t) - \max(\mathbf{V^j(t)})}{\tau(t)}]$ times the joint density of the observed and true wage, where the j superscript denotes the vector of value functions over all alternatives. The first term in the kernel is the smoothed simulator of the probability that $d_k(t) = 1$, with $\tau(t)$ the smoothing parameter, which is allowed to vary with t. Specifically, $\tau = \tau_0 \cdot (1 + .25(49 - t))$, where τ declines with age because the magnitude of the value functions also decline with age. See Keane and Wolpin (1997) and Eckstein and Wolpin (1999) for further applications.

Let $V^{\neg k}(t, \Omega^{\neg k}(t))$ denote the maximum of the alternative-specific value functions excluding that of alternative k. Then, the decision rule for the optimal choice takes the form

$$(10) \quad \begin{aligned} d^k(t) &= 1 \quad \text{iff} \quad V^k(t,\ \Omega^k(t)) - V^{\neg k}(t,\ \Omega^{\neg k}(t)) = F^k(t,\ \Omega(t)) > 0 \\ &= 0 \qquad \text{otherwise} \ . \end{aligned}$$

Note that although the alternative-specific value functions may depend on specific subsets of the state space, the decision rule depends on the entire state space. Given the state variables that enter the model, to estimate approximate decision rules, then, requires choosing functions F^k.

Table 6 presents logit regressions showing the relationship between the state variables and the probability of becoming pregnant. In the first specification, the only initial conditions included are the woman's age of marriage and the distance of the household from a secondary school; all of the other state variables evolve with time, either exogenously, e.g., the woman's age at the decision period, the parent's income, or as a function of past choices, e.g., the sequence of prior births as represented by the age distribution of children at the decision period. In the second specification, initial conditions are augmented to include the parents' education and the state of residence, incorporated in our structural estimation as unobserved permanent components of preferences and constraints. Although these additional initial conditions are jointly statistically significant, the results do not differ greatly.

As seen in the table, the probability of a pregnancy declines with both the wife's and husband's age and increases with the wife's age at marriage, given her current age, past fertility and the other state variables. The model would not predict the existence of an effect of wife's age at marriage except through its effect on past decisions, which are being held constant (in a regression sense) as is the remaining fecund period given her current age. We allow for such an effect through the correlation between age at marriage and unobserved heterogeneity. As is consistent with table 4, a woman who was pregnant in the previous period (has a current newborn) is considerably less likely to become pregnant, reflecting biological constraints as well as, perhaps, preferences. Overall, the relationship between the probability of becoming pregnant and the age distribution of existing children follows an inverted u-shape, peaking at age 7 (although there is an aberrant increase if there is 13-year old). In a homogeneous population, and in the context of our model, the average schooling of existing children, at any parent ages, will reflect the accumulation of past shocks to preferences and constraints. The results in table 6 would imply, under this interpretation, that those families that experienced shocks that induced them to choose more frequently to send their children to school at least until age 11, rather than to choose the work or home alternative, would be less likely to choose to have an additional pregnancy (consistent with the quality-quantity fertility model of Becker and Lewis (1973)) . An alternative interpretation is that families differ in their preferences or constraints not captured in either specification that leads to this inverse relationship (see Rosenzweig and Wolpin (1980)). The further the distance that the family lives from a secondary school the greater is the likelihood of having an additional child, which is also consistent with a quality-quantity tradeoff through one of the above mechanisms. Higher parent's income is associated with a lower pregnancy likelihood, consistent with a quality-quantity trade-off.[37] Finally, the control group, conditional on all of the state variables, does not differ from the treatment group (in the pre-program year, 1997) in their pregnancy likelihood.

Table 7 presents similar logit regressions for the approximate decision rule regarding school attendance. As in the previous table, the two specifications do not differ greatly. In the regression, we treat each child in the family as a separate observation, although in the model the decision is jointly made about all of the children. As is reflected in table 2, the probability of attending school declines with the child's age. It also declines with the mother's age (given age at marriage and the other state variables). Attendance also declines with the existence of an additional child at most ages of that child, with the largest reductions occurring with the existence of a 5 or 6-year old and children age 12 and over.[38] The higher is the child's own attained

[37] Transitory income shocks could affect decisions only if there are borrowing constraints. Recall that the model assumes that the family's budget constraint is satisfied period by period..

[38] A test of coefficient equality among all of the coefficients on the age-specific child dummies clearly rejects equality, i.e., the age distribution as well as the total number of children matters to school attendance.

schooling at the decision period, the more likely is the child to attend school for an additional year. Also, the higher is the average school attainment of 7-11 year old children the greater the probability that a given child will attend school. As with the relationship between schooling and pregnancies, there are multiple interpretations for these results as well. Living further from a secondary school is associated with a lower propensity to attend school and parent's income, although imprecisely estimated, with a higher propensity.

V. Estimation Results:

VI. Using the Experiment to Test the Validity of the Model:

Given the parameter estimates, it is straightforward to predict the impact of the school subsidy program on school attendance. A subsidy paid to the family for each child that attends school augments family income and affects the family's school attendance and fertility decisions by changing the family budget constraint (2). Resolving the optimization problem for each family in the presence of the subsidy will lead to a different pattern of school attendance and fertility decisions. Comparing the decisions of the treatment group predicted by the model to their actual decisions (at the same life cycle stage and actual state variables) will provide a direct out-of-sample test of the model's validity.

In order to be able to forecast the impact of the program, it is necessary that the (relevant) parameters of the model be identified. To show how identification is achieved and which parameters are critical for making such a forecast, we present a very simplified version of the model. Consider a household with one child making a single period (myopic) decision about whether to send the child to school or to work, the only two alternatives. Let utility of the household be separable in consumption (C) and school attendance (s), namely u = C + $(\alpha+\epsilon)$s, where s=1 if the child attends school and 0 otherwise and ϵ is a preference shock. Assume that the preference shock is distributed as normal with mean zero and variance σ^2. The family's income is y + w(1-s), where y is the parent's income and w is the child's earnings if working. The family chooses to have the child attend school if and only if $\epsilon \geq w - \alpha$. The unknown parameters of the model are thus α and σ. In this simple model, the probability of the child attending school is $1 - F((w-\alpha)/\sigma)$. Clearly, it is both necessary and sufficient to obtain estimates of α and σ that child wages vary among families and that we observe those wages.[39]

Now, suppose the government is contemplating a program to increase school attendance of children through the introduction of a subsidy to parents of b if they send their child to school. Under such a program, the probability that a child attends school will increase by $F((w-\alpha-b)/\sigma)-F((w-\alpha)/\sigma)$. As this expression indicates, knowledge of α and σ is sufficient for the government to forecast the impact of the program.[40] Moreover, it is also sufficient in order to be able to forecast the effect of varying the amount of the subsidy on school attendance.

A potentially important qualification is that in order to obtain a prediction of the program's effect, it is necessary to make an assumption about the extent to which the families in the treatment villages viewed the program as permanent. If we err in our assumption about the perception of the program's permanence, our forecast of the program's effect based on the estimated model will be incorrect even if the underlying behavioral model is not. One approach to this problem is to estimate the degree to which the program is viewed as permanent by the treatment villages. A practical method would be to treat the duration of the program as a random variable and to estimate its distribution using the observed outcomes in the treatment villages by maximizing the same likelihood as for the control villages but conditioning on the estimates of the behavioral model from the control villages. We could then compare the model's fit to the actual outcomes in the treatment villages.

---

[39] More precisely, in order to use the probability statement above, we need to observe child wage offers. If we observe only accepted wages, that is, the wages of children who work, then we need to able also to identify the parameters of the offered wage distribution togther with α and σ. In the more complete model, identification is achieved through distributional and exclusionary assumptions; child wages are log normal and are not affected by, for example, the age at which the mother married.

[40] See Wolpin (2000) for a similar analysis of the informational content of probabilistic subjective expectations.

A more serious problem arises if the families in the control villages were aware of the program's existence and assigned some non-negligible probability to being included in the program at some future time. Indeed, all of the control villages have now been included in the program. If families anticipated the introduction of the program in the control villages, then the experiment is itself contaminated. An evaluation of the program that did not account for behavioral changes in the control villages induced by anticipating the program would not estimate the treatment effect against the benchmark of having no program. It is possible to adapt the behavioral model and the likelihood function for the control population to account for this anticipation. Upon estimation, tests of the validity of the behavioral model can be conducted as described above.[41]

If the predictions based on estimates of the behavioral model about the program's effect on the school attendance decisions of the families in the treatment villages are close to the effects estimated under the experiment, then the estimated behavioral model can be exploited for additional evaluative purposes that are impossible within the strict experimental design, as previously noted. We can obtain estimates of the program's impact that account for longer term effects on family fertility, vary the size of the subsidy, determine how the ineligible sub-populations vary in their response to the subsidy (e.g., families with higher income) and compare the subsidy program to alternative programs, such as setting legal limits on the use of child labor.

VII. Conclusions:

[41] There is the additional problem concerning the perceptions about the program's permanence by the families in the control villages.

References

Becker, Gary and Lewis, H. Greg (1973): "On the Interaction between the Quantity and Quality of Children" Journal of Political Economy, 81, 4, Part 2, S143-62.

Behrman, Jere and Petra Todd (2000): "Randomness in the Experimental Samples of PROGRESA"

International Food Policy Research Institute, Washington, D.C.

Behrman, Jere, Sengupta, Piyali, and Petra Todd (2000): "Progressing through PROGRESA: An Impact Assessment of Mexico's School Subsidy Experiment," unpublished manuscript.

Behrman, Jere, Sengupta, Piyali, and Petra Todd (2000): "The Impact of PROGRESA on Achievement Test Scores in the First Year" *International Food Policy Research Institute,* Washington, D.C.

Eckstein Zvi and Kenneth I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities and Abilities," *Econometrica* 67(6),1295-1340.

Gomez de Leon, J. and Susan Parker (2000): "The Impact of Anti-Poverty Programs on Children's Time Use," *International Food Policy Research Institute,* Washington, D.C.

Heckman, James J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete-Time Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. Manski and D. Mcfadden, 179-197.

Heckman James J. and V. Joseph Hotz (1989): "Choosing Among Alternative Methods of Estimating the Impact of Social Programs: The Case of Manpower Training ," *Journal of the American Statistical Association*, 84(408), 862-974.

Heckman, James J., Ichimura, Hidehiko and Petra Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 64(4), 605-654.

Heckman, James J., Ichimura, Hidehiko, Smith, Jeff, and Petra Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5), 1017-1098.

Keane, Michael and Wolpin, Kenneth I. (1994): “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence,” *Review of Economics and Statistics* 1994(4), 648-672. .

Keane, Michael and Wolpin, Kenneth I. (1997): “Career Decisions of Young Men” *Journal of Political Economy*, 105, 473-522.

Keane, Michael and Wolpin, Kenneth I. (1999): “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment” *International Economic Review*, forthcoming.

LaLonde, Robert (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *American Economic Review*, 76, 604-620.

Lumsdaine, Robin L., Stock, James H. and Wise, David A. (1992): “Pension Plan Provisions and Retirement: Men and Women, Medicare, and Models,” in D. A. Wise (ed.) *Studies in the Economics of Aging*, Chicago: University of Chicago Press.

McFadden, Daniel (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration” *Econometrica*, 57, 5, 995-1026.

Rosenzweig, Mark R. and Wolpin, Kenneth I. (1980): “Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment,” Econometrica, 48, 227-40.

Schultz, T. Paul (2000a): “Progresa's Impact on School Enrollments from 1997/98 to 1998/99,” *International Food Policy Research Institute*, Washington, D.C.

Schultz, T. Paul (2000b): “Progresa's Impact on School on School Attendance Rates in the Sampled Population,” *International Food Policy Research Institute*, Washington, D.C.

Schultz, T. Paul (2000c): “School Subsidies for the Poor: Evaluation a Mexican Strategy for Reducing Poverty,” unpublished manuscript, Yale University.

Skoufias, Emmanuel and Susan Parker (2000): “The Impact of PROGRESA on Work, Liesure and Time Allocation,” *International Food Policy Research Institute*, Washington, D.C.

Steinbrickner, Todd E. "Compensation Policies and Teacher Decisions" *International Economic Review*, 42(3), 751-780.

Wolpin, Kenneth I. "Public Policy Uses of Dynamic Programming Models" *American Economic Review, Papers and Proceedings*, 86(2), 427-432.

Table 1
Means and standard deviations of selected variables

	Mean	Standard Deviations
Wife's age in 1997	30.5	8.1
Husband's age in 1997	34.4	9.5
Wife's age at marriage	18.1	3.4
Number of children ever born (1997)	3.01	1.92
Number of children ever born (1998)	3.06	1.84
Number of children ever born to women age 35-49 (1997)	4.05	2.14
Highest grade completed of children age 7-11	2.39	1.41
Highest grade completed of children aged 12-15	5.79	1.76
Highest great completed of children at age 16 +		
All	6.60	2.81
Males	6.64	2.82
Females	6.56	2.81
Percentage with secondary school in Village	26.7	-
Distance to secondary school if not in Village (km)	2.82	1.60
Distance to a city (km)	136	74
Parents' income (pesos) in 1997	12,012	13,073

Market income in 1997 of working children age:		
12	6,337	2,789
13	6,314	2,878
14	7,375	5,524
15	8,072	4,941

Table 2
Percent of children attending school, working and home by age and sex [a]

Age	Attends school (Oct. 1997)		Attends school (revised)		Works		At home	
	M	F	M	F	M	F	M	F
6	91.0	94.9	92.9	95.3	-	-	7.1	4.7
7	97.8	97.4	98.9	97.8	-	-	1.1	2.2
8	97.5	97.3	98.6	99.2	0.0	0.0	1.4	0.8
9	99.6	98.4	99.6	99.2	0.4	0.0	0.0	0.8
10	97.2	97.9	97.6	98.7	0.4	0.0	2.0	1.2
11	97.7	95.9	98.6	96.9	0.0	0.0	1.4	3.1
12	89.2	89.3	88.7	90.0	2.5	1.1	8.8	8.9
13	78.1	67.5	78.1	70.9	8.6	4.0	13.4	25.2
14	66.9	58.8	67.3	60.4	16.1	10.1	16.7	29.5
15	48.7	38.5	47.7	40.2	27.5	15.6	24.8	44.3

a. Control and treatment groups in 1997

Table 4
Distribution of the duration of marriage to first, second and third pregnancies[a]

Duration of marriage (years)	First Pregnancy	Second Pregnancy	Third Pregnancy
1	52.4	-	-
2	18.5	10.6	-
3	9.7	24.5	1.7
4	4.6	18.6	8.9
5	3.4	11.9	16.0
6	2.5	8.3	15.4
7	1.5	6.4	12.4
8	1.4	4.2	10.1
9	1.3	3.3	8.0
10	1.1	2.2	6.6
11	0.8	1.8	4.5
12+	2.7	8.2	16.5

a. Control and treatment groups in 1997

Table 3
Distribution of highest grade completed at ages 15 and 16[a]

Years of schooling	Male Children	Female Children
0	2.9	2.3
1	1.0	0.8
2	2.3	1.6
3	3.6	1.2
4	4.5	3.1
5	7.8	8.2
6	32.0	38.5
7	10.7	5.8
8	12.3	12.1
9+	23.0	26.5

a. Control and treatment group in 1997

Table 5
One period transition rates by sex: age(a) 13 to 15

	Home (a)	Work (a)	School (a)
		Males	
Home (a-1)	44.4	40.7	14.8
Work (a-1)	25.0	62.5	12.5
School (a-1)	8.3	5.5	86.2
		Females	
	Home (a)	Work (a)	School (a)
Home (a-1)	92.5	7.5	0.0
Work (a-1)	40.0	20.0	20.0
School (a-1)	21.5	1.5	76.9

Table 6
Approximate decision rule: Pregnant logit regression[a]

	Specification	
	(1)	**(2)**
Wife's age	-0.098	-0.109
	(0.020)	(0.020)
Husband's age	-0.022	-0.108
	(0.012)	(0.012)
Woman's age at marriage	0.083	0.093
	(0.023)	(0.023)
Has a newborn	-1.14	-1.14
	(0.169)	(0.172)
Has a 1 year-old	0.126	0.113
	(0.123)	(0.125)
Has a 2 year-old	-0.162	-0.190
	(0.130)	(0.134)
Has a 3 year-old	0.184	0.203
	(0.116)	(0.117)
Has a 4 year-old	0.068	0.076
	(0.123)	(0.125)
Has a 5 year-old	0.238	0.214
	(0. 121)	(0.124)
Has a 6 year-old	0.347	0.351
	(0.130)	(0.130)
Has a 7 year-old	0.400	0.407
	(0.138)	(0.140)
Has a 8 year-old	-0.004	-0.015
	(0.163)	(0.164)

Has a 9 year-old	0.361 (0.170)	0.335 (0.171)
Has a 10 year-old	0.173 (0.201)	0.135 (0.203)
Has a 11 year-old	0.297 (0.213)	0.258 (0.216)
Has a 12 year-old	0.181 (0.233)	0.159 (0.234)
Has a 13 year-old	0.486 (0.249)	0.479 (0.251)
Has a 14 year-old	-0.152 (0.286)	-0.171 (0.291)
Has a 15 year-old	-0.115 (0.298)	-0.207 (0.305)
Number of children 16 +	-0.211 (0.180)	-0.216 (0.181)
Avg schooling of children 7-11	-0.218 (0.070)	-0.203 (0.070)
Avg schooling of children 12-15	0.021 (0.048)	0.024 (0.048)
Avg schooling of children 16 +	0.006 (0.043)	0.006 (0.043)
Distance to a secondary school	4.86E-5 (2.43E-5)	4.20E-5 (2.74E-5)
Parents income	-7.38E-5 (3.48E-5)	-6.11E-5 (3.55E-5)
Year = 1998	-0.153 (0.124)	-0.136 (0.131)

1997 dummy $\times$ control group dummy	-0.011 (0.117)	0.003 (0.119)
Wife's schooling	-	-0.073 (0.021)
Husband's schooling	-	0.030 (0.020)
State dummies	no	Yes
Constant	15.0 (12.1)	14.1 (12.7)
Pseudo-R^2	0.109	0.116

a. Robust standard errors in parentheses

Table 7
Approximate decision rule:
Logit regression of school attendance[a]

	Specification	
	(1)	**(2)**
Child's age	-0.620 (0.070)	-0.604 (0.073)
Child's schooling	0.399 (0.052)	0.361 (0.053)
Child's sex	0.313 (0.110)	0.309 (0.112)
Wife's age	-0.039 (0.016)	-0.019 (0.016)
Woman's age at marriage	0.028 (0.023)	0.010 (0.024)
Birth order	0.575 (0.189)	0.574 (0.195)
Has a newborn	-0.176 (0.204)	-0.182 (0.203)
Has a 1 year-old	0.040 (0.195)	0.015 (0.201)
Has a 2 year-old	-0.121 (0.172)	-0.148 (0.174)
Has a 3 year-old	-0.128 (0.162)	-0.116 (0.159)
Has a 4 year-old	-0.187 (0.145)	-0.154 (0.148)
Has a 5 year-old	-0.412 (0.140)	-0.337 (0.142)

Has a 6 year-old	-0.355 (0.147)	-0.303 (0.151)
Has a 7 year-old	0.023 (0.148)	0.110 (0.154)
Has a 8 year-old	0.235 (0.141)	0.305 (0.140)
Has a 9 year-old	0.073 (0.136)	0.162 (0.134)
Has a 10 year-old	-0.112 (0.164)	-0.014 (0.165)
Has a 11 year-old	-0.056 (0.174)	0.025 (0.176)
Has a 12 year-old	-0.331 (0.142)	-0.272 (0.145)
Has a 13 year-old	-0.614 (0.157)	-0.571 (0.164)
Has a 14 year-old	-0.650 (0.174)	-0.616 (0.179)
Has a 15 year-old	-0.747 (0.194)	-0.759 (0.197)
No. of children 16 +	-0.622 (0.193)	-0.597 (0.197)
Avg schooling of children 7-11	0.131 (0.059)	0.113 (0.059)
Average schooling of children 12-15	-0.002 (0.042)	0.003 (0.042)
Average schooling of children 16 +	0.025 (0.022)	0.020 (0.023)

Distance to a secondary school	-1.28E-4 (2.78E-5)	-9.64E-5 (2.99E-5)
Parents income	1.53E-4 (2.30E-4)	5.69E-5 (2.29E-4)
Year = 1998	-0.230 (0.131)	-0.167 (0.134)
1997 dummy × control group dummy	-0.158 (0.132)	-0.076 (0.135)
Wife's schooling	-	0.049 (0.028)
Husband's schooling	-	0.097 (0.029)
State dummies	no	yes
Constant	30.6 (12.8)	24.3 (13.1)
Pseudo-R^2	0.308	0.323

a. Robust standard errors in parentheses

Table A.1
Monthly transfers for school attendance under the PROGRESA program

School Level	Grade	Monthly Payment in Pesos	
		Females	Males
Primary	3	70	70
	4	80	80
	5	105	105
	6	135	135
Secondary	1	210	200
	2	235	210
	3	255	225

(a) Source: Schultz (1999a, Table 1). Corresponds to first term of the 1998-99 school year.

Table A.2
Program impacts on proportion enrolled (all children)
(t-statistics in parentheses)

Variable (all are indicator variables)	Oct. 1997	Oct. 1998	Nov. 1999
age 6	0.89 (80.60)	0.92 (83.19)	0.96 (80.30)
age 7	0.94 (85.16)	0.97 (90.87)	0.97 (89.92)
age 8	0.94 (84.71)	0.96 (91.37)	0.98 (95.45)
age 9	0.95 (80.19)	0.96 (87.42)	0.98 (91.09)
age 10	0.94 (84.14)	0.96 (86.40)	0.96 (92.84)
age 11	0.92 (80.72)	0.93 (84.56)	0.95 (84.95)
age 12	0.83 (71.26)	0.81 (74.63)	0.86 (82.51)
age 13	0.69 (55.40)	0.74 (65.17)	0.74 (67.66)
age 14	0.57 (42.89)	0.55 (45.40)	0.63 (57.25)
age 15	0.39 (26.52)	0.43 (33.46)	0.44 (36.85)
age 16	0.31 (16.91)	0.29 (20.52)	0.32 (25.47)
age 17	0.26 (8.06)	*	0.23 (16.83)
age 18	0.07 (0.81)	*	0.16 (10.39)

treatment*age 6	0.02 (1.09)	0.00 (0.16)	0.00 (0.32)
treatment*age 7	0.00 (0.19)	0.00 (0.27)	0.01 (0.73)
treatment*age 8	0.01 (0.64)	0.01 (0.55)	0.01 (0.62)
treatment*age 9	0.00 (0.18)	0.02 (1.42)	0.00 (0.03)
treatment*age 10	0.00 (0.27)	0.0 (0.35)	0.02 (1.52)
treatment*age 11	0.01 (0.59)	0.03 (2.08)	0.02 (1.37)
treatment*age 12	-0.01 (-0.52)	0.09 (6.27)	0.06 (4.48)
treatment*age 13	0.03 (1.66)	0.07 (4.60)	0.10 (7.54)
treatment*age 14	-0.01 (-0.78)	0.16 (10.68)	0.09 (6.67)
treatment*age 15	0.04 (2.21)	0.05 (2.94)	0.09 (6.10)
treatment*age 16	0.02 (0.98)	0.06 (3.10)	0.07 (4.40)
treatment*age 17	-0.01 (-0.12)	*	0.04 (2.17)
treatment*age 18	0.04 (0.36)	*	0.01 (0.35)
p-value from chi-square test that impacts are 0 for all ages	0.5652	<0.0001	<0.0001
p-value from chi-square test that impacts are 0 for ages 12 and older	0.3456	<0.0001	<0.0001